
Математическое моделирование

УДК 517+519.2:80

Динамическая нелинейная модель распространения и изменения лингвистической информации в индоевропейском модельном языковом сообществе

А. А. Егоров*, М. А. Егорова†

* *Институт общей физики им. А.М. Прохорова РАН, Москва, Россия*

† *Российский университет дружбы народов, Москва, Россия*

В работе рассмотрена нелинейная динамическая математическая модель, описывающая распространение и изменение лингвистической информации в индоевропейском языковом сообществе. При построении математической модели распространения и изменения лингвистической информации в языковом сообществе в качестве априорной информации использовали данные независимых исследований, как из лингвистики, так и из других научных областей, например, из истории, генетики и археологии. В рамках этой модели было проведено численное исследование распространения лингвистической информации в некотором модельном индоевропейском языковом сообществе, в том числе на начальном этапе его формирования. Даны результаты предварительного теоретического анализа и компьютерного моделирования. Установлено, что данная математическая модель процесса распространения и изменения лингвистической информации демонстрирует как регулярное, так и типично хаотическое поведение. В качестве одной из количественных характеристик рассматриваемого нелинейного процесса распространения лингвистической информации предложено рассматривать число возникающих циклов в качестве числа возникших новых языков в данном языковом сообществе. Результаты компьютерного моделирования показывают, что из двух основных гипотез формирования праиндоевропейцев — Анатолийской и Курганной — последняя лучше соответствует полученным нами временным оценкам.

Ключевые слова: математическая модель, нелинейное уравнение, математическая лингвистика, индоевропейские языки, индоевропейская хронология, индоевропейское распространение, Анатолийская и Курганная гипотезы

1. Введение

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

Обычно для информации используют следующее определение. *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 бифуркации происходит

перестройка характера движения исследуемой системы, например, физической, химической, биологической или социальной. Нередко требуется только ограниченный объём информации, причём качественного характера, которая и представляет основной интерес при изучении конкретных объектов и описывающих их систем уравнений. Бифуркации происходят в точках бифуркации (критические точки); при этих значениях параметров системы она меняет своё поведение, — происходит качественное изменение свойств системы. Если состояние системы можно описать некоторой потенциальной функцией, то в бифуркационном множестве происходит переход из одного локального минимума (устойчивое состояние системы) в другой или фазовый переход.

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 данная математическая модель процесса распространения и изменения лингвистической информации демонстрирует как регулярное, так и типично хаотическое поведение.

2. Динамическая нелинейная модель распространения и изменения лингвистической информации в сообществе (древовидная модель)

Динамическая модель распространения и изменения «лингвистической информации» в сообществе может быть описана нелинейным уравнением [1, 2]:

$$I_{m+1} = [a_1 I_m (M - I_m) + a_2 (M - I_m)^2] \lambda, \quad (1)$$

где I — величина анализируемой лингвистической информации, $m = 1, 2, \dots$ ($m = 1$ соответствует первому «измерению», т. е. I_1 — начальное значение данной информации, например, в некоторый начальный момент времени t_1); a_1 — коэффициент, характеризующий распространение лингвистической информации при контактах «незнающих» со «знающими» данную информацию I_n ; a_2 — коэффициент, характеризующий воздействие только на «незнающих»; M — максимальное значение данной лингвистической информации; λ — параметр задачи (в соответствии с теорией катастроф он может быть назван управляющим параметром). Нелинейное уравнение (1) позволяет исследовать процесс изменения распространяющейся лингвистической информации, как в зависимости от времени, так и от параметров входящих в данное уравнение. Данное уравнение и его варианты могут использоваться в частности при исследовании процесса обучения, например, детей взрослыми в некотором лингвистическом сообществе.

Преобразуем уравнение (1) к следующему виду:

$$y = \lambda_1 x(1 - x) + \lambda_2 x(1 - x)^2, \quad (2)$$

где сделана замена переменных: $y = I_{m+1}/M$, $x = I_m/M$; и $\lambda_1 = a_1 M \lambda$, $\lambda_2 = a_2 M \lambda$ — новые управляющие параметры системы; $0 \leq x \leq 1$.

