

УДК: 519.765

Диссипативная стохастическая динамическая модель развития языковых знаков

В. В. Поддубный^{1,a}, А. А. Поликарпов^{2,b}

¹Томский государственный университет,
Россия, 634050, г. Томск, пр. Ленина, 36, ТГУ

²Московский государственный университет им. М. В. Ломоносова,
Россия, 119991, ГСП-1, г. Москва, Ленинские горы, МГУ

E-mail: ^a vvpoddubny@gmail.com, ^b polikarp@philol.msu.ru

Получено 17 марта 2011 г.

Предлагается диссипативная стохастическая динамическая модель эволюции языковых знаков, удовлетворяющая принципу «наименьшего действия» — одному из фундаментальных вариационных принципов природы. Модель предполагает пуассоновский характер потока рождения языковых знаков, экспоненциальное (показательное) распределение ассоциативно-семантического потенциала (АСП) знака и оперирует разностными стохастическими уравнениями специального вида для диссипативных процессов. Получаемые из модели распределения полисемии и частотно-ранговые распределения языковых знаков статистически значимо (по критерию Колмогорова–Смирнова) не отличаются от эмпирических распределений, полученных из представительных толковых и частотных словарей русского и английского языков.

Ключевые слова: языковой знак, эволюция, ассоциативно-семантический потенциал, значение знака, полисемия, частотно-ранговое распределение, диссипативная стохастическая динамическая модель

Dissipative Stochastic Dynamic Model of Language Signs Evolution

V. V. Poddubny¹, A. A. Polikarpov²

¹Tomsk State University, 36 Lenin av., Tomsk, 634050, Russia

²Lomonosov Moscow State University, Leninskie Gory, Moscow, 119991, Russia

Abstract. — We offer the dissipative stochastic dynamic model of the language sign evolution, satisfying to the principle of the least action, one of fundamental variational principles of the Nature. The model conjectures the Poisson nature of the birth flow of language signs and the exponential distribution of their associative-semantic potential (ASP). The model works with stochastic difference equations of the special type for dissipative processes. The equation for momentary polysemy distribution and frequency-rank distribution drawn from our model do not differs significantly (by Kolmogorov-Smirnov's test) from empirical distributions, got from main Russian and English explanatory dictionaries as well as frequency dictionaries of them.

Keywords: language sign, evolution, associative semantic potential, sign meanings, polysemy, frequency rank distribution, dissipative stochastic dynamic model

Citation: *Computer Research and Modeling*, 2011, vol. 3, no. 2, pp. 103–124 (Russian).

Введение

Проблема возникновения и эволюции языка как средства человеческого общения издавна привлекает внимание многих исследователей. Этой проблеме посвящено огромное количество публикаций. Однако только малая их часть связана с привлечением математического аппарата, математических методов и математических моделей для исследования этой проблемы. Объясняется это чрезвычайной сложностью объекта исследования — живого человеческого языка.

Наиболее продвинутыми в части применения математических методов к исследованию проблем эволюции естественного языка являются, на наш взгляд, работы [Nowak, Krakauer, 1999; Plotkin, Nowak, 2000; Nowak, Komarova, Niyogi, 2002; Ahlswede и др., 2004; Vogt, 2009] и др., развивающие динамические вероятностно-информационные подходы к построению математических моделей процессов языковой коммуникации и эволюции языка. Эти подходы используют дифференциальные и разностные уравнения, а также модели детерминированного хаоса (например, [Mitchener, Nowak, 2004]) для описания вероятностных и динамических характеристик процессов эволюции.

Однако в мировой научной литературе практически отсутствуют работы по математическому моделированию развития *полисемии* языковых знаков. По-видимому, впервые проблема математического описания развития полисемии знака, его жизненного цикла как унимодальной кривой, определяющей зависимость числа значений знака от времени, была поставлена А. А. Поликарповым в 1993 г. [Polikarpov, 1993]. Первоначальная идея того, что каждый акт языковой коммуникации является одновременно и тактом развития языка (точнее, употребленного в данном акте знака: расширение смысловой области употребленного значения и его соответствующая абстрактивизация, деспецификация, субъективизация и грамматикализация), сформулирована А. А. Поликарповым еще в 1988 г. в докладе [Поликарпов, 1988], а затем в статье [Поликарпов, 1990]. Развитие этой идеи получило отражение в докладе [Polikarpov, 1991], в статье [Polikarpov, 1993], а затем в статьях [Поликарпов, 1994; Поликарпов, 1995]. Первый вариант собственно математической модели развития жизненного цикла языкового знака на основе использования математической теории ветвящихся процессов был представлен в работе [Khmelev, Polikarpov, 2000]. Большую роль в становлении представлений авторов о процессах эволюции языковых знаков сыграли важные работы Ю. К. Крылова по статистическому анализу полисемии [Крылов, Якубовская, 1977] и по марковской модели эволюции лексической многозначности (полисемии) [Krylov, 1995].

В настоящей работе мы развиваем динамический подход к построению математической модели эволюции полисемии языковых знаков, основанный на гипотезе о диссипативном характере этой эволюции и приводящий к дифференциальным уравнениям, удовлетворяющим **«принципу наименьшего действия»** («the least action principle» — LAP) в форме Лагранжа для диссипативных процессов. Принцип наименьшего действия, открытый Пьером Мопертюи в 1744 году, является одним из фундаментальных принципов, выражающих законы Природы. Он определяет систему дифференциальных уравнений, описывающих закон изменения состояния системы (процесса), из вариационного принципа минимума так называемого функционала действия [Айзерман, 1980; Маркеев, 1999]:

$$S = \int_0^{\infty} L(x(t), \dot{x}(t)) dt, \quad x(t) = \arg \min_{x(t)} S,$$

где $x(t)$ — вектор обобщенных координат, описывающий состояние системы (процесса) в момент времени t , $\dot{x}(t)$ — вектор обобщенных скоростей (производных от обобщенных координат

по времени), $L(x(t), \dot{x}(t))$ — функция Лагранжа (функция координат $x(t)$ и скоростей $\dot{x}(t)$). В частности, для диссипативных систем (процессов) [Лоскутов, Михайлов, 2007] эта функция имеет смысл разности между кинетической $T(\dot{x}(t))$ и потенциальной $U(x(t))$ энергиями на траектории системы (процесса), так что ее минимум, равный нулю, обеспечивает полное рассеяние (диссипацию) энергии. Из этого принципа (при различных функциях Лагранжа) вытекают многие законы физики (оптики, классической и квантовой механики, механики сплошных сред, классической и квантовой теории поля и др.). В настоящей работе мы показываем, что этот принцип применим также и в лингвистике для описания процессов эволюции языковых знаков.

Другая группа работ (например, [Ferrer i Cancho, Sole, 2003; Li, 1992; Крылов, 1996; Крылов, Кудрин, 1999; Орлов, 1980; Орлов, 2001] и др.), также использующая вероятностно-информационные, но не динамические подходы, направлена на изучение частотно-ранговых распределений различных статистических ансамблей, реализующих «**принцип наименьшего усилия**» («the least effort principle» — LEP). В этих работах исследуются границы применимости для различных вероятностных моделей широко известных в психологии, биологии, социологии, экономике, лингвистике эмпирических законов Ципфа [Zipf, 1949], Ципфа–Мандельброта [Mandelbrot, 1965; Мандельброт, 1973], Ципфа–Парето («закон 80/20»).

Обратим внимание, что «принцип наименьшего действия» (LAP) имеет строгую математическую формулировку (минимум функционала от функции Лагранжа), является математическим вариационным принципом, основанным на вариационном исчислении, в то время как «принцип наименьшего усилия» (LEP) не имеет строгой математической формулировки, является скорее философско-психологическим, чем математическим принципом. Понятие *усилия* в нем строго не определено, достаточно расплывчато, трудно формализуемо, является скорее качественным, чем количественным понятием. Кроме того, в чистом виде этот принцип, по-видимому, нигде в природе не реализуется [Li, 1992; Василенко, 2004]. Как будет показано ниже, результаты моделирования эволюции знаковых ансамблей на основе использования принципа наименьшего действия (LAP) дают хорошее совпадение распределений полисемии и частотно-ранговых распределений с данными толковых и частотных словарей без использования принципа наименьших усилий (LEP).

