

## Волновая модель распространения и изменения лингвистической информации в индоевропейском модельном языковом сообществе

А. А. Егоров\*<sup>†</sup>, М. А. Егорова<sup>‡</sup>

\* *Институт общей физики им. А.М. Прохорова РАН  
ул. Вавилова, д. 38, Россия, Москва, 119991*

<sup>†</sup> *Учебно-научный Центр Совета ВОИР РУДН*

<sup>‡</sup> *Кафедра иностранных языков факультета гуманитарных и социальных наук  
Российский университет дружбы народов  
ул. Миклуто-Маклая, д. 6, Москва, Россия, 117198*

В работе рассмотрена волновая математическая модель, описывающая распространение и изменение лингвистической информации в индоевропейском модельном языковом сообществе. Дана краткая информация об анатолийской и курганной гипотезах формирования праиндоевропейцев. Математическая модель процесса волнового распространения и изменения информации описывается системой интегро-дифференциальных уравнений. Даны результаты предварительного теоретического анализа и компьютерного моделирования. В частности, получена оценка максимально возможного времени ( $\sim 24000$  лет) развития/возникновения рассматриваемой индоевропейской языковой «семьи» в рамках данной модели. Результаты компьютерного моделирования показывают, что из двух основных гипотез формирования праиндоевропейцев — анатолийской и курганной — последняя лучше соответствует полученным нами временным оценкам. В заключение получены данные о гипотетическом праиндоевропейском алфавите, а именно — о числе возможных «символов/знаков» (или «букв») алфавита праиндоевропейцев на основании полученных нами данных. Исходя из полученных результатов установлено, что диапазон возможных значений для символов алфавита ограничен примерно значениями:  $3 \div 32$ . Кроме того, есть хорошая корреляция наших данных с выводами генетиков о возможном времени появления (примерно 14000–20000 лет назад) гаплогруппы Y-DNA R1a (считается, что носителями праиндоевропейских языков были в первую очередь мужчины с данной гаплогруппой).

**Ключевые слова:** математическая модель, интегрально-дифференциальное уравнение, математическая лингвистика, индоевропейские языки, индоевропейская хронология, индоевропейское распространение, анатолийская и курганная гипотезы

### 1. Введение

В настоящей работе рассмотрена волновая математическая модель, описывающая распространение и изменение лингвистической информации в некоторых сообществах [1, 2]. В рамках этой модели было проведено численное исследование распространения лингвистической информации в модельном индоевропейском (ИЕ) языковом сообществе, в том числе на начальном этапе его формирования.

*Информация* — это (от лат. *informare* — сообщать, изображать, составлять понятие о чём-либо; *informatio* — разъяснение, изложение) сведения, содержащиеся в конкретном высказывании и представляющие собой объект передачи, хранения, переработки, воспроизведения. В более узком смысле под *лингвистической информацией* могут пониматься, например, семантика (определяет соотношение между словами и их значениями) и грамматика (правила, выражающие общие синтаксические свойства слов и групп слов, позволяющие производить и/или описывать правильные предложения языка) некоторого языка; в принципе это может быть и количество групп языков или просто число языков в некоторой языковой семье.

*Бифуркация* (от латинского *bifurcus* — раздвоенный) — появление нового качества в поведении динамической системы при малом изменении её параметров [3, 4]. Если при перестройке объекта он не изменяется во времени, то вместо термина бифуркация можно использовать термин «катастрофа».

При построении математических моделей распространения и изменения лингвистической информации в языковых сообществах в качестве априорной информации использовали данные независимых исследований, как из лингвистики, так и из других научных областей, например, из истории, генетики и археологии [5–20].

В статье даны некоторые результаты предварительного теоретического анализа и компьютерного моделирования. Установлено, что для рассмотренной математической модели процесса распространения и изменения лингвистической информации наиболее характерной чертой является известный в математической лингвистике логистический характер поведения моделируемой динамической системы.

## 2. Число языков. Анатолийская и Курганная гипотезы формирования протоиндоевропейцев

Рассмотрим данные, следующие из оценки временных рамок возникновения и развития языков и их семей. Известно, что сейчас существует по разным данным примерно от 6000 до 8000 языков [9, 11], из которых собственно индоевропейских примерно 300–500. Так, по данным [11], живых индоевропейских языков сейчас примерно 445. В [12] справедливо отмечено о числе живых языков: «В последней версии международного лингвистического каталога “Этнолог” [11] приводится цифра в 7106 единиц, но абсолютной точности здесь быть не может, потому что не существует единого строгого критерия, который позволял бы определить, где кончаются диалекты, то есть региональные варианты, одного языка и начинаются, собственно говоря, разные языки».

Приведённые данные по числу языков позволяют получить следующие временные оценки, в основе которых лежит сделанное в нашей работе по динамической нелинейной модели распространения и изменения лингвистической информации в индоевропейском модельном языковом сообществе предположение [21]: использование нелинейной модели, где есть циклы вида  $S^{2^p}$ , позволяет получить не только «встроенный» внутренний временной масштаб с шагом в 500 лет, но и определить временную «длину» данной лингвистической временной «линейки», вдоль которой развивается динамика исследуемой лингвистической системы. Действительно, для циклов вида  $S^{2^p}$ , где  $p = 0, 1, 2, 3, 4, \dots$  имеем следующий ряд значений для числа языков  $L$ , соответствующего данному ряду последовательных их удвоений, начиная с одного возможного ИЕ праязыка, первоосновы всех современных ИЕ языков: 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024 и т. д., исходя из этого несложно получить следующую простую формулу для оценки возможного числа языков в таком динамическом процессе:  $L = 2^p$ . Откуда получаем формулу для численной оценки количества соответствующих данному числу языков циклов  $p$ :  $p = \log_2 L$ . Например, для 6000 всех языков получаем:  $p = \log_2 6000 \approx 12,55$ , т. е. примерно 13 поколений, что соответствует примерно времени  $13 \cdot 500$  лет = 6500 лет. Только для индоевропейских языков имеем:  $p = \log_2 500 \approx 8,97$ , т. е. примерно 9 поколений или 4500 лет. Учёт вероятных исчезнувших языков и всевозможных диалектов увеличит эти оценки, особенно в случае всех языков, что вполне естественно. С этой точки зрения можно рассматривать данные оценки как оценки времён, когда различные языки окончательно разошлись.