Типичный график зависимости функции $y(x)$ приведён на рис. 1. Видно, что максимум функции $y_{\max} \approx 0,83$ наблюдается при $x = x_{\max} \approx 0,44$. При увеличении параметра $\lambda_1 \in (0, 4)$ в системе наблюдаются следующие закономерности. При $\lambda_1 < 1$ $y(x) \rightarrow 0$ (итерации I_n сходятся к $I^* \rightarrow 0,3$ — устойчивая неподвижная точка, см. рис. 2).

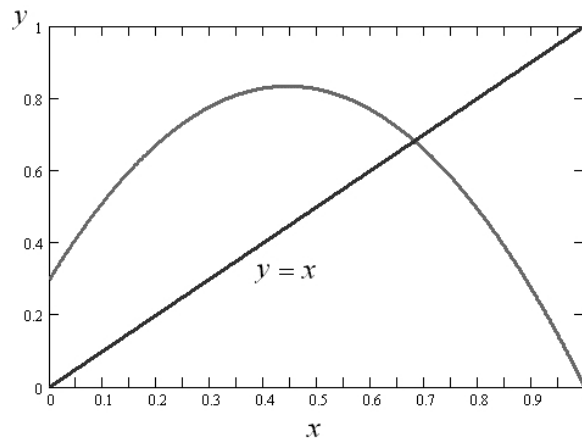


Рис. 1. График зависимости функции $y = \lambda_1 x(1-x) + \lambda_2 x(1-x)^2$; $\lambda_1 = 3$ и $\lambda_2 = 0,3$

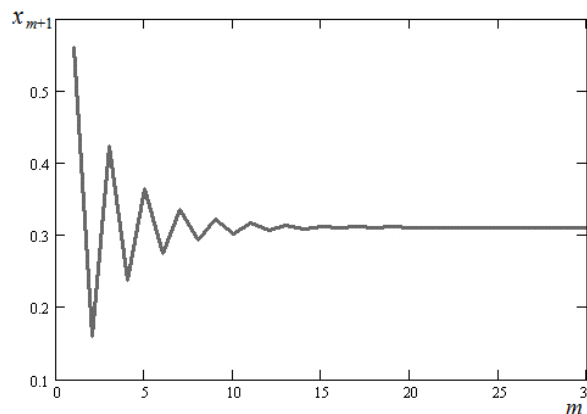


Рис. 2. График зависимости функции $x_{m+1} = \lambda_1 x_m(1-x_m) + \lambda_2 x_m(1-x_m)^2$ от m при $\lambda_1 = 0,21$, $\lambda_2 = 0,56$ и $p = 0$. Наблюдается цикл S^1

При дальнейшем росте управляющего параметра λ_1 происходит бифуркация: точка $I^* \rightarrow 0,3$ теряет устойчивость, а вновь появившаяся точка становится устойчивой (итерированные значения $I_m \rightarrow \text{const}$, т. е. динамический режим является стационарным или имеет период, равный единице — наблюдается цикл S^1).

Дальнейшее увеличение параметра λ_1 снова приводит к изменению системы: происходит бифуркация удвоения периода (неподвижная точка расщепляется и I_m осциллирует между двумя значениями, образующими устойчивый аттрактор с периодом 2). В этом случае отображение (2) имеет устойчивый цикл с периодом 2: S^2 (см. рис. 3). При дальнейшем увеличении параметра λ_1 возникают циклы вида S^{2^p} : $S^1, S^2, S^4, S^8, S^{16}, S^{32}, S^{64} \dots$, и т. д. (где $p = 0, 1, 2, 3, 4, \dots$). В результате можно наблюдать ветвящийся процесс: исходная ветвь разделяется на две, новые две снова делятся на две и т. д. Поэтому, применительно к процессу распространения лингвистической информации эту модель можно ещё назвать моделью «дерева» по аналогии с родословным деревом (в англоязычной литературе подобные модели называют «tree-model»). При таком подходе в качестве одной из количественных характеристик рассматриваемого процесса распространения «лингвистической информации»