Первоначальный вариант предлагаемой диссипативной стохастической динамической модели развития языковых знаков, позволяющей моделировать их эволюцию и получать распределение полисемии, длительности жизни и возраста знаковых ансамблей с эмпирической проверкой ее адекватности только по полисемии, опубликован в трудах трех конференций [Поддубный, Поликарпов, 2010; Поддубный, Поликарпов, 2010а; Поддубный, Поликарпов, 2011]. В настоящей работе излагается наиболее полный вариант модели, позволяющей моделировать не только процессы развития полисемии знаков, строить распределения полисемии знаковых ансамблей и проверять их адекватность на материале представительных словарей русского и английского языков, но и моделировать частотно-ранговые распределения знаковых ансамблей с эмпирической проверкой их адекватности по данным частотных словарей русского и английского языков, а также распределения длительностей жизни языковых знаков и их возрастов с эмпирической проверкой последних по данным возрастно-частотно-хронологических и толковых словарей русского и английского языков.

Феноменологическое описание процессов эволюции знаков и знаковых ансамблей естественного языка

Основные определения и качественные допущения

Предлагаемый подход к математическому моделированию эволюции лексической системы базируется на тех онтологических представлениях, которые составляют *лингвистическую модель*

жизненного цикла знака [Поликарпов, 1998; Поликарпов, 2007; Поликарпов, 2007а; Поликарпов, 2009а]. Основные положения этой модели заключаются в следующем.

(1) *Мышление* любого индивидуума — это операции над образами. Образы являются отражением внешнего мира в сознании индивида и с теоретико-множественной точки зрения характеризуются множествами (наборами) определенных признаков (возможно, нечеткими множествами).

(2) Мир образов (или, иначе, смыслов) в сознании каждого индивида — это иерархический набор, некая *пирамида родо-видовых абстракций* с широким основанием, состоящим из самых конкретных первичных образов-смыслов, с постепенным сужением количества разных образов на каждом последующем, все более высоком уровне их обобщения (абстракции).

(3) Пара «обобщенный образ ситуации — обобщенный образ акта поведения в ней» есть *элементарная кодовая единица поведения*. Прижизненно нарабатываемый набор таких кодовых единиц — *жизненный опыт индивидуума*. Жизненный опыт позволяет индивиду выживать в тех жизненных ситуациях, в которые он попадает.

(4) Кодовые единицы поведения *много-многозначны*. Это означает, во-первых, что одному образу ситуации жизнедеятельности может соответствовать несколько относительно равноценных, *изофункциональных* образов полезного поведения в ней. Во-вторых, один и тот же *полифункциональный* образ полезного поведения с равной эффективностью может быть применен в нескольких различных типах ситуаций жизнедеятельности.

(5) Особый тип ситуаций под названием «*коммуникативные ситуации жизнедеятельности*» характеризуются тем, что целями жизнедеятельности индивидов в них является не получение каких-либо жизненных благ (с целью повышения своих шансов на выживание), а «обмен» образами-смыслами, *обмен индивидов жизненным опытом, который далее может использоваться в практических ситуациях*. Коммуникативные ситуации являются вспомогательными для практических ситуаций жизнедеятельности, зависят от них, отображают их основные информационные характеристики. Возможность обмена смыслами достигается за счет выработки в каждом коллективе некоторого относительно *единообразного коммуникативного кода поведения*, состоящего из значений — относительно обобщенных образов ситуаций жизнедеятельности (берущихся, как правило, с какого-нибудь достаточно обобщенного уровня индивидуальных пирамид абстракций) и образов коммуникативного поведения в них (образов «артикуляции» для создания акустических, оптических, электронных и иных внешних посредников между артикулирующими и воспринимающими органами индивидуумов).

Каждая пара «значение (или несколько значений) — образ артикуляции» — это *элементарная коммуникативная кодовая единица*, или элементарная знаковая единица, *знак*.

Каждая такая пара может быть как *многозначной* (то есть полифункциональной), так и *синонимичной* (то есть изофункциональной) со значениями каких-либо еще других знаков.

Системный набор знаков — *ЯЗЫК*. Язык — коммуникативная часть мыслительного механизма индивидуума, коммуникативная часть его жизненного опыта.

(6) Процесс коммуникации между индивидуумами базируется на *принципе намекания*. Каждый знак может быть *многозначным*, вырабатывая в своей истории или получая в наследство от производящего слова определенный набор значений. Каждое значение каждого знака является в определенной степени *многозначным*, покрывает некоторую область смыслов, каждый из которых, как правило, является более конкретным, чем значение, но имеет какую-то часть общих с ним (повторяющихся) признаков. На основе повторяемости части признаков и возможно эвристическое ассоциативное возбуждение (*выбор*) соответствующего значения знака при возбуждении одного из смыслов данной смысловой области, предназначенного к «передаче» (у отправителя сообщений) и, обратно, эвристическое возбуждение (*выбор*) сначала соответствующего значения, а потом и возбуждение (*выбор*) соответствующего смысла после правильного

выбора значения полученного знака при «приеме» (у получателя сообщений). Успешность этих *намекательных операций и операций догадки* базируется на *общности апперцепционной базы общающихся*, что *создает определенный уровень избыточности их общей смысловой системы и ведет к возможности успешного употребления ими намека и догадки в ходе общения.*

(7) Формируемый в каждом национальном языке набор значений знаков является производным от набора смыслов, является его *редуцированным отображением*. Редукция заключается одновременно и в том, что *значений существенно меньше, чем смыслов*, и в том, что *каждое значение — более схематичное, более абстрактное отображение содержания смыслов, подпадающих под него.*

(8) Каждый акт общения одновременно и акт эволюции языка (точнее, микроэволюции), так как существует определенная вероятность выхода очередного значения-намека за прежде сформированные границы его смысловой области и закрепления расширившегося объема этой области. Это определяет *неизбежность постоянного со временем разрастания смысловой области каждого значения, увеличения общего числа разных смыслов в ней и повышения общего разнообразия признаков в этих смыслах.*

(9) *Повышение разнообразия смыслов и разнообразия признаков в смыслах данного значения* знака сужает область пересечения в наборах признаков смыслов, что со временем ведет к *усилению определенного более узкого круга наименее специфических признаков в значении и к ослаблению остальных, более специфических, менее распространенных, менее представленных в разных смыслах признаков значения.* В конечном счете это приводит к выпадению из значения части совсем ослабевших, наиболее специфических признаков, формируя *тенденцию к абстрактивизации во времени любого значения. И чем дольше значение существует, тем оно оказывается более абстрактным, наполненным все менее специфическими признаками.*

Наиболее специфические признаки значения в наибольшей степени перекликаются с теми признаками его смыслов, которые характеризуют особую предметную, денотативную область смыслов. Эти признаки значений и смыслов могут быть названы *денотативными*. Наименее специфические признаки, встречаясь в составе самых различных смыслов данного значения, выполняют в них в большей степени *характеризующую роль, роль аспектуальных качественных определителей.* Эти признаки значений и смыслов могут быть названы *качественными*.

(10) В ходе процесса абстрактивизации значений могут не только теряться какие-то из их исходных признаков, но (хотя и в существенно меньшей степени) приобретаться новые. Однако качество приобретаемых признаков совершенно иное, чем теряемых. Дело в том, что абстрактивизация значений приводит к расширению круга коммуникативных контекстов их употребления и, следовательно, к ассоциативному воздействию на набор признаков каждого значения наиболее повторяющихся смысловых элементов из контекстов употребления значения. Наиболее же повторяющимися элементами смысловых контекстов общения между людьми являются их субъективные оценки по отношению друг к другу, другим людям, предметам и явлениям. Это и приводит со временем к заметному усилению в значениях слов оценочных, модальных и иных их субъективных признаков. То есть со временем значения должны претерпевать заметную субъективизацию.

(11) *Появление нового значения — частный случай расширения (скачком) смысловой области материнского значения.* Дело в том, что часть новых добавляемых смыслов (в результате указанного выше процесса неизбежного расширения смысловой области каждого значения) может походить на другие смыслы столь непрочно, поверхностно, что рядом с прежней смысловой областью может выделяться как особая новая смысловая область, которая хотя и ассоциативно связана с прежней, но может обозначать особый денотат.