Напомним кратко об Анатолийской и Курганной гипотезах формирования протоиндоевропейцев. *Анатолийская гипотеза* локализует индоевропейскую прародину

в западной Анатолии (современная Турция). Данные, полученные Греем и Аткинсоном методами байесовского анализа, по их мнению, указывают на возраст праиндоевропейского языка в интервале от 8000 до 9500 лет и на анатолийское происхождение языка (см., например, [13]). *Курганная гипотеза* была предложена Марией Гимбутас в 1956 г., чтобы соединить данные археологических и лингвистических исследований для определения местонахождения прародины народов носителей праиндоевропейского языка. Гипотеза является наиболее популярной в отношении происхождения ПИЕ. По ней протоиндоевропейские народы существовали в причерноморских степях и юго-восточной Европе примерно с V по III тысячелетие до н.э. (а возможно и ранее) [5, 6, 13, 14, 18]. Важнейшим этапом в развитии курганной культуры было одомашнивание лошади и использование повозок, что сделало носителей культуры мобильными и существенно расширило их влияние [6, 14]. В курганной гипотезе полагается, что причерноморские степи были прародиной протоиндоевропейцев и по всему этому региону говорили на поздних диалектах праиндоевропейского языка.

Используем полученные данные при численном исследовании распространения и изменения лингвистической информации в некотором модельном индоевропейском (ИЕ) языковом сообществе, в том числе на начальном этапе его формирования (см. также [21]). При этом будем полагать, что время начала разделения (т. е. по сути «исчезновения») гипотетического праиндоевропейского (ПИЕ) языка (в англоязычной литературе используется термин Proto-Indo-European language — PIE) произошло приблизительно не позднее 6500 (курганная гипотеза) или не позднее 9500 (анатолийская гипотеза) лет назад [5, 14].

### 3. Волновая модель распространения и изменения лингвистической информации в сообществе

Как известно, при описании языковых явлений в сообществах наиболее часто употребляются термины: передача и распространение (см., например, [16, 17]). Действительно, лингвисты давно обратили внимание на волновую природу многих языковых явлений (см., например, [16, 17]). Однако в этой области пока предложено мало математических моделей, позволяющих учитывать именно волновой характер различных лингвистических явлений. Многообразие волновых процессов в природе приводит к тому, что никаких абсолютных общих свойств волн выделить не удаётся. Среди многообразия волн можно выделить некоторые типы, которые возникают во многих ситуациях из-за математического сходства описывающих их теоретических моделей. Целый ряд таких моделей описывается волновыми уравнениями.

Для описания процессов распространения информации в языковом сообществе предлагается применить математический аппарат, который широко используется при исследовании различных волновых явлений (см, например, [22–25]). Как известно, скалярное волновое уравнение

$$c^2 \frac{\partial^2 \Psi(z, t)}{\partial z^2} - \frac{\partial^2 \Psi(z, t)}{\partial t^2} = 0,$$

являющееся уравнением гиперболического типа, описывает математические модели различных реальных процессов в биологических, экологических и социальных системах (см., например, [3, 23–25]). В рассматриваемом одномерном случае пространственная координата (ось  $z$ ) направлена вдоль направления распространения волны;  $c$  — некоторая положительная константа (скорость волны). Решение  $\Psi(z, t)$  волнового уравнения известно и даётся формулой Даламбера [22].

Предлагаемая нами математическая модель, описывающая процессы волнового распространения и изменения некоторой информации в анализируемой распределённой системе, может быть представлена в одномерном виде как модификация

исходного волнового уравнения, которая позволяет учесть процессы изменения информации, распространяющейся в системе, в результате взаимодействия как между членами данного ИЕ языкового сообщества, так и между членами данного ИЕ языкового сообщества и членами других (например, соседних) не ИЕ языковых сообществ [1, 2]. Заметим, что в общем случае процесс волнового распространения информации может сопровождаться процессами диффузии информации в окрестности ареалов, куда доходят волны лингвистической информации. Напомним, что по определению *информация* — сведения об объектах и явлениях окружающей среды (живой и неживой природы), а также об их параметрах, свойствах и состояниях. В данном случае интересующие нас сведения о тех или иных информационных волновых процессах в рассматриваемой системе могут быть получены с помощью функции  $\Psi$ . На данном этапе исследования из всех возможных характеристик волнового процесса нас интересуют в первую очередь две: амплитуда волны  $\Psi$  и интенсивность волны  $|\Psi|^2$ . Более того, полученное в итоге решение, как среднее по ансамблю статистически идентичных лингвистических систем (из данного языкового сообщества), описывает по сути интересующую нас информацию  $I$  (скалярная функция), которая определяется с помощью величины  $|\Psi|^2$ . Для удобства решения все величины могут быть приведены к безразмерному виду с использованием характерных параметров системы: размера  $L_c$  (критический размер области, занятый сообществом, ниже которого начинается его вымирание), частоты  $f_c = 1/T_c$  (в качестве  $T_c$  может быть взято, например, время разделения 2-х языков, т. е. 500 лет) и времени  $T_c$ . Для модельных расчётов можно положить начальное значение  $|\Psi_0|^2 = 1$  либо использовать полученные в первой модели [22] значения  $I$  как некоторые начальные значения.