можно рассматривать число возникающих циклов S^{2^p} в качестве числа возникших новых языков (или диалектов) в данном языковом сообществе; так циклы S^1 , S^2 , S^4 , S^8 , S^{16} , S^{32} дают соответственно: 1, 2, 4, 8, 16 и 32 языка. Тогда вполне естественно можно выбрать некоторый «внутренний» («встроенный») масштаб времени в данном процессе, а именно — воспользуемся известными в лингвистике данными: каждый цикл в динамике лингвистического сообщества соответствует на временной шкале величине 500 лет — время расхождения двух родственных языков (см, например, [10]). Таким образом, данная нелинейная модель, где есть ветвящиеся (древовидные) циклы типа S^{2^p} , позволяет нам использовать в качестве временного масштаба априорную лингвистическую информацию о времени расхождения языков. Например, между циклами S^2 и S^4 расстояние по времени составляет 500 лет, при этом каждый из двух языков предшественников в исходном цикле S^2 разошёлся за 500 лет на 2 новых, родственных исходным, языка (две новые лингвистические популяции), и в итоге их стало 4; и т. д. Отметим, что на начальных этапах развития языкового сообщества это, возможно, могут быть и языковые семьи или языковые ареалы типа «сатем-кентум».

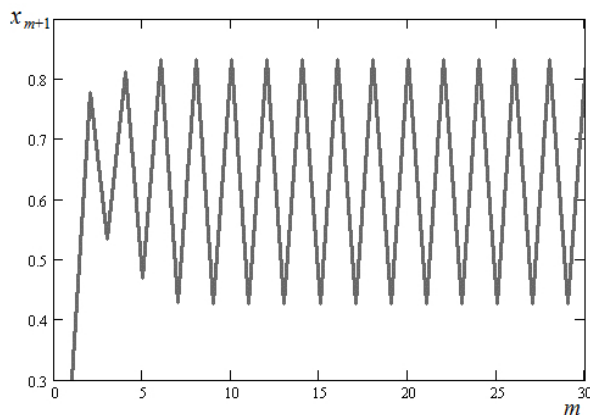


Рис. 3. График зависимости функции $x_{m+1} = \lambda_1 x_m(1 - x_m) + \lambda_2 x_m(1 - x_m)^2$ от m при $\lambda_1 = 3$, $\lambda_2 = 0,3$ и $p = 1$. Наблюдается цикл S^2

Устойчивый цикл S^4 показан на рис. 4. Можно также построить функцию $y = f(f(x))$, которая пересекается с прямой $y = x$ в точках, принадлежащих циклу S^4 (они и определяют данный цикл). Дальнейший рост управляющего параметра λ_1 позволяет наблюдать переход системы в хаотический режим (см. рис. 5): непериодический, случайный процесс возникает как предел всё более сложных структур (циклов вида S^{2^p}). Таким образом, хаос возникает как предел сверхсложной организации. Наконец, при некотором значении управляющего параметра $\lambda_1 \rightarrow \lambda_\infty$ отображение (2) даёт уже непериодическую последовательность $y(x)$ типа цикла x_{m+1} , показанного на рис. 5.

Между порядком и хаосом наблюдается глубокая внутренняя связь. Подобные закономерности существуют в любых системах, где наблюдаются последовательности бифуркаций удвоения периода [1–5]. Анализ зависимостей, аналогичных полученным выше, при широком варьировании параметров позволяет установить области изменения параметров системы и соотношение между ними, при которых происходит как переход от стационарного к периодическому, а затем хаотическому поведению системы, так и последующий возможный переход — от хаотического к стационарному или периодическому поведению.

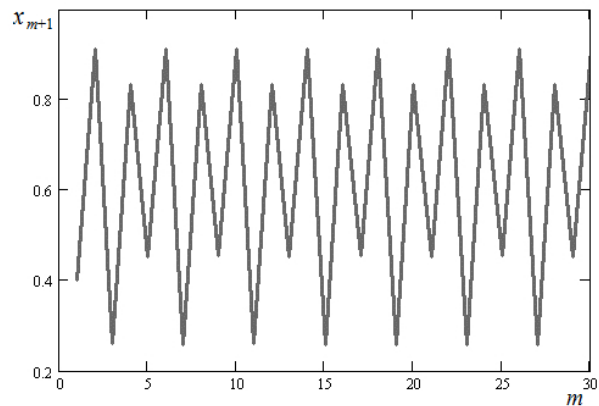


Рис. 4. График зависимости функции $x_{m+1} = \lambda_1 x_m(1 - x_m) + \lambda_2 x_m(1 - x_m)^2$ от m при $\lambda_1 = 3,2$, $\lambda_2 = 0,4$ и $p = 2$. Наблюдается цикл S^4