(12) В качестве нового значения каждую такую новую смысловую область начинает представлять один из наиболее обобщенных образов среди добавившихся новых смыслов. По-

этому *каждое новое значение в тенденции является относительно более абстрактным, чем материнское значение.*

(13) Если это так, то в среднем *пропорционально увеличению порядкового номера появляющихся новых значений должна нарастать степень абстрактности* (уменьшаться степень наполненности числом признаков) *значения.*

(14) Чем значение более абстрактно, тем оно менее активно в порождении следующих значений. Это происходит из-за того, что у *более абстрактного значения меньше специфических признаков*, которыми можно было бы ему «зацепиться» за какой-то новый смысл, который может стать новым значением. Вследствие этого *процесс прибавки новых значений в истории знака постепенно затухает.*

(15) Чем значение относительно абстрактнее, тем относительно большая *стабильность, длительность жизни для него свойственна* (в основном за счет более широкой смысловой области приложения так изменяющегося значения). Вследствие этого в истории знака *последовательность выпадения значений существенно коррелирует с последовательностью их появления*, так как исходные, самые конкретные значения являются наиболее неустойчивыми, наиболее зависящими от изменений в мире смыслов и от конкурентных притязаний значений других знаков на обозначение некоторой смысловой области.

(16) Наложение друг на друга двух процессов — приобретения новых и потери ранее приобретенных значений — определяет характер кривой развития (актуальной) полисемии знака во времени. В случае развития полисемии знака от исходной однозначности должна наблюдаться *асимметричная кривая со сдвигом пика развития полисемии к началу процесса.*

(17) На основе абстрактивизации значений, последовательно появляющихся у знака (расширения покрываемой ими смысловой области), пропорционально этому *растет степень употребительности знака в каждом из последующих значений*, что определяет характер кривой развития общей частоты знака в ходе его жизненного цикла, аналогичный характеру кривой развития полисемии, но с относительным сдвигом пика развития частоты в более поздний период жизненного цикла знака.

(18) *Сообразно абстрактивизации, окачествлению, деспецификации и субъективизации значений знаков в ходе их старения повышается объективная склонность значений к синонимизации со значениями других знаков.* Это связано с тем, что увеличение степени неспецифичности значений разных знаков определяет все большую степень их совпадения почти всеми признаками и возможность синонимического противопоставления субъективно-оценочными признаками (удельный вес которых в значениях, наоборот, нарастает), в чем, собственно, и заключается природа синонимии.

(19) Процесс появления у знака новых значений протекает на основе реализации того потенциала развития, который будем называть *ассоциативно-семантическим потенциалом знака* (АСП). Ассоциативно-семантический потенциал заложен в исходном значении (исходных значениях) знака. Вся история знака — это история перераспределения этого потенциала и его растрачивание.

Перераспределение потенциала осуществляется за счет отпочкования от каждого значения новых значений. При этом наблюдается занятие свободных ассоциативных валентностей материнского значения, что постепенно приводит к затуханию способности каждого значения к рождению от него новых значений. Кроме того, новые значения (то есть значения каждой очередной ступени деривации), будучи более абстрактными, менее способны к «деторождению», чем их родители. Таким образом, процесс перераспределения ассоциативно-семантического потенциала, рождения новых значений у данного знака на каком-то этапе замирает полностью.

Растрата потенциала осуществляется как в ходе собственной абстрактивизации каждого значения в его истории, так и в ходе неизбежных потерь значений. Потери значений могут

происходить как из-за выпадения из жизни смысла, им обозначавшегося, так и из-за возможного перехвата функции обозначения этого смысла значением другого знака. Этот второй случай возможен в силу постоянной конкуренции знаков как наиболее подходящих намекаателей на тот или иной смысл. Конкуренентоспособность старых, достаточно абстрактивизировавшихся значений, в сравнении с более конкретными значениями более молодых знаков, явно ниже, так как они становятся все менее определенными (менее эффективными) намекаателями на тот или иной смысл. Кроме того, процессы эвфемизации и дисфемизации тоже постоянно заставляют носителей языка менять знаковые обозначения для эмфатических смыслов [Поликарпов, Селезнева, 2004; Поликарпов, 2009].

(20) Разумно предположить, что знак (в случае начала знаком своей жизни с однозначности) характеризуется тремя параметрами — АСП, активностью и стабильностью. Параметр АСП ответственен за общее число значений, которое может быть порождено в течение жизни знака. Параметр активности ответственен за ту скорость, с которой значения порождаются. Наконец, параметр стабильности ответственен за степень долгожительства тех значений, которые появляются у знака.

(21) Более сложная картина семантической динамики наблюдается в случае начала жизненного цикла производными словами от некоторого унаследованного от материнского слова набора значений. Сам процесс образования новых слов — логическое продолжение уже указанных тенденций к абстрактивизации каждого значения знака и к появлению у него в его истории новых, относительно более абстрактных значений. Все это в совокупности ведет к *общей тенденции перехода меняющегося знака от исходного обозначения предметов, веществ и материалов к обозначению тех свойств и качеств, которые присущи этим и, возможно, другим предметам, веществам и материалам, а также к обозначению процессов, в которые эти объекты и их свойства могут быть вовлечены*. Растущая в таких случаях частота признакового употребления значений слова, видимо, заставляет говорящих подсознательно подыскивать для такого слова и более подходящую категориальную признаковую форму, а именно — *заменять данное слово производным от него более признаковым словом*. Самые явные примеры этого рода — словообразование относительных прилагательных от существительных, наречий — от прилагательных и т. п. То есть сутью словообразовательного процесса во многих случаях является не только и не столько получение новых знаковых оболочек с совершенно новыми значениями для обозначения новых смыслов, сколько переобозначение (переименование) уже обозначаемых в языке смыслов, но с добавлением в форме (звучании и написании) и значениях новых производных слов (в сравнении с производящими) либо тех или иных дополнительных категориально-грамматических компонентов (дерево — деревянный: *синтаксическая деривация, транспозиция*), либо дополнительных категориально-лексических (дерево — деревцо: *модификация*), либо дополнительных чисто лексических компонентов, наряду с грамматическими (забегать — забегаловка: *мутация*). Случаи синтаксической деривации и близкой ей модификации составляют среди всех актов словообразования, как по данным [Земская, 1992], так и по нашим данным [Поликарпов, Кукушкина, Токтонов, 2009], до 70% всех случаев. Мутационная разновидность лексической деривации (обозначение новым производным словом чего-то совсем нового в сравнении с любым значением его производящего) составляет оставшуюся, существенно меньшую долю среди всех лексических новообразований.

Итак, словообразовательное переоформление признаковыми номинациями (например, прилагательными, глаголами) плана выражения слов, исходно бывших «оболочками» значений предметных номинаций (существительных), а также словообразовательная замена некоторых знаменательных признаковых номинаций служебно-признаковыми номинациями (служебными словами) становится для каждого из слов, в среднем, все более вероятным в связи с их общим старением, общей абстрактивизацией, депредметизацией и окачествлением, а также субъективизацией их значений в их истории.

Переоформление слова в таком случае заключается в создании нового производного слова на основе прежнего (производящего) с добавлением к основе слова того или иного суффиксального классифицирующего средства. Тем самым, можно констатировать, что в большинстве случаев при образовании нового слова наблюдается его относительное *удлинение* в сравнении с производящим словом.

Следует добавить, что, как правило, суффиксально-производное слово должно наследовать от производящего не все, а только часть значений, в наибольшей степени те, которые относительно более абстрактны, более признаковы среди его прочих. То есть производное слово, как правило, должно начинать свою семантическую эволюцию с меньшей величины полисемии и с более абстрактного состояния своих значений, чем его материнское (производящее) слово. Поэтому производное слово с самого начала должно иметь более высокий потенциал к образованию от него в будущем еще менее предметных по категориальному и семантическому качеству слов, чем его деривационный предшественник.

Чем на более отдаленной ступени деривации от начала словообразовательной цепи находится данное слово, тем, соответственно, оно длиннее, более признаково и имеет относительно меньшую исходную полисемию.

При необходимости подбора знакового обозначения для совсем новых смыслов, ранее не обозначавшихся в данном коллективе, носители любого языка чаще должны пользоваться такими способами как образование нового словосочетания, образование нового сложного слова, а также заимствование знака из другого языка, а не образование аффиксального производного.

В случае с заимствованиями мы имеем дело с новым для данного языка смыслом, но уже получившим в другом языке однословное наименование, которое в готовом виде с полным или неполным набором значений используется и в заимствующем языке.

Поскольку слова и словосочетания совместно покрывают весь набор смыслов данного общества, постольку они и составляют единый системный набор знаков, единый системный ансамбль номинативных средств языка. Этот принцип совместного функционирования знаковых единиц разных уровней организации языка необходимо учитывать при попытках моделирования распределений (в разных отношениях) этих единиц. Они существуют не по отдельности, а только совместно. При этом каждый из этих типов знаков выполняет свою функцию — слова покрывают более ядерную, наиболее активную часть смыслов общества, словосочетания — более периферийную.