Математическая модель процесса волнового распространения и изменения информации в рассматриваемой системе описывается следующей системой интегродифференциальных уравнений:

$$\frac{\partial^2 a_n}{\partial z^2} - \dot{\beta}_n \frac{\partial a_n}{\partial z} + \sum_n a_m B_{mn}(z) + \int q(\rho, z) D_n(\rho, z) d\rho = 0, \quad (1)$$

где  $a$  и  $q$  — некоторые коэффициенты (собственные значения);  $\dot{\beta}_n$  — комплексный волновой параметр системы;  $B_{mn}$ ,  $D_n$  — некоторые функции, определяемые с помощью собственных (базисных) скалярных функций системы.

При выводе (1) предполагалось, что рассматриваемая система имеет определённые ограничения в пространстве, оказывающие влияние на рассматриваемый волновой процесс, а именно полагалось, что информационный поток распространяется в ограниченном по ширине канале. Применительно к некоторым экологическим и социальным системам подобные волновые и квазиволноводные явления были достаточно подробно рассмотрены нами ранее.

Приближённое искомое решение ищется с помощью спектрального метода в следующем виде:

$$\Psi(X, t) = \sum_n a_n(\rho, t) \Psi_n(X_n, t) + \int q(\rho, t) \Psi_p(X_p, t) d\rho, \quad (2)$$

где первое слагаемое в правой части (2) определяет величину «информации» в распределённой системе, обусловленную контактами «незнающих» со «знающими» (сумма по дискретному набору значений информации у её  $N < +\infty$  носителей); второе информационное слагаемое определяется как интеграл (т. е. в общем случае это может быть некоторый континуум, включающий только «незнающих»);  $\rho \geq 0$ .

Коэффициенты  $a$  и  $q$  могут рассматриваться как некоторые эффективные амплитуды (весовые коэффициенты) в соответствующих (дискретном и интегральном)

разложениях. Например, в случае соперничества в рассматриваемом лингвистическом сообществе формирующих его этносов коэффициент  $q \geq 0$  может характеризовать утрату «пассионарной» энергии [25]. В самом простом случае коэффициент  $q$  можно задать в виде матрицы, элементами которой являются действительные положительные числа. В общем случае коэффициенты  $a_n$  и  $q_i$  могут меняться как во времени, так и в зависимости от места (ландшафта) действия, описываемого с помощью множества  $X$ . Учитывая положительную определённость информации  $I \geq 0$  [26], мы, естественно, определяем информацию следующим образом:  $I \propto |\Psi|^2$ , поскольку в общем случае  $\Psi$  — волновая функция, которая по определению является комплекснозначной величиной; в частном случае это может быть действительная функция  $U$ , характеризующая некоторое возмущение (волну) в рассматриваемой системе. Заметим, что наш теоретический подход в некоторой степени аналогичен подходу, использованному, в частности, в работах [23, 25, 27] при построении математических моделей динамики полей различной природы.

Ниже на рис. 1 приведены два типичных графика, характеризующие некоторые стационарные распределения информации (среднее по ансамблю статистически идентичных лингвистических систем), получающиеся в рамках данной комплексной нелинейной модели распространения лингвистической информации в данной системе. Здесь [21] проявляется явление самоорганизации, которое в общем случае не следует сводить только к установлению пространственного порядка в рассматриваемой открытой распределённой системе.

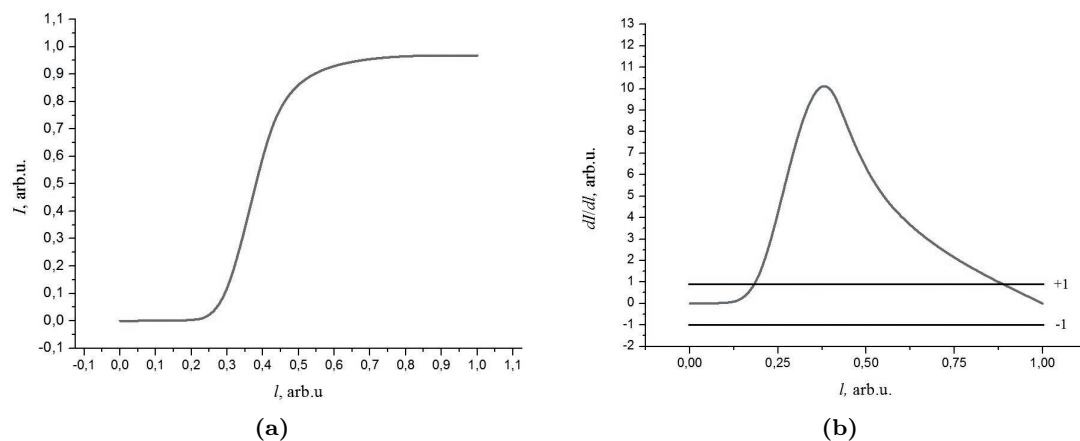


Рис. 1. (а) — график зависимости функции  $I(X)$  от нормированной величины  $l$ ; (б) — график первой производной зависимости, приведённой слева

Используя методы теории катастроф [3, 4], можно пояснить характер поведения рассматриваемой открытой лингвистической системы в зависимости, например, от параметра системы  $l$  (нормированная длина участка, на котором распространяется исходная информация). С этой целью исследуем график первой производной данной зависимости. Эту функцию по аналогии с теорией катастроф можно рассматривать как некоторую потенциальную функцию диссипативной системы, имеющую минимумы (локальные и глобальные). Для определения интервала значений функции, где модуль первой производной не превышает 1, проводятся горизонтальные линии  $+1$  и  $-1$ . В этом интервале соблюдается условие устойчивости особой точки (точек), где  $|dI/dl| \leq 1$ .