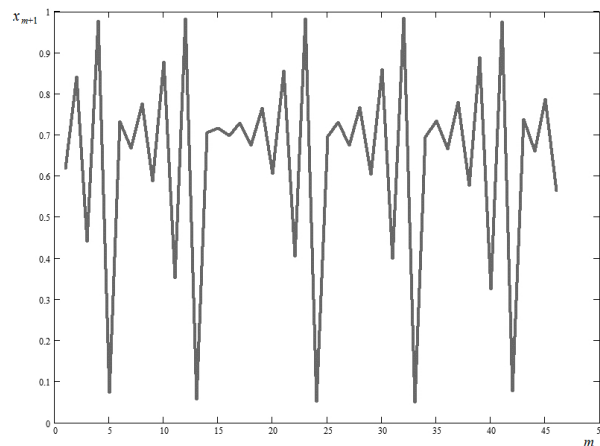


Рис. 5. График зависимости функции $x_{m+1} = \lambda_1 x_m(1 - x_m) + \lambda_2 x_m(1 - x_m)^2$ от m при $\lambda_1 = 3,2$, $\lambda_2 = 0,6$ и $1 \ll p < +\infty$. В системе наблюдается хаос

Используем полученные выше данные при численном исследовании распространения и изменения лингвистической информации в некотором модельном индоевропейском (ИЕ) языковом сообществе, в том числе на начальном этапе его формирования. 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. Результаты численных расчётов и их анализ. Выводы

Ниже на рис. 6, 7 приведены наиболее характерные из полученных нами результатов моделирования в рамках данной нелинейной модели. Видно, что на начальных этапах развития лингвистической системы в ней преобладает хаотическое состояние. При дальнейшем развитии лингвистического сообщества внутри областей хаоса

появляются окна периодического движения: узкие — на графиках слева и достаточно протяжённые — на графиках справа. Виден переход от хаотического «движения» динамической системы к периодическому (проявление определённых самоорганизующих свойств данной системы). Мы связываем предположительно именно этот переход с началом формирования собственно индоевропейских языков (когда Анатолийская ветвь уже отделилась), т. е. известных нам сейчас современных языков этой языковой семьи. Начальный же период, мы считаем, соответствует периоду совместного развития предков данных языков до их окончательного разделения (в пространстве и во времени). Как следует из графиков, приведённых слева, диапазон хаотического «движения» наблюдается от 0-го до 5–6-го «поколения» (итерации по m), т. е. примерно от 6500 до 2500–3000 лет назад. У графиков, приведённых справа, диапазон хаотического «движения» также наблюдается от 0-го до 5–6-го поколения, но в данном случае он составляет примерно отрезок от 9500 до 6500–7000 лет назад.