Резюме

Итак, согласно современным представлениям о реализации жизненного цикла языкового знака, любой языковой знак (в том числе и слово) после своего появления в языке в некотором начальном значении может либо сохранять это значение в течение всей своей жизни, либо претерпевать эволюцию, последовательно рождая новые значения, все более и более абстрактные по смыслу, пока не будет полностью израсходован его так называемый ассоциативно-семантический потенциал (АСП) — способность порождать новые значения (своя для каждого знака). Если знак появляется в языке сразу с некоторой многозначностью (модифицированный знак, унаследовавший ряд значений от ранее использовавшегося знака, или знак, заимствованный из другого языка с соответствующей многозначностью), происходит аналогичный процесс эволюции: либо знак не порождает больше новых значений (когда его начальная многозначность равна АСП), либо продолжает последовательно порождать новые, все более абстрактные значения (когда его начальная многозначность меньше, чем АСП). При этом скорость рождения новых значений на начальном участке процесса эволюции знака максимальна (хотя и своя для каждого знака), а затем постепенно падает до нуля. Через некоторое время τ_0 (свое для каждого знака) начинается

аналогичный процесс выпадения из употребления сначала наименее абстрактных (наиболее конкретных) значений знака, а затем все более абстрактных, пока все значения знака не выйдут из употребления. При этом начальный участок процесса выпадения из употребления значений знака характеризуется наибольшей скоростью, которая затем постепенно уменьшается до нуля. Весь процесс выпадения значений знака из употребления идет более медленно, чем процесс рождения новых значений. В результате в каждый данный момент текущего времени после момента возникновения знака число актуальных (живущих) значений знака (его полисемия) сначала растет, достигает максимума, а затем постепенно падает. Через некоторое время (время жизни знака) это число становится равным нулю — знак исчезает из употребления. Кривая развития этого процесса во времени (кривая жизненного цикла языкового знака) — унимодальная кривая с максимумом, смещенным влево, к началу процесса (как схематично показано на рис. 1).

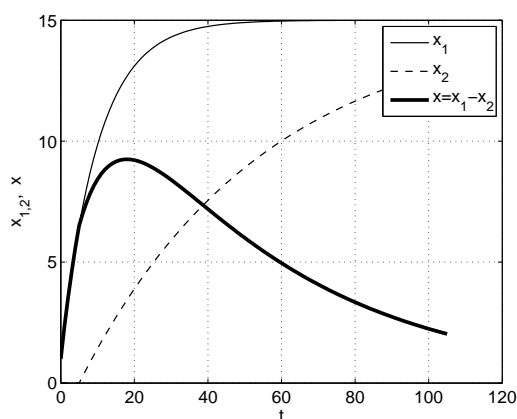


Рис. 1. Кривая жизненного цикла языкового знака

Кривая $x_1(t)$ — процесс роста числа значений знака, $x_2(t)$ — процесс выпадения значений знака из употребления, $x(t) = x_1(t) - x_2(t)$ — процесс развития активной полисемии знака, t — условное время. Максимальное значение $x_1(t)$ определяет АСП знака, скорость роста кривой $x_1(t)$ — активность знака, отставание роста кривой $x_2(t)$ от $x_1(t)$ — стабильность знака. Параметры кривой развития полисемии $x(t)$ (АСП, активность и стабильность знака) меняются от знака к знаку в знаковом ансамбле языка случайным образом, но в соответствии с определенными законами распределения.

Далее, каждое значение знака в соответствии с принципом намекания представляет («намекает на») некоторую область смыслов и может быть охарактеризовано некоторым смысловым объемом, монотонно возрастающим с ростом уровня абстрактности значения. Чем больше этот объем, тем чаще употребляется знак в этом значении. Доля суммарного смыслового объема всех значений знака, живущих в некоторый данный момент времени и составляющих активную полисемию знака, в общем смысловом объеме ансамбля живущих в этот момент знаков языка определяет относительную частоту употребления этого знака на данный момент среди всех знаков языка. В процессе эволюции знака его смысловый объем сначала растет (соответственно росту полисемии и абстрактности составляющих ее значений), а затем постепенно спадает до нуля (до выхода знака из употребления). Смысловый объем начального значения знака случайным образом меняется от знака к знаку в соответствии с определенным законом распределения. Это распределение вместе с распределениями АСП, активности и стабильности знаков определяет частоты употребления знаков, живущих на любом одномоментном срезе эволюции знакового ансамбля языка.

Возникает вопрос: какой математической моделью может быть описан процесс развития языкового знака и знаковых ансамблей языка?

Детерминированная диссипативная динамическая модель эволюции языкового знака

Модель процесса рождения новых значений знака

Введем переменные, характеризующие процесс роста числа значений отдельного языкового знака. Обозначим через G максимальный АСП знака; через k — номер значения, появившегося на k -м шаге эволюции значений знака, $k = 1, 2, \dots, G$; через t_k — момент появления k -го значения ($t_1 = 0$ — начальный момент, момент появления знака в его начальном значении); через v_k — скорость роста числа значений знака на k -м шаге эволюции (v_1 — начальная скорость).

Очевидно,

$$v_k = ((k + 1) - k)/(t_{k+1} - t_k) = 1/\Delta t_k, \quad (1)$$

где $\Delta t_k = t_{k+1} - t_k$ — промежуток времени между рожденьями $k + 1$ -го и k -го значений знака. В процессе рождения новых значений знака его АСП растрачивается, уменьшаясь на единицу при каждом рождении нового значения, так что на k -м шаге (уровне) эволюции АСП знака оказывается равным $G - k$ (после рождения k -го значения знака может родиться только $G - k$ новых значений).

Естественно предположить, что скорость рождения новых значений знака v_k пропорциональна АСП знака на k -м уровне эволюции:

$$v_k = a(G - k), \quad k = \overline{1, G}, \quad a > 0. \quad (2)$$

Тогда максимальная скорость роста равна $v_1 = a(G - 1)$, а минимальная — нулю: $v_G = a(G - G) = 0$. Из соотношений (1)–(2) получаем рекуррентную формулу для моментов появления новых значений знака:

$$t_{k+1} = t_k + 1/(a(G - k)), \quad k = \overline{1, G - 1}, \quad t_1 = 0. \quad (3)$$

Отсюда видно, что интервал времени $\Delta t_k = t_{k+1} - t_k$ обратно пропорционален АСП k -го значения знака, причем чем больше величина коэффициента пропорциональности $1/a$, тем длиннее этот интервал. Следовательно, коэффициент $1/a$ имеет смысл некоторой «постоянной времени» $\tau = 1/a$ роста числа новых значений знака (чем больше τ , тем медленнее рост, так как больше Δt_k , и наоборот). Поэтому вместо коэффициента a более резонно использовать обратную величину τ , так что рекуррентное соотношение (3) примет вид:

$$t_{k+1} = t_k + \tau/(G - k), \quad k = \overline{1, G - 1}, \quad t_1 = 0. \quad (4)$$

Процесс рождения новых значений знака как диссипативный процесс, удовлетворяющий принципу наименьшего действия

К сожалению, проанализировать характер поведения решения рекуррентного уравнения (4) затруднительно. Однако если перейти временно к непрерывной переменной k , то рекуррентное уравнение (2) с учетом (1) и обозначения $\tau = 1/a$ можно записать в виде дифференциального уравнения:

$$dk(t)/dt = (G - k(t))/\tau, \quad k(0) = 1. \quad (5)$$

Это линейное дифференциальное уравнение с разделяющимися переменными. Его решение имеет вид:

$$k(t) = G - (G - 1) \exp(-t/\tau), \quad t \geq 0. \quad (6)$$

Как видим, $k(t)$ экспоненциально растет с ростом t от значения $k(0) = 1$ до $k(\infty) = G$. Очевидно,

$$v(t) = dk(t)/dt = (G - k(t))/\tau, \quad (7)$$

$$d^2k(t)/dt^2 = -(1/\tau)dk(t)/dt = -(1/\tau)v(t). \quad (8)$$

Последнее равенство означает, что «сила инерции» $d^2k(t)/dt^2$ процесса роста числа значений знака в каждый момент текущего времени t уравнивается «силой вязкого трения» $-(1/\tau)v(t)$. Следовательно, процесс роста новых значений знака является диссипативным процессом [Лоскутов, Михайлов, 2007]. Первоначальная «кинетическая энергия» такого процесса, равная $T(0) = v^2(0)/2 = (G - 1)^2/(2\tau^2)$, растрачивается на преодоление «сил вязкого трения», так что к моменту времени $t > 0$ ее остается

$$T(t) = v^2(t)/2 = (G - k(t))^2/(2\tau^2) < T(0). \quad (9)$$

С другой стороны, этот остаток «кинетической энергии» способен совершить «работу» против «сил вязкого трения», равную

$$U(t) = - \int_{k(t)}^G (1/\tau)v(t)dk(t) = (G - k(t))^2/(2\tau^2). \quad (10)$$