Полученный на рис. 1 справа график первой производной соответствует известному случаю бифуркации состояния равновесия («частица» в потенциальной яме с

барьером или полочкой). Здесь возможны два состояния равновесия системы: при  $l < 0,25$  и при  $l > 0,75$ , а между ними есть высокий потенциальный барьер. Здесь можно применить понятие фазового перехода, при котором происходит качественное изменение системы. Например, если система начинает движение из любой точки участка  $0 < l < 0,35$ , то итерации  $I_n$  сходятся к  $I^* = 0$  — устойчивая неподвижная точка. Аналогично, если система начинает движение из любой точки участка  $0,35 < l < 0,5$ , то интегрированные значения  $I_n \rightarrow \text{const}$ , т. е. динамический режим становится стационарным или имеет период, равный единице: возникает цикл  $S^1$ . Проведённые расчёты показали, что при вариации параметров в системе возможны также и другие циклы типа  $S^{2p}$ .

Из рис. 1 видно, что для этой модели наиболее характерной чертой является известный в математической лингвистике логистический характер поведения моделируемой динамической системы [2, 5–8]. Полученные результаты позволяют в частности высказать предположение, что около 3500–4000 лет назад в рассматриваемом модельном ИЕ языковом сообществе могли возникнуть 2 основных лингвистических популяции, характеризующиеся сейчас как деление индоевропейских языков на так называемые языковые ареалы «сатем-кентум» [6, 8]. Лингвисты предполагают, что такое деление могло возникнуть около 2000–2500 гг. до н.э. Действительно, если две лингвистических популяции стартуют в окрестности точки экстремума  $l = 0,37$  одна слева, а другая справа от неё, то они полностью расходятся примерно за 8–10 суммарных итераций (количество поколений в ансамбле лингвистических систем), что соответствует примерно 2000 г. до н.э. Однако заметим, что возможны и другие изоглоссы, которые делят на две подгруппы индоевропейские языки иначе, чем «сатем-кентум» [8], что, очевидно, представляет интерес при дальнейшем исследовании.

Заметим, что в данной модели мы используем в качестве «встроенного» внутреннего временного масштаба (с шагом в 500 лет) последовательность итераций, при которых происходит последовательное увеличение площади, занимаемой развивающимся сообществом. При этом предполагаем, что каждая итерация даёт последовательное увеличение языкового разнообразия в сообществе, т. е. рост числа языков является также следствием роста площади, на которую распространяется исходное сообщество. Как следствие, можно также определить временную «длину» лингвистической «линейки», вдоль которой развивается динамика динамической распределённой системы.

В волновой модели для ряда полученных графиков (характерных S-кривых), подобных графику слева на рис. 1, величина «лингвистической информации», после перехода динамической системой критической точки  $l = 0,37$ , может достигать максимума примерно за 4–8 поколений в случае цикла  $S^2$ , и за примерно 8–48 итераций в случае цикла  $S^1$ , т. е. в среднем чуть больше, чем в первой модели [22]. Скорее всего, число 48 определяет возможное максимальное время ( $\sim 24000$  лет) развития (или существования) рассматриваемой языковой «семьи» в рамках данной теоретической модели. Подчеркнём, что в дальнейшем планируется уточнение сделанных при исследовании предположений, а также выбранных параметров, в том числе с учётом новых лингвистических и генетических данных.

Приведём ещё одну полезную оценку для возможного времени возникновения данного сообщества. С этой целью воспользуемся методикой определения времени диффузии  $T_{\text{dif}}$  системы в минимум по величине кривизны ( $d^2I/dl^2$ ) потенциальной функции системы в точках максимума и минимума [4]. В результате находим нижнюю оценку для времени диффузии сообщества:  $T_{\text{dif}} \approx 2600 \div 4900$  лет. Такой большой разброс определяется в первую очередь погрешностью определения критического размера  $L_c$  области, занятого лингвистической популяцией, ниже которого начинается её вымирание, поскольку имеющиеся исторические и археологические данные не позволяют определить эту величину абсолютно точно (если это вообще возможно).

В этой связи полезно напомнить, что с формальной точки зрения можно вести отсчёт времени существования языков от момента появления речевого аппарата у “*Homo sapiens*” около 100 000 лет назад (по данным археологии) [9]. Однако предполагается, что «язык в его современном виде мог возникнуть на земле около 40 000 лет назад» [9]. Однако там же в [9] отмечено, что «на самом деле язык вполне мог возникнуть одновременно с появлением *Homo sapiens* более 100 000 лет назад» [9]. С этой точки зрения наша временная оценка в 24000 лет уже не выглядит чрезмерно завышенной, но тогда остаётся вопрос о типе языковой семьи. Скорее всего, в то время это могло быть сообщество по крайней мере трёх находящихся в контакте языковых семей: праиндоевропейской, прауральской (возможно прауральско-алтайской) и пракавказской.

В заключение посмотрим, возможно, ли что-то сказать о гипотетическом ПИЕ алфавите, а именно — о числе возможных «символов/знаков» (или «букв») алфавита праиндоевропейцев на основании полученных нами данных.

Получим оценку исходя из данных, полученных для волновой модели. С этой целью используем следующую формулу хорошо известную в теории информации [28]:

$$I_0 = (2F_m T + 1) \log_2 \sqrt{(P_S + P_N)/P_N}, \quad (3)$$

где  $F_m$  — (максимальная) частота, с которой модулируется передаваемое сообщение, например, речевое;  $T$  — длительность сообщения; число различных символов алфавита  $\mathfrak{S}$  при средних мощностях сигнала  $P_S$  (величина некоторой передаваемой в сообществе лингвистической информации  $I_0$ ) и шума  $P_N$  (величина некоторых «помех» передаваемой в сообществе лингвистической информации) принято равным величине:  $\mathfrak{S} = \sqrt{(P_S + P_N)/P_N}$ . Величина  $P_S/P_N$  в (3) — отношение сигнала к шуму.