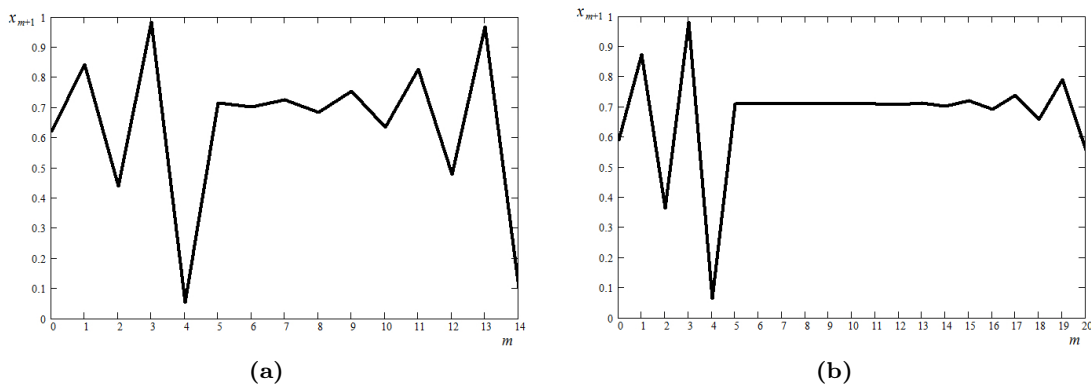


Рис. 6. График зависимости функции $x_{m+1} = \lambda_1 x_m(1 - x_m) + \lambda_2 x_m(1 - x_m)^2$ от m :
 (а) — при параметрах: $\lambda_1 = 3,2$ и $\lambda_2 = 0,62$, начальное значение $x_0 = 10^{-4}$;
 (б) — при параметрах: $\lambda_1 = 3,2$ и $\lambda_2 = 0,59$, начальное значение $x_0 = 10^{-4}$

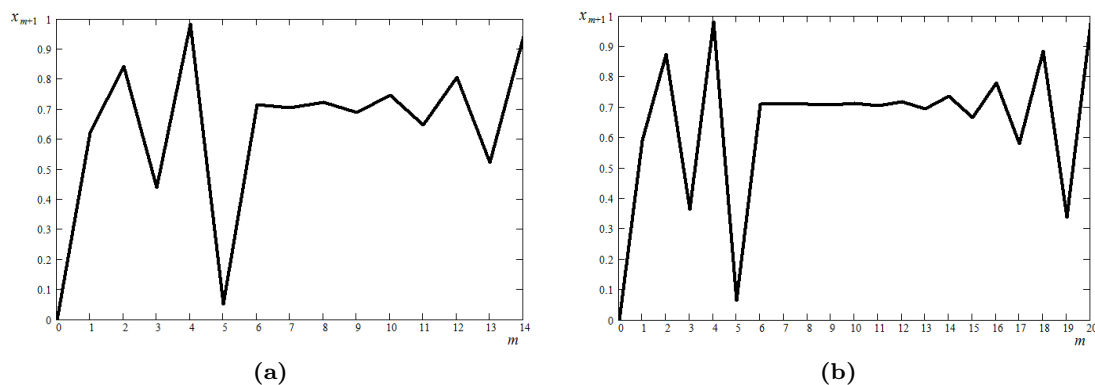


Рис. 7. График зависимости функции $x_{m+1} = \lambda_1 x_m(1 - x_m) + \lambda_2 x_m(1 - x_m)^2$ от m :
 (а) — при параметрах: $\lambda_1 = 3,2$ и $\lambda_2 = 0,62$, начальное значение $x_0 = 1$;
 (б) — при параметрах: $\lambda_1 = 3,2$ и $\lambda_2 = 0,59$, начальное значение $x_0 = 1$

Анализ приведённых на рис. 6, 7 зависимостей и сравнение полученных из них временных оценок с данными независимых исследований [5] позволяет сделать вывод, что графики слева больше соответствует Курганной гипотезе, а графики справа — Анатолийской. Здесь следует сделать одно важное замечание: количественное и качественное изменения рассматриваемой модели произошли при незначительном количественном изменении одного параметра, а именно: λ_2 изменилось с 0,62 до 0,59, т. е. всего в 1,05 раза, или примерно на 5%. Как следует из рассмотрения данной математической модели, всего 5% могут стать решающими в истории развития исследуемого лингвистического сообщества.

Теперь приведём данные, следующие из другой 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 сейчас примерно 300–500 языков. 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 выше предположение: использование данной нелинейной модели, где есть циклы вида S^{2^p} , позволяет нам 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, 2048 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 (как и для 8000 языков: $p = \log_2 8000 \approx 12,97$), 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. Для уточнения модели можно использовать следующий факт: мы предполагали выше, что в начальный период (от 0-го до 5–6-го поколения) был возможен некоторый период совместного развития предков данных ИЕ языков до их окончательного разделения (в пространстве и во времени), как внутри своей индоевропейской семьи, так и, возможно, с другими языковыми семьями. Вполне вероятно, что в этот период завершилось окончательное разделение ИЕ-семьи языков с Уральской языковой семьёй, в частности с финно-угорскими языками.