Эта «работа» называется «потенциальной энергией» (потенциальной функцией). Как видим, «потенциальная энергия» знака $U(t)$ на каждом уровне $k(t)$, то есть в каждый момент времени t , равна «кинетической энергии» знака $T(t)$, $U(t) = T(t)$, так что его «функция действия» $S(t) = |T(t) - U(t)| \equiv 0$ и, соответственно, «функционал действия»

$$S = \int_0^\infty |T(t) - U(t)|dt = 0$$

принимают минимальные (именно, равные нулю) значения. Следовательно, процесс $k(t)$ роста числа новых значений знака в рассматриваемой математической модели подчиняется принципу «наименьшего действия» [Айзерман, 1980; Маркеев, 1999], одному из фундаментальных вариационных принципов Природы, открытому Пьером Мопертюи в 1744 году.

Частная производная $\partial U/\partial k(t) = -(1/\tau)v(t)$ от потенциальной функции U по координате $k(t)$ определяет «силу вязкого трения», «силу сопротивления движению», что характерно для диссипативных систем. Применительно к языковому знаку можно сказать, что она определяет силу сопротивления языковой среде рождению новых значений знака. Возвращаясь к дискретной модели роста числа новых значений знака, можно сказать, что процесс эволюции языкового знака есть дискретный аналог диссипативного процесса.

Модель процесса выхода из употребления значений знака и общая детерминированная модель эволюции знака

Естественно предположить, что аналогично, но более медленно протекает процесс выхода из употребления значений знака. Тогда, снабдив верхними индексами процесс рождения новых

значений (индекс 1) и процесс выпадения значений из употребления (индекс 2), для i -го знака получим из (4):

$$\begin{aligned} t_{i,k+1}^{(1,2)} &= t_{i,k}^{(1,2)} + \tau_i^{(1,2)} / (G_i - k), \quad k = \overline{1, G_i - 1}, \\ t_{i,1}^{(1)} &= t_i, \quad t_{i,1}^{(2)} = \tau_{0i} + t_i, \quad \tau_i^{(2)} > \tau_i^{(1)}, \quad i = \overline{1, N}, \end{aligned} \quad (11)$$

где t_i — момент появления в языке i -го знака (слова), N — число слов в языке. Очевидно, $L_{i,k} = t_{i,k}^{(2)} - t_{i,k}^{(1)}$ — длительность жизни k -го значения i -го знака. Нетрудно видеть, что эта величина подчиняется рекуррентному соотношению:

$$L_{i,k+1} = L_{i,k} + (\tau_i^{(2)} - \tau_i^{(1)}) / (G_i - k), \quad L_{i,1} = \tau_{0i}, \quad k = \overline{1, G_i - 1}, \quad i = \overline{1, N}, \quad (12)$$

так что $L_{i,k+1} > L_{i,k}$, то есть длительность жизни каждого значения любого i -го знака увеличивается с ростом k .

Заметим, что в общем случае каждый i -й знак может появиться в языке как в единственном начальном значении, так и сразу с некоторым набором $k_{0i} > 1$ значений, то есть с некоторой начальной полисемией k_{0i} в диапазоне $1 \leq k_{0i} \leq G_i$, $i = \overline{1, N}$. При этом моменты появления всех этих k_{0i} значений совпадают с моментом появления знака в языке, то есть с моментом t_i . Полисемия i -го знака развивается с момента $t_{i,1}^{(1)} = t_i$ появления i -го знака в языке до момента $t_{i,G_i}^{(2)} = t_i + \tau_{0i} + L_{i,G_i}$ выхода из употребления последнего (G_i -го) значения i -го знака. Интервал времени длиной $L_i = \tau_{0i} + L_{i,G_i}$ от $t = t_{i,1}^{(1)}$ до $t = t_{i,G_i}^{(2)}$ — интервал жизненного цикла i -го знака, L_i — время его жизни.

Если в заданный момент времени T (на данном временном сечении) i -ый знак существует, то есть уже появился, но еще не вышел из употребления, так что $t_i \leq T < t_{i,G_i}^{(2)}$, то разность $A_i = T - t_i$ определяет возраст i -го знака в момент времени T , а все множество живущих в этот момент знаков с их возрастными определяет одномоментное распределение знаков по возрастам.

Смысловый объем знака и относительная частота его употребления

Поскольку знаков и их значений в языке существенно меньше, чем образов-смыслов, формирующихся в сознании носителей языка, каждый знак и его значение намекают в процессе коммуникации на множество смыслов из некоторой смысловой области. Размер этой области (и соответствующий «смысловый объем» знака или его значения) можно характеризовать числом содержащихся в ней смыслов (мощностью множества смыслов), если это число удастся установить, или величиной, пропорциональной этой мощности. На момент T временного сечения смысловый объем $V_i(T)$ i -го знака можно определить как число образов (в относительных единицах — десятках, сотнях, тысячах и т. п.), на которые намекает i -й знак в актах коммуникации на данном временном сечении жизни языка. Естественно рассматривать смысловый объем знака как величину, равную сумме смысловых объемов $V_{i,k}(T)$ всех значений i -го знака, живущих в момент времени T :

$$V_i(T) = \sum_{k \in M_i(T)} V_{i,k}(T). \quad (13)$$

Здесь $M_i(T)$ — множество значений i -го знака, составляющих активную полисемию знака в момент времени T :

$$M_i(T) = \overline{k_{i1}(T), k_{i2}(T)} \subset \overline{1, G_i}, \quad (14)$$

где $k_{i1}(T)$ — минимальный уровень, $k_{i2}(T)$ — максимальный уровень значений i -го знака, живущих в момент времени T .

В модели предполагается, что смысловой объем $V_{i,k}(T)$ k -го значения i -го знака монотонно растет с ростом уровня k значения знака (с ростом уровня абстрактности значения) от некоторого начального (первого, при $k = 1$) смыслового объема $V_{i,1}(T)$ до максимально возможного смыслового объема этого знака. Можно допустить, что этот рост подчиняется, например, степенному закону с неотрицательным показателем, необязательно целочисленным:

$$V_{i,k}(T) = V_{i,1}(T)k^{c_i}, \quad c_i > 0, \quad k \in M_i(T), \quad i = \overline{1, N}. \quad (15)$$

Вид этого закона подлежит уточнению по данным частотных словарей, но предположение о монотонности, по-видимому, должно оставаться в силе, так как оно вытекает из феноменологической модели развития языкового знака. Начальный смысловой объем $V_{i,1}(T)$ и показатель c_i роста смыслового объема i -го знака — свои для каждого знака.

Модель (11)–(15) является детерминированной дискретной диссипативной динамической моделью развития жизненного цикла языкового знака и ансамбля живущих в каждый данный момент времени знаков языка.

Очевидно, параметры жизненного цикла i -го знака (их 8 в нашей детерминированной модели) $t_i, G_i, \tau_{0i}, \tau_i^{(1)}, \tau_i^{(2)}, k_{0i}, V_{i,1}, c_i$ различны для различных знаков. Этими параметрами определяются и моменты рождения $t_{i,k}^{(1)}$, и моменты выхода из употребления $t_{i,k}^{(2)}$, и длительности жизни $L_{i,k}$, и смысловой объем $V_{i,k}$ каждого k -го значения i -го знака, и, наконец, само число G_i различных значений i -го знака (его АСП) и смысловой объем V_i каждого знака. Параметры $\tau_i^{(1)}$ и G_i определяют активность i -го знака, параметры τ_{0i} и L_{i,G_i} (то есть $\tau_{0i}, \tau_i^{(2)} - \tau_i^{(1)}, G_i$) — его стабильность. Параметры $V_{i,1}$ и c_i определяют смысловой объем i -го знака.