Для ненормированной зависимости  $I(X)$  на графике слева на рис. 1 находим:  $I_{0,\max} = 1493$  бита (участок кривой близкий к насыщению справа). В средней части кривой, сразу после перехода динамической системой критической точки  $l = 0,37$  вправо, получаем искомую оценку:  $I_0 \approx I_{0,\max}/2 = 747$  бит, т. е. несколько больше, чем аналогичная величина (примерно 523 бита), полученная в первой модели в аналогичной точке (однако в пределах ансамбля есть реализации, где эти отличия минимальны). При оценке величины  $I_0$  использованы: табличные расчётные данные, величина  $P_N = 1$  (погрешность вычисления данных завышена с учётом возможных реальных «помех»), частоту основного тона (у разных людей — мужчин, женщин, детей — находится примерно в диапазоне 50–450 Гц) взяли небольшой  $F_m = 75$  Гц (или  $s^{-1}$ ), а  $T = 3$  с. В результате получили:  $I_0 = (2 \cdot 75 \cdot 3 + 1) \log_2 \sqrt{(490 + 5)/5} \approx 1493$  бита.

Возьмём снова данные с рис. 1 для величины  $I(X)$  и оценим возможное число символов алфавита  $\mathfrak{S}$ . При варьировании уровня сигнала и фиксированном уровне шума мы можем получить ряд значений для  $\mathfrak{S}_a = \sqrt{(P_{S_a} + P_N)/P_N}$ . Рассмотрим для примера диапазон изменения уровня сигнала  $P_{S_a}$  от 1 до 200 при  $P_{N_1} = 0,01$ . Возьмём ряд значений  $P_{S_n}$  из данного диапазона и найдём возможное число знаков в алфавите  $\mathfrak{S}$ :  $P_{S_1} = 1$ ,  $\mathfrak{S}_{S_1} \approx 10$ ;  $P_{S_{10}} = 10$ ,  $\mathfrak{S}_{S_{10}} \approx 32$ ;  $P_{S_{20}} = 20$ ,  $\mathfrak{S}_{S_{20}} \approx 45$ . Для сравнения приведём аналогичные значения при более реальной величине  $P_{N_2} = 0,1$ :  $P_{S_1} = 1$ ,  $\mathfrak{S}_{S_1} \approx 3$ ;  $P_{S_{10}} = 10$ ,  $\mathfrak{S}_{S_{10}} \approx 10$ ;  $P_{S_{30}} = 30$ ,  $\mathfrak{S}_{S_{30}} \approx 17$ ;  $P_{S_{50}} = 50$ ,  $\mathfrak{S}_{S_{50}} \approx 22$ ;  $P_{S_{80}} = 80$ ,  $\mathfrak{S}_{S_{80}} \approx 28$ ;  $P_{S_{100}} = 100$ ,  $\mathfrak{S}_{S_{100}} \approx 32$ .

Мы можем оценить некоторую минимальную величину «помех» при передаче лингвистической информации в данном модельном сообществе. Получаем, что уровень  $P_{N,\min}$  составляет в данном случае примерно  $1/500 = 0,002$  от заданного уровня «сигнала» и не может быть ниже этой величины (хотя в численных расчётах в отсутствие «помех» эта величина, конечно, может быть ниже, поскольку ограничивается в основном точностью численных расчётов). Как видно из приведённых примеров,

в первом случае  $P_{S_a}$  удовлетворяет этому ограничению ( $P_{S_a} \geq P_{N,\min}$ ) только при уровне  $P_{S_a} < 20$ , начиная с которого  $P_{N_1}/P_{S_a} \geq 0,002$ . А во втором случае — при уровне  $P_{S_a} \leq 50$ . При низком уровне «помехи» эти диапазоны будут немного шире.

Исходя из полученных результатов приходим к выводу, что диапазон возможных значений для символов алфавита ограничен следующими значениями: в первом случае это  $\mathfrak{Z}_n \approx 10 \div 32$ , а во втором —  $\mathfrak{Z}_n \approx 3 \div 22$ . Как следствие, получаем, что  $\mathfrak{Z}_n$  ограничено диапазоном значений:  $10 \div 22$ . В случае же объединения двух множеств получим более широкий диапазон:  $\mathfrak{Z}_n = 3 \div 32$ . Полученные данные соответствуют данным нашей работы [21].

Для сравнения приведём известные в научной литературе данные по числу знаков в алфавитах некоторых языков.

*Финикийское письмо* является одной из первых алфавитных письменностей в мире. Появилась около 3500 лет назад. Финикийская письменность дала начало нескольким ветвям алфавитных письменностей, и на сегодняшний день практически все алфавитные письменности мира имеют корни именно в финикийском письме. Отличительные особенности: использовался консонантный принцип, то есть для записи слов использовались только согласные звуки, а значение гласных оставлялось на понимание читателя; текст записывался справа налево. Число знаков в алфавите: 22.

*Греческий алфавит*. Время создания алфавита: около VIII века до н.э. Число знаков в алфавите: 24 буквы. В доклассическую эпоху в некоторых диалектах греческого применялись ещё 6 букв. Происхождение: из финикийского алфавита. Греческий алфавит развился в следующие виды письменности: готское письмо, глаголица, кириллица, коптское письмо, армянское письмо, этрусский алфавит, латинский алфавит.