Как видно из графиков, полученные результаты компьютерного моделирования хорошо соответствуют по времени двум основным гипотезам о формировании протоиндоевропейцев: Анатолийской и Курганной [5, 14]. Напомним кратко об этих двух гипотезах. *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]. Этот факт был положен нами в основу при построении наших теоретических моделей. В частности, при компьютерном моделировании соотношение коэффициентов λ_1 и λ_2 выбиралось исходя из соотношения средних скоростей движения (в км/час) всадников v_1 и пешеходов v_2 в индоевропейских сообществах: $\lambda_1 : \lambda_2 \propto v_1 : v_2 \approx 15 : 3$. В курганной гипотезе полагается, что причерноморские степи были прародиной протоиндоевропейцев и по всему этому региону говорили на поздних диалектах праиндоевропейского языка.

Итак, из сравнения полученных нами данных с известными данными о возможном возрасте праиндоевропейского языка (в Анатолийской и Курганной гипотезах) следует, что они хорошо соответствуют друг другу. При этом сравнение полученных нами данных по обеим гипотезам с данными независимых исследователей (см., например, [5]) позволяет сделать однозначный вывод о предпочтительности Курганной гипотезы. Действительно, как видно из рис. 6, 7 (слева), в случае реализации сценария развития языков по Курганной гипотезе период существования 1–2 языков составляет примерно 1000–1500 лет (с примерно 4500 до 3500 лет назад), что вполне соответствует, например, данным [5]. Если считать, что полученные нами данные соответствуют формированию групп ИЕ языков (например, Славянская, Италийская, Кельтская, Германская, Индо-Арийская и др., то сходство наших результатов с результатами [5] будет более широким. В случае же Анатолийской гипотезы (см. рис. 6, 7 (справа)) период существования 1–2 ИЕ языков (или групп ИЕ языков) составляет примерно 4000–4500 лет (с примерно 7500 до 2000 лет назад), что не соответствует результатам большинства последних исследований (см., например, [5]). Таким образом, как видно из проведённого анализа полученных данных, более предпочтительной в данной модели является Курганная гипотеза, поскольку именно она полнее и точнее соответствует данным независимых исследований.

Приведём ещё результаты анализа некоторых данных, полученных по данной древоподобной модели. Например, как видно на графиках слева на рис. 6 и 7, величина «лингвистической информации» в первой (ветвящейся) модели после перехода от хаотического «движения» динамической системы к периодическому (предположительно это время формирования собственно индоевропейских языков — этап их окончательного расхождения или *дивергенции* (divergence); начальный хаотический период соответствует некоторому периоду совместного развития, определённой *конвергенции* (convergence) предков данных ИЕ языков до их окончательного разделения) достигает первый раз максимума на 13–14-м поколениях. В этой области возникают циклы, где p может превышать значения 16, 32, 64.

Подчеркнём, что в дальнейшем планируется уточнение сделанных при исследовании предположений, а также выбранных параметров, в том числе с учётом новых лингвистических и генетических данных.

Теперь 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. Для этого используем данные численных расчётов возможного количества информации примерно на 5–7-й итерациях по m у графиков, приведённых слева, после хаотического «движения», т. е. примерно 4000 ± 500 лет назад.

Для оценки используем хорошо известную в теории информации формулу для полной информации I_Σ , содержащейся в некотором сообщении [21]:

$$I_\Sigma = n \log_2 \mathfrak{S}, \quad (3)$$

где n — общее число символов в сообщении (непрерывный сигнал заменён дискретной последовательностью отсчётов); \mathfrak{S} — число различных символов алфавита.

Для одного из полученных в численных расчётах по первой модели значения $I_{\Sigma} = 523$ бита (двоичные единицы), при некотором фиксированном заданном начальном значении информации в сообществе, получаем ряд значений для $\mathfrak{S} = 2^{I_{\Sigma}/n}$ при вариации параметра¹ n : $n = 80$, $\mathfrak{S}_{80} = 93$ символа; $n = 100$, $\mathfrak{S}_{100} = 38$ символов; $n = 120$, $\mathfrak{S}_{120} = 21$ символов; $n = 150$, $\mathfrak{S}_{150} = 11$ символов; $n = 200$, $\mathfrak{S}_{200} = 6$ символов. Скорее всего, разумная оценка \mathfrak{S}_n должна быть ближе к диапазону: $\mathfrak{S}_n = 11 \div 38$ символов. На следующей итерации значения I_{Σ} возрастают, и соответственно возрастают значения \mathfrak{S}_n .