Стохастическая диссипативная динамическая модель эволюции ансамбля языковых знаков

Поскольку параметры модели характеризуют каждый знак, а все знаки разные, можно считать эти параметры для случайно выбранного знака случайными, подчиняющимися некоторым устойчивым законам распределения (в стационарном режиме появления, развития жизненного цикла и смены языковых знаков, то есть в стационарном процессе функционирования языка). Эти законы распределения должны проявляться во временных сечениях процесса функционирования языка. Именно на таких сечениях могут быть проверены (путем сравнения со словарями или путем анализа текстов) предполагаемые законы распределения параметров, а также предсказаны путем компьютерного моделирования законы распределения характеристик, определяемых этими параметрами (частотного распределения полисемии, распределения уровней абстрактности значений знаков, времени жизни, возраста знаков, частотно-рангового распределения знаков и пр.).

В стохастической модели развития во времени процессов функционирования больших ансамблей знаков на основе предложенной модели жизненного цикла отдельного языкового знака предполагается, что знаки возникают в языке в случайные моменты времени, образуя пуассоновский поток событий некоторой интенсивности λ (стационарный ординарный поток редких событий). В этом случае интервал времени $\tau_i = t_{i+1} - t_i$ между появлением соседних по времени знаков языка имеет экспоненциальное распределение с плотностью

$$p(\tau_i) = (1/\langle\tau\rangle) \exp(-\tau_i/\langle\tau\rangle), \quad \langle\tau\rangle = 1/\lambda, \quad (16)$$

с математическим ожиданием $\langle\tau\rangle$, обратным интенсивности потока λ .

Предполагается также, что постоянные времена $\tau_i^{(1)}$ и $\tau_i^{(2)}$ обратно пропорциональны нормированному G_i (то есть $G_i/\langle G \rangle$, где $\langle G \rangle$ — среднее значение максимальных АСП всех знаков

языка) для всех знаков, то есть $\tau_i^{(1,2)} = c^{(1,2)}\langle G \rangle / G_i$, где $c^{(1)}$, $c^{(2)}$ — параметры модели. Заметим, что предположение об обратной пропорциональности $\tau_i^{(1,2)}$ величине G_i является гипотезой, которая, как будет показано в следующем разделе, подтверждается статистически значимой близостью модельных и эмпирических распределений полисемии, полученных из представительных словарей. Однако возможно, что более адекватными будут другие гипотезы, например, гипотеза об обратной пропорциональности для $\tau_i^{(1)}$, но прямой пропорциональности для $\tau_i^{(2)}$, соответствующая высокой активности и высокой стабильности знака. Для проверки этой гипотезы необходимы достаточно точные эмпирические данные о распределении возрастов и длительностей жизни знаков.

Предполагается, что сами G_i образуют статистический ансамбль с экспоненциальным (точнее, с показательным, так как G_i могут принимать только целочисленные значения) законом распределения:

$$P(G_i) = (1 - \vartheta)\vartheta^{G_i-1}, \quad (17)$$

где параметр ϑ и математическое ожидание $\langle G \rangle$ величины G_i связаны соотношениями: $\langle G \rangle = 1/(1 - \vartheta)$, $\vartheta = (\langle G \rangle - 1)/\langle G \rangle$.

Предполагается, что задержки τ_{0i} начала выхода значений знаков из употребления также подчиняются экспоненциальному закону распределения с плотностью

$$p(\tau_{0i}) = (1/\langle \tau_0 \rangle) \exp(-\tau_{0i}/\langle \tau_0 \rangle) \quad (18)$$

с математическим ожиданием $\langle \tau_0 \rangle$ и не зависят статистически от значений G_i .

Кроме того, в стохастической модели предполагается, что моменты $t_{i,k}^{(1)}$ рождения новых значений и $t_{i,k}^{(2)}$ выхода из употребления этих значений тоже флуктуируют для каждого знака и значения, но так, что рождение значения более высокого уровня $k + 1$ происходит не раньше рождения значения предыдущего, более низкого уровня k , то есть так, что $t_{i,k+1}^{(1)} \geq t_{i,k}^{(1)}$, и аналогично $t_{i,k+1}^{(2)} \geq t_{i,k}^{(2)}$, $\forall k, \forall i$. Эти флуктуации в модели описываются независимыми равномерно распределенными случайными величинами с нулевыми средними значениями и полуширинами

$$t_{i,k+1}^{(1,2)} - t_{i,k}^{(1,2)} = \tau_i^{(1,2)} / (G_i - k) \quad (19)$$

плотностей распределения соответственно.

В стохастической модели развития знакового ансамбля языка предполагается также, что каждый i -й знак может появиться в языке либо в единственном начальном значении, либо сразу с некоторым набором $k_{0i} > 1$ значений, то есть с некоторой начальной полисемией k_{0i} в диапазоне $1 \leq k_{0i} \leq G_i$, $i = \overline{1, N}$. Естественно предположить, что наиболее вероятна единичная начальная полисемия (единственное значение знака), а более высокие значения начальной полисемии k_{0i} все менее и менее вероятны. Следовательно, условное (при фиксированном G_i) вероятностное распределение начальной полисемии является усеченным и монотонно спадающим с ростом k_{0i} распределением. В качестве такого распределения удобно взять усеченное показательное (дискретное экспоненциальное) распределение, обеспечивающее максимальную вероятность появления в языке однозначных слов и невозможность появления знаков, начальная полисемия которых превышает их максимальный АСП G_i :

$$P(k_{0i}|G_i) = \frac{1 - \gamma_i}{1 - \gamma_i^{G_i}} \gamma_i^{k_{0i}-1}, \quad k_{0i} = \overline{1, G_i}, \quad i = \overline{1, N}. \quad (20)$$

Параметр этого распределения $\gamma_i < 1$ и при $G_i = 1$ несущественен, так как $P(1|1) = 1$. При $G_i > 1$ $\gamma_i = P(k_{0i} - 1|G_i)/P(k_{0i}|G_i) < 1$ является параметром модели, определяющим скорость спада распределения. Условное математическое ожидание случайной величины k_{0i} имеет довольно

сложный вид: $\langle k_0 | G_i \rangle = 1/(1 - \gamma_i) - G_i \gamma_i^{G_i} / (1 - \gamma_i^{G_i})$. Эта величина не может быть сделана независимой от индекса i , то есть одинаковой для всех знаков, так как условное среднее значение начальной полисемии знака существенно зависит от его АСП G_i . Однако параметр γ_i можно связать с G_i с помощью некоторого инварианта, который можно считать общим для всех знаков параметром распределения (20). Действительно, если предположить, что усеченное распределение начальной полисемии знаков имеет одинаковую природу — усечение экспоненциального распределения таким образом, чтобы площадь «хвоста» распределения правее точки G_i была одинаковой, равной некоторой малой постоянной величине α для всех знаков, то есть чтобы величина G_i была квантилью уровня $1 - \alpha$ экспоненциального распределения, то мы получим следующий инвариант: $\gamma_i^{G_i} = \alpha$. Примем α в качестве параметра распределения (20).

Наконец, в стохастической модели предполагается, что начальный смысловый объем $V_{i,1}(T)$ знака, входящий в выражение (15), является случайной величиной и подчиняется экспоненциальному закону распределению с плотностью

$$p(V_{i,1}) = (1/\langle V_1 \rangle) \exp(-V_{i,1}/\langle V_1 \rangle), \quad (21)$$

так что его математическое ожидание $\langle V_1 \rangle$ становится еще одним параметром стохастической модели. Выбор экспоненциального закона распределения объясняется естественным соображением, вытекающим из феноменологической модели, что малый начальный смысловый объем знака более вероятен, чем большой.

Неотрицательный показатель c_i роста смыслового объема знака принимается в данной модели для простоты постоянным параметром $\langle c \rangle$ модели, одинаковым для всех знаков. Однако не исключено, что в дальнейших исследованиях его придется рассматривать как случайную величину, распределенную также, например, по экспоненциальному закону с плотностью $p(c_i) = (1/\langle c \rangle) \exp(-c_i/\langle c \rangle)$ с математическим ожиданием $\langle c \rangle$ в качестве параметра модели.

Можно допустить, что относительная частота $p_i(T)$ употребления i -го знака в актах общения (в момент времени T) равна доле его смыслового объема в суммарном смысловом объеме $V(T)$ всех живущих в момент времени T знаков языка:

$$p_i(T) = V_i(T)/V(T), \quad V(T) = \sum_{i \in I(T)} V_i(T), \quad i = \overline{1, N}, \quad (22)$$

где $I(T)$ — множество знаков, живущих в момент T . Если все живущие в момент времени T знаки ранжировать в порядке убывания относительной частоты их употребления, получится частотно-ранговое распределение знаков языка на этот момент времени.