*Глаголица* — одна из первых славянских азбук. Языки: старославянский, церковнославянский, ранние славянские языки. Тип письма: консонантно-вокалическое письмо. Время создания алфавита: примерно IX век н.э. Направление письма: слева направо. Число знаков в алфавите: 41. Происхождение: греческий (финикийский).

*Кириллица* — кириллический (или кирилловский) алфавит; один из двух (наряду с глаголицей) древних алфавитов для старославянского языка. Тип письма: консонантно-вокалическое письмо. Языки: старославянский, церковнославянский, русский, сербский, болгарский, македонский, украинский, молдавский и многие другие. Время создания алфавита: примерно IX век; в современном виде с конца XVI века. Состав первоначальной кириллической азбуки неизвестен; «классическая» старославянская кириллица из 43 букв.

*Санскрит* — древний литературный язык Индии со сложной синтетической грамматикой. Само слово «санскрит» означает «обработанный, совершенный». Возраст ранних памятников доходит до 3500 лет. Классификация: индоевропейская семья, индоиранская ветвь, индоарийская группа. Единой системы письменности для санскрита не существует. Это объясняется тем, что литературные произведения передавались в основном устно. С примерно XIX века используется письменность деванагари и др. Классический санскрит насчитывает около 36 фонем<sup>1</sup>. Если учитывать аллофоны<sup>2</sup> (а система письма их учитывает), то общее количество звуков в языке увеличивается до 48.

С учётом данных по древнему санскриту (примерно 3700 лет назад) из книги [29] можно оценить минимальное возможное число «знаков» в алфавите прасанскрита величиной:  $\mathfrak{Z} = 9$ .

<sup>1</sup>Фонема (от др.-греч. *φώνημα* — «звук») — минимальная смысловозначительная единица в языке. Фонема не имеет самостоятельного лексического или грамматического значения, она служит для различения и отождествления значимых единиц языка, таких как морфемы и слова. Морфема — наименьшая единица языка, имеющая некоторый смысл.

<sup>2</sup>Аллофон (от греч. *ἄλλος* — другой, и *φωνή* — звук) — реализация фонемы, её вариант, обусловленный конкретным фонетическим окружением. В отличие от фонемы аллофон является не абстрактным понятием, а конкретным речевым звуком.



Как видно из приведённых данных для этих языков, они в основном удовлетворяют заданному диапазону:  $\mathfrak{S}_n = 3 \div 32$  символов.

Полученные результаты, по сути, позволяют сделать вывод, что предложенная нами математическая волновая модель распространения и изменения лингвистической информации в сообществе достаточно адекватно описывает исследуемое явление. Вместе с тем проведённое исследование показало, что необходима дальнейшая тонкая настройка моделей, в том числе с учётом априорной лингвистической и другой информации об исследуемых языковых сообществах.

Итак, в настоящей работе приведены результаты компьютерного моделирования, которые показывают в частности хорошее соответствие двум основным гипотезам о формировании праиндоевропейцев: анатолийской и курганной. Кроме того, есть хорошая корреляция наших данных с выводами генетиков о возможном времени появления (примерно 14000–20000 лет назад) гаплогруппы Y-DNA R1a (считается, что носителями ПИЕ-языков были в первую очередь мужчины с данной гаплогруппой) [30, 31]. При этом совокупность данных, полученных при исследовании обеих моделей, говорит о предпочтительности курганной гипотезы. Данный вывод хорошо согласуется с выводами большинства исследователей во всём мире: лингвистов, генетиков, историков и археологов.

#### 4. Заключение

В данной работе изложены теоретические принципы нового метода исследования лингвистических сообществ (и процессов распространения в них лингвистической информации) как динамических диссипативных систем. Этот подход может стать особенно перспективным в первую очередь для качественного анализа поведения подобных лингвистических систем. Результаты компьютерного моделирования показывают, что курганная гипотеза формирования праиндоевропейцев лучше соответствует полученным временным оценкам.

Полученные в статье результаты позволяют также высказать предположение, что около 3500–4000 лет назад в рассматриваемом модельном ИЕ языковом сообществе могли возникнуть две основные лингвистические популяции, характеризующие сейчас как деление индоевропейских языков на так называемые языковые ареалы «сатем-кентум». Хотя возможны и другие изоглоссы, которые делят на две подгруппы индоевропейские языки иначе, чем «сатем-кентум».

#### Литература

1. Егоров А. А. Некоторые закономерности распространения информации в обществе // Сборник тезисов докладов Научно-технической конференции «Сети связи и сетевые технологии», 24–26 июня 1997, Суздаль. — М.: ЦНИИС, 1997. — С. 46–49.
2. Егоров А. А., Егорова М. А. О моделях распространения лингвистической информации в языковом сообществе // Тезисы докладов XXI-й Всероссийской конференции «Теоретические основы конструирования численных алгоритмов и решение задач математической физики». 5–11 сентября 2016 г., Новороссийск, Абрау-Дюрсо, Россия. — М.: ИПМ им. М.В. Келдыша РАН, 2016. — С. 82–83.
3. Компьютеры и нелинейные явления: Информатика и современное естествознание. — 1988.
4. *Gilmore R.* Catastrophe Theory for Scientists and Engineers. — New York: Wiley, 1981.
5. Ancestry-Constrained Phylogenetic Analysis Supports the Indo-European Steppe Hypothesis / W. Chang, C. Cathcart, D. Hall, A. Garrett // *Language*. — 2015. — Vol. 91, No 1. — Pp. 194–244.
6. *Anthony D. W.* The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World. — Princeton: Princeton University Press, 2007.