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. Отличительные особенности: тип письма — консонантно-вокалическое; направление письма — слева направо. 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 (деванагари: *samskr̥tā vās*, «литературный язык») — 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 фонем². Если учитывать аллофоны³ (а система письма их учитывает), то

¹Для простоты под этим числом может пониматься некоторое число условных «символов», соответствующих числу знаков в некотором списке, например, в 100-словном списке Сводеша. «Символами» в доисторическую эпоху могли быть, например, как фонемы, так и целые предложения.

²Фонема (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.

³Аллофон (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.

общее количество звуков в языке увеличивается до 48. С учётом данных по древнему санскриту (примерно 3700 лет назад) из книги [22] можно оценить минимальное возможное число «знаков» в алфавите пра-санскрита величиной: $\mathfrak{S} = 9$.

Как видно из приведённых данных для этих языков, они в основном удовлетворяют заданному диапазону: $\mathfrak{S}_n = 11 \div 38$ символов.

Полученные результаты позволяют сделать вывод, что предложенная нами математическая нелинейная модель распространения и изменения лингвистической информации в модельном ИЕ сообществе достаточно адекватно описывает исследуемое явление. Вместе с тем проведённое исследование показало, что необходима дальнейшая тонкая настройка модели, 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 пра-индоевропейцев: Анатолийской и Курганной. При этом совокупность данных, полученных при исследовании, 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.

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. *David W. A.* 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.
12. Старостин Г. С. и др. К истокам языкового разнообразия. — М.: Издательский дом «Дело» РАНХиГС, 2015.
13. Russell D. G., Quentin D. A. 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. Mitnik 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. Гоноровский И. С. Радиотехнические цепи и сигналы. Ч. 1. — М.: Радио, 1966.
22. Borsboom W. Alphabet or Abracadabra? Reverse Engineering the Western Alphabet. — Victoria, BC, Canada: FriesenPress, 2015.

UDC 517+519.2:80

Dynamic Non-Linear Model of Distribution and Changes of Linguistic Information in the Indo-European Model Language Community

A. A. Egorov*, M. A. Egorova†

* *General Physics Institute, Russian Academy of Sciences, Moscow, Russia*

† *RUDN University (Peoples' Friendship University of Russia), Moscow, Russia*

The paper considers the nonlinear dynamic mathematical model describing the distribution and variation of linguistic information in the Indo-European linguistic community. When constructing a mathematical model of linguistic information propagation and changes in the linguistic community as a priori information data from independent studies both linguistics and other scientific fields, such as history, archeology and genetics were used. Within the framework of this model the spread of linguistic information in a model Indo-European language community, including at the initial stage of its formation was numerically studied. The preliminary results of theoretical analysis and computer simulation are given. It was found that the mathematical model of the distribution and modification of linguistic information shows both regular and typical chaotic behavior. As one of quantitative characteristics of considered nonlinear process of distribution of the linguistic information it is offered to consider number of arising cycles as number of the arisen modern languages, in the given language community. Results of computer modeling show, that from two main hypotheses of formation of the Proto-Indo-Europeans – Anatolian and Kurgan, the latter better matches temporary estimates obtained by us.

Key words and phrases: mathematical model, nonlinear equation, mathematical linguistics, Indo-European languages, Indo-European chronology, Indo-European dispersal, Anatolian and Kurgan hypotheses

References

1. A. A. Egorov, Some Patterns of Dissemination of Information in Society, in: 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: 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. 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. W. A. David, *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
12. G. S. Starostin, et al., *To the Source of Linguistic Diversity*, Publishing House “Delo” RANHiGS, Moscow, 2015, in Russian.
13. D. G. Russell, D. A. Quentin, 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. I. S. Gonorovsky, *Radio Circuits and Signals. Part 1, Radio*, Moscow, 1966, in Russian.
22. W. Borsboom, *Alphabet or Abracadabra? Reverse Engineering the Western Alphabet*, FriesenPress, Victoria, BC, Canada, 2015.