Статистическое моделирование эволюции ансамбля языковых знаков

Параметры стохастической модели

Предложенная выше стохастическая диссипативная динамическая модель эволюции языковых знаков содержит 8 параметров:

- средний интервал времени $\langle \tau \rangle$ между соседними знаками в пуассоновском потоке знаков, рождающихся в процессе эволюции языка (величина, обратная интенсивности потока);
- среднее значение $\langle G \rangle$ АСП знаков (параметр показательного распределения АСП);
- коэффициент $c^{(1)}$ обратной пропорциональной зависимости постоянной времени $\tau^{(1)} = c^{(1)} \langle G \rangle / G$ роста новых значений каждого знака от его нормированного АСП $G/\langle G \rangle$ (параметр, характеризующий скорость рождения новых значений знаков);
- коэффициент $c^{(2)}$ обратной пропорциональной зависимости постоянной времени $\tau^{(2)} = c^{(2)} \langle G \rangle / G$ выхода из употребления значений каждого знака от его нормированного

АСП $G/\langle G \rangle$ (параметр, характеризующий скорость выхода значений знаков из употребления, $c^{(2)} \gg c^{(1)} > 0$);

— среднее время $\langle \tau_0 \rangle$ запаздывания начала выхода значений знака из употребления относительно момента появления знака в языке (параметр экспоненциального распределения запаздывания);

— средний начальный смысловый объем знака $\langle V_1 \rangle$ (параметр экспоненциального распределения начального смыслового объема);

— инвариантный параметр $\langle \alpha \rangle$ усеченного показательного распределения начальной полисемии знаков;

— средний показатель $\langle c \rangle$ роста смыслового объема значений знаков.

Значения этих параметров определяют конкретный вид результатов статистического моделирования эволюции языковых знаков. Однако математическая модель может быть полезной только тогда, когда она достаточно адекватно предсказывает реальные распределения знаков по числу значений (полисемии), по длительности жизни, по возрасту, частотно-ранговое распределение употребительности знаков и так далее, то есть когда результаты моделирования согласуются с реальными данными представительных словарей того или иного языка или с данными представительных корпусов текстов.

Идентификация модели и проверка ее адекватности по распределениям полисемии и частотно-ранговым распределениям, полученным из представительных словарей

Для проверки адекватности модели по отношению к распределению полисемии использовались шесть представительных словарей русского и английского языков (четыре — русского и два — английского): **СЯП** — «Словарь языка А. С. Пушкина»; **ССРЛЯ** — «Словарь современного русского литературного языка» в 17-ти т.т. (1948–1965), (большой словарь); **МАС** — «Словарь русского языка» под ред. А. П. Евгеньевой (1957–1961), (средний словарь); **СО** — «Словарь русского языка» С. И. Ожегова в 4-х т.т. (1972, 9-е издание), (краткий словарь); **Shorter** — «Shorter Oxford English Dictionary» (1962), (средний словарь); **Hornby** — A. S. Hornby. «Oxford Advanced Learner's Dictionary of Current English» (1982), (краткий словарь).

Для проверки адекватности модели по отношению к частотно-ранговому распределению использовались три частотных словаря русского и английского языков (два — русского и один — английского): **СЯП** — Длинно-возрастно-частотно-семантический словарь языка А. С. Пушкина; **ЧСГТЯ** (RusFrqDict-Jadernyj Corpus) — Частотный словарь по Ядерному корпусу русских газет конца XX века (объем корпуса — 1350 тыс. словоупотреблений; создан в 2001 году в Лаборатории общей и компьютерной лексикологии и лексикографии филологического факультета МГУ); **БЧСАЯ** — (BNC-WholeCorp-FrDict) — Частотный словарь английского языка, созданный на основе British National Corpus.

На рис. 2–4 в качестве примера приведены в логарифмическом масштабе по обеим осям эмпирические (по словарям) и теоретические (по временным сечениям потока знаков, получаемых в модели) распределения актуальной полисемии m знаков русского и английского языков при соответствующем подборе параметров модели (ее идентификации). Моделировалась динамика развития ансамбля 10000 знаков. Среднее значение интервалов между появлением знаков выбиралось равным $\langle \tau \rangle = \langle t_{i+1} - t_i \rangle = 0.05$ (в условных единицах измерения времени). Идентификация по четырем параметрам модели $\langle G \rangle$, $c^{(1)}$, $c^{(2)}$, $\langle \tau_0 \rangle$ производилась методом стохастической аппроксимации [Вазан, 1972]. При этом проверялась нулевая гипотеза H_0 об идентичности эмпирического и теоретического законов распределения полисемии против альтернативной гипотезы об их различии (по критерию Колмогорова–Смирнова). Во всех случаях (для всех шести перечисленных выше словарей) уровень значимости критерия был $p \gg 5\%$, так что нулевая гипотеза H_0 не отвергалась.

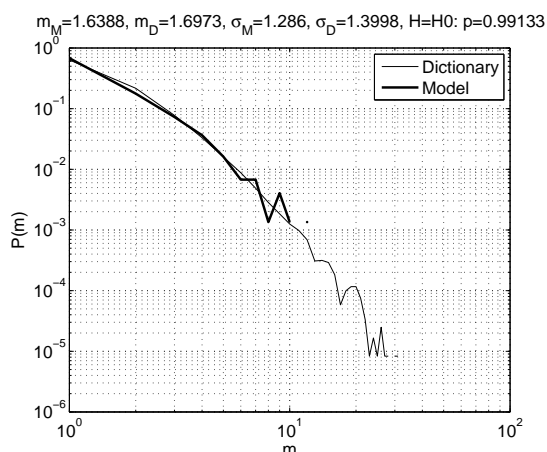


Рис. 2. Распределение полисемии по ССРЛЯ

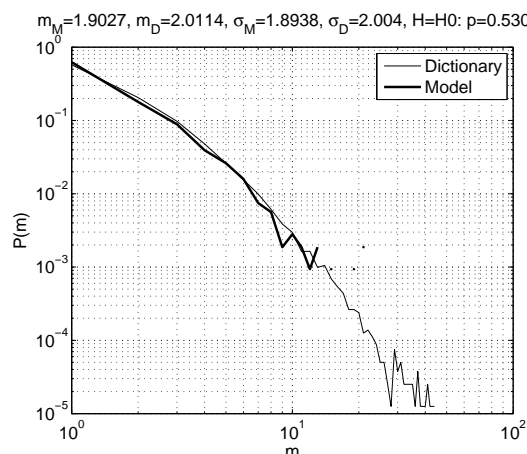


Рис. 3. Распределение полисемии по Shorter

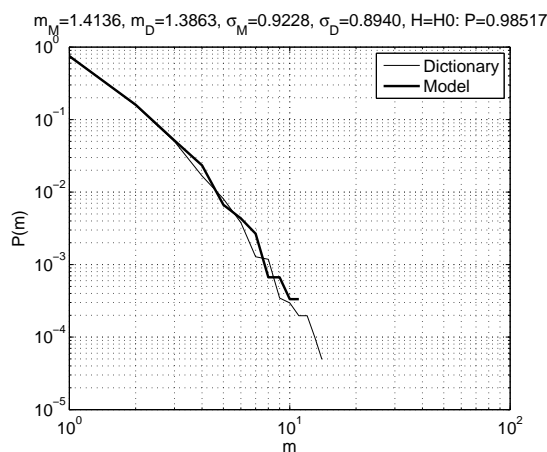


Рис. 4. Распределение полисемии по СЯП

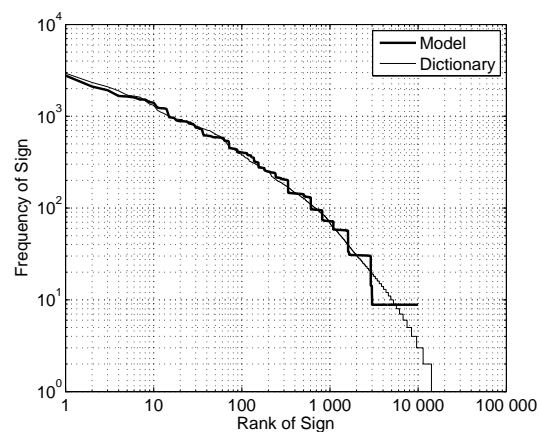


Рис. 5. Частотно-ранговое распределение по СЯП

Как видно из приведенных рисунков, кривые распределений полисемии, полученные по предложенной диссипативной стохастической динамической модели эволюции языковых знаков, отлично согласуются с кривыми эмпирических распределений, полученными из представительных словарей русского и английского языков. По-видимому, можно ожидать согласия модельных распределений полисемии также и с данными представительных словарей других языков.

На рис. 5–6 представлены частотно-ранговые распределения знаков, полученные упорядочением (ранжированием) знаков, живущих в момент времени T , в порядке убывания частоты их использования в языке по данным частотных словарей и по модельным данным после настройки параметров модели на соответствующие словари. Идентификация модели с частотными словарями осуществлялась дополнительно по трем параметрам $\langle V_1 \rangle, \alpha, \langle c \rangle$ также методом стохастической аппроксимации [Вазан, 1972]. Из рис. 5 видно хорошее совпадение модельного частотно-рангового распределения с соответствующим распределением знаков в языке А. С. Пушкина (по СЯП). Достаточно хорошим является также совпадение частотно-ранговых распределений знаков модели и больших словарей русского и английского языков (рис. 6). Модельный участок распределения довольно короткий из-за трудностей моделирования очень больших ансамблей знаков (процесс моделирования оказывается слишком длительным). Модель позволяет исследовать условия и границы выполнимости известного закона Ципфа–Мандельброта для частотно-ранговых распределений знаков из различных видов текстов.

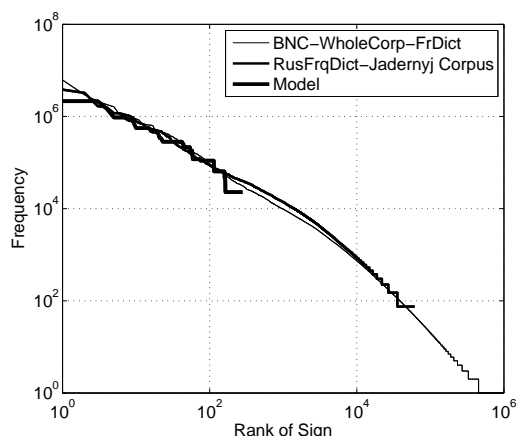


Рис. 6. Частотно-ранговое распределение по БЧС

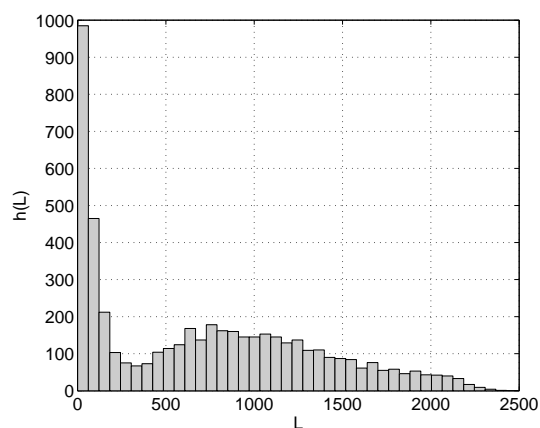


Рис. 7. Распределение длительности жизни знаков

Распределение знаков по длительности жизни и возрасту

На временном сечении (разрезе) моделируемой эволюции статистического ансамбля (потока) языковых знаков можно получить не только распределения их полисемии и частотно-ранговые распределения, но и распределения их длительности жизни, возраста и др.

В качестве примера на рис. 7 представлена одномоментная гистограмма модельного распределения длительности жизни знаков при значениях параметров модели, соответствующих ССРЛЯ. На рис. 8–11 представлены модельные и эмпирические распределения возраста знаков при значениях параметров модели, соответствующих СЯП, причем на рис. 8 приведены возрастные распределения знаков по возрастным зонам (возрастным интервалам) типа 1 (коротко, зонам 1), соответствующим табл. 1 (№ зоны, интервал в годах, число знаков в зоне по СЯП и по модели), а на рис. 9–11 — распределения знаков по первым 10 возрастным зонам типа 2 (коротко, зонам 2), соответствующим постоянной длительности в 50 лет и равномерному увеличению возраста каждой зоны на 50 лет, начиная с 1801 года. Таким образом, на рис. 9–11 представлены распределения знаков, возраст которых не превышает 500 лет (из-за ограниченного объема компьютерного эксперимента).

Таблица 1. Возрастные зоны типа 1

№ зоны	1	2	3	4	5	6	7	8
Интервал	0–30	31–80	81–180	181–380	381–780	781–1180	1181–2880	2881–5880
СЯП	2767	3333	3880	3155	3489	1339	2179	128
Модель	1891	2034	3707	6523	5789	326	0	0

Насколько модельное (теоретические) распределение длительностей жизни знаков будет соответствовать эмпирическим распределениям, полученным из представительных словарей или больших корпусов текстов, покажут дальнейшие исследования (в настоящее время авторы не располагают достаточно надежными данными по длительности жизни знаков).

Что же касается распределения возрастов знаков, то, как видно из рис. 8, модельные и эмпирические (по СЯП) распределения по возрастным зонам 1 довольно сильно (статистически значимо) различаются. Однако в представлении данных по возрастным зонам 2 после соответствующего усреднения, приведения данных к коротким интервалам этих зон (рис. 9–11) различие становится значительно меньше. Сравнение модельных и эмпирических распределений

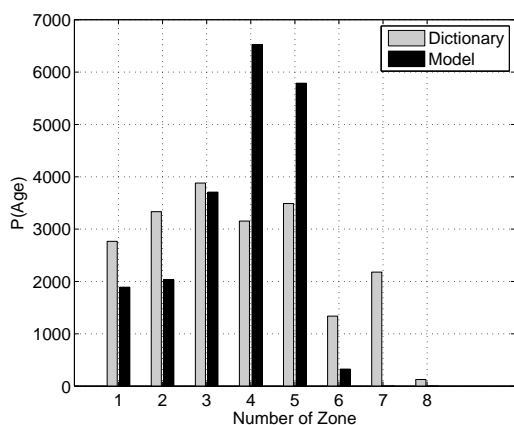


Рис. 8. Распределение возрастов по зонам 1 (СЯП)

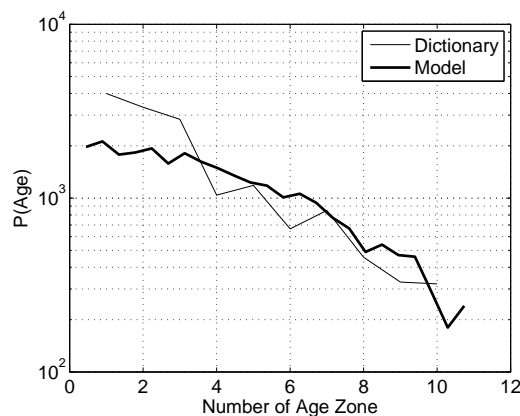


Рис. 9. Возрастное распределение знаков по СЯП

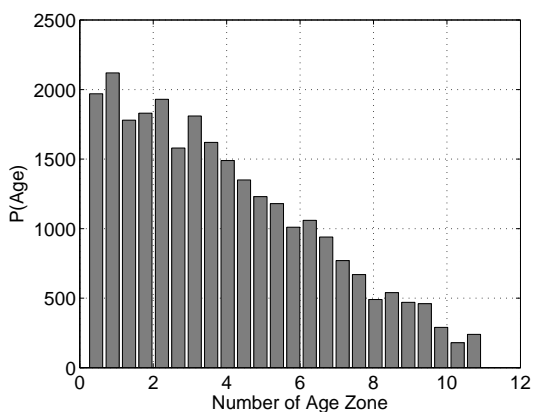


Рис. 10. Распределение возрастов по зонам 2 (модель)

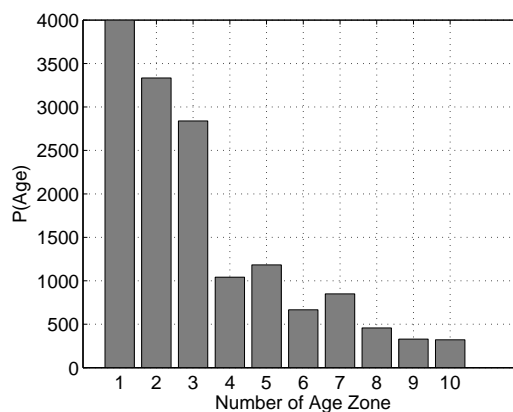


Рис. 11. Распределение возрастов по зонам 2 (СЯП)

возрастов знаков на более продолжительном интервале времени требует значительного увеличения объема компьютерного эксперимента. Но уже и при этом объеме данных из рис. 9 видно, что распределение возрастов знаков близко к экспоненциальному (кривые распределений в логарифмическом масштабе близки к прямым линиям).

Заключение

Итак