7. *Kornai A.* Mathematical Linguistics. — London: Springer, 2008.
8. *Бурлак С. А., Старостин С. А.* Сравнительно-историческое языкознание. — М.: Издательский центр «Академия», 2005.
9. Атлас языков мира. Происхождение и развитие языков во всем мире. — 1998.
10. *Яхонтов С. Е.* Теоретические основы классификации языков мира. — М., 1980. — С. 148–157.
11. Ethnologue: Languages of the World. — [www.ethnologue.com](http://www.ethnologue.com).
12. *Старостин Г. С. и др.* К истокам языкового разнообразия. — М.: Издательский дом «Дело» РАНХиГС, 2015.
13. *Gray R. D., Atkinson Q. D.* Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origin // *Nature*. — 2003. — Vol. 426. — Pp. 435–439.
14. *Pereltsvaig A., Lewis M. W.* The Indo-European Controversy: Facts and Fallacies in Historical Linguistics. — Cambridge: Cambridge University Press, 2015.
15. Развитие модели распространения информации в социуме / А. П. Михайлов, А. П. Петров, Н. А. Маревцева, И. В. Третьякова // Математическое моделирование. — 2014. — Т. 26, № 3. — С. 65–74.
16. *Labov W.* Transmission and Diffusion // *Language*. — 2007. — Vol. 83. — Pp. 344–387.
17. *Heggarty P., Maguire W., McMahon A.* Splits or waves? Trees or webs? How Divergence Measures and Network Analysis Can Unravel Language Histories // *Phil. Trans. R. Soc. B*. — 2010. — Vol. 365. — Pp. 3829–3843.
18. Massive Migration from the Steppe is a Source for Indo-European Languages in Europe / W. Haak, I. Lazaridis, N. Patterson et al. // *Nature*. — 2015. — Vol. 522. — Pp. 207–211.
19. Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans / I. Lazaridis, N. Patterson, A. Mittnik et al. // *Nature*. — 2014. — Vol. 513. — Pp. 409–428.
20. Аркаим – Синташта: древнее наследие Южного Урала. Сборник научных трудов. — Челябинск: изд. Челябинского гос. ун-та, 2010.
21. *Егоров А. А., Егорова М. А.* Динамическая нелинейная модель распространения и изменения лингвистической информации в индоевропейском модельном языковом сообществе // Вестник РУДН. Серия «Математика. Информатика. Физика». — 2016. — С. 37–48.
22. *Арсенин В. Я.* Методы математической физики и специальные функции. — М.: Наука, 1974.
23. *Нахушев А. М.* Уравнения математической биологии. — М.: Высшая школа, 1995. — С. 301.
24. *Мари Д.* Нелинейные дифференциальные уравнения в биологии. — М.: Мир, 1983.
25. Математические модели социальных систем / А. К. Гуц, В. В. Коробицын, А. А. Лаптев, Л. А. Паутова, Ю. В. Фролова. — Омск: ОГУ, 2000.
26. *Яглом А. М., Яглом И. М.* Вероятность и информация. — М.: Наука, 1973.
27. *Mathematical Modeling* / Ed. by J. G. Andrews, R. R. McLone. — London: Butterworths, 1976.
28. *Гоноровский И. С.* Радиотехнические цепи и сигналы. Ч. 1. — М.: Радио, 1966.
29. *Vorsboom W.* Alphabet or Abracadabra? Reverse Engineering the Western Alphabet. — Victoria, BC, Canada: FriesenPress, 2015.
30. Separating the Post-Glacial Coancestry of European and Asian Y Chromosomes within Haplogroup R1a / P. A. Underhill, N. M. Myres, S. Rootsi, M. Metspalu, M. A. Zhivotovsky, R. J. King et al. // *European Journal of Human Genetics*. — 2010. — Vol. 18. — Pp. 479–484.
31. *Klyosov A. A., Rozhanskii I. L.* Haplogroup R1a as the Proto Indo-Europeans and the Legendary Aryans as Witnessed by the DNA of Their Current Descendants // *Advances in Anthropology*. — 2012. — Vol. 2, No 1. — Pp. 1–13.

UDC 517+519.2:80

DOI: 10.22363/2312-9735-2017-25-1-69-80

## The Wave Model of the Propagation and Changes of Linguistic Information in the Indo-European Model Language Community

A. A. Egorov\*<sup>†</sup>, M. A. Egorova<sup>‡</sup>

\* *General Physics Institute, Russian Academy of Sciences  
38, Vavilov str., Moscow, Russia, 119991*

<sup>†</sup> *Educational and Scientific Center of the RUDN' VOIR Council*

<sup>‡</sup> *Foreign Languages Department of Humanitarian and Social Sciences Faculty  
RUDN University (Peoples' Friendship University of Russia)  
6, Miklukho-Maklaya str., Moscow, Russia, 117198*

Paper considers the wave mathematical model describing the distribution and change of linguistic information in Indo-European model linguistic community. Brief information about the Anatolian and Kurgan hypothesis of formation of Proto-Indo-Europeans (PIE) is given. The mathematical model of wave propagation and changing of information is described by system of integral-differential equations. Results of the preliminary theoretical analysis and computer modeling are given. In particular, an estimate of the maximum possible time ( $\sim 24000$  years) of development/emergence of considered Indo-European “family” is obtained in the framework of this model. Results of computer modeling show, that from two main hypotheses of formation of the PIE — Anatolian and Kurgan, the latter better matches obtained by us temporary estimates. Finally, data on a hypothetical PIE alphabet — namely, the number of possible “characters/symbols” (or “letters”) alphabet of PIE on the basis of our data were obtained. Based on the results, it was found that the range of possible values for alphabetic characters is limited to about values:  $3 \div 32$ . In addition, there is a good correlation of our data with the findings of geneticists on a possible time of occurrence ( $\sim 14000$ – $20000$  years ago) of the haplogroup Y-DNA R1a (it is believed that PIE language speakers were primarily men with this haplogroup).

**Key words and phrases:** mathematical model, integral-differential equation, mathematical linguistics, Indo-European languages, Indo-European chronology, Indo-European dispersal, Anatolian and Kurgan hypothesis

## References

1. A. A. Egorov, Some Patterns of Dissemination of Information in Society, in: Proc. of Conference “Communication Networks and Network technologies”, June, 24–26, 1997, Moscow, 1997, pp. 46–49, in Russian.
2. A. A. Egorov, M. A. Egorova, About Models of Dissemination of Linguistic Information in the Language Community, in: Proc. of XXI-th All-Russian Conference “Theoretical Foundations and Designing of Numerical Algorithms of Solution of Mathematical Physics Problems”, September 5–11, 2016, Novorossiysk, Russia, IAM of RAS, Moscow, 2016, pp. 82–83, in Russian.
3. E. of foreword A. Samarskii (Ed.), Computers and Nonlinear Phenomena: Information Technology and Modern Science, Science, Moscow, 1988, in Russian.
4. R. Gilmore, Catastrophe Theory for Scientists and Engineers, Wiley, New York, 1981.
5. W. Chang, C. Cathcart, D. Hall, A. Garrett, Ancestry-Constrained Phylogenetic Analysis Supports the Indo-European Steppe Hypothesis, *Language* 91 (1) (2015) 194–244.
6. D. W. Anthony, The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World, Princeton University Press, Princeton, 2007.
7. A. Kornai, Mathematical Linguistics, Springer, London, 2008.
8. S. A. Burlak, S. A. Starostin, Comparative and Historical Linguistics, Publishing center “Academy”, Moscow, 2005, in Russian.

9. Atlas of the World's Languages. The Origin and Development of Languages Worldwide, Leake press, Moscow, 1998.
10. S. E. Yahontov, Assessment of the Degree of Closeness of Related Languages, Moscow, 1980, Ch. Theoretical Bases of Classification of World Languages, pp. 148–157, in Russian.
11. Ethnologue: Languages of the World.  
URL [www.ethnologue.com](http://www.ethnologue.com)
12. G. S. Starostin, et al., To the Source of Linguistic Diversity, Publishing House “Delo” RANHiGS, Moscow, 2015, in Russian.
13. R. D. Gray, Q. D. Atkinson, Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origin, *Nature* 426 (2003) 435–439.
14. A. Pereltsvaig, M. W. Lewis, The Indo-European Controversy: Facts and Fallacies in Historical Linguistics, Cambridge University Press, Cambridge, 2015.
15. A. P. Mikhailov, A. P. Petrov, N. A. Marevtseva, I. V. Tretyakova, Development of a Model of Information Dissemination in Society, *Mathmod.* 26 (3) (2014) 65–74, in Russian.
16. W. Labov, Transmission and Diffusion, *Language* 83 (2007) 344–387.
17. P. Heggarty, W. Maguire, A. McMahon, Splits or Waves? Trees or Webs? How Divergence Measures and Network Analysis Can Unravel Language Histories, *Phil. Trans. R. Soc. B.* 365 (2010) 3829–3843.
18. W. Haak, I. Lazaridis, N. Patterson, et al., Massive Migration from the Steppe is a Source for Indo-European Languages in Europe, *Nature* 522 (2015) 207–211.
19. I. Lazaridis, N. Patterson, A. Mittnik, et al., Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans, *Nature* 513 (2014) 409–428.
20. Arkaim – Sintashta: the Ancient Heritage of the Southern Urals. Coll. of Scientific Works, Chelyabinsk State University, Chelyabinsk, 2010, in Russian.
21. A. A. Egorov, M. A. Egorova, Dynamic Non-Linear Model of Distribution and Changes of Linguistic Information in the Indo-European Model Language Community, *Bulletin of Peoples' Friendship University of Russia. Series “Mathematics. Information Sciences. Physics”* (2016) 37–48 In Russian.
22. V. Y. Arsenin, *Methods of Mathematical Physics and Special Functions*, Science, Moscow, 1974, in Russian.
23. A. M. Nahushev, *Equations of Mathematical Biology*, Higher School, Moscow, 1995, in Russian.
24. J. Murray, *Non-Linear Differential Equations in Biology*, Mir, Moscow, 1983, in Russian.
25. A. K. Guts, V. V. Korobitsyn, A. A. Laptev, L. A. Pautova, J. V. Frolova, *Mathematical Models of Social Systems*, OSU, Omsk, 2000, in Russian.
26. A. M. Yaglom, I. M. Yaglom, *Probability and Information*, Science, Moscow, 1973, in Russian.
27. J. G. Andrews, R. R. McLone (Eds.), *Mathematical Modeling*, Butterworths, London, 1976.
28. I. S. Gonorovsky, *Radio Circuits and Signals. Part 1, Radio*, Moscow, 1966, in Russian.
29. W. Borsboom, *Alphabet or Abracadabra? Reverse Engineering the Western Alphabet*, FriesenPress, Victoria, BC, Canada, 2015.
30. P. A. Underhill, N. M. Myres, S. Rootsi, M. Metspalu, M. A. Zhivotovsky, R. J. King, et al., Separating the Post-Glacial Coancestry of European and Asian Y Chromosomes within Haplogroup R1a, *European Journal of Human Genetics* 18 (2010) 479–484.
31. A. A. Klyosov, I. L. Rozhanskii, Haplogroup R1a as the Proto Indo-Europeans and the Legendary Aryans as Witnessed by the DNA of Their Current Descendants, *Advances in Anthropology* 2 (1) (2012) 1–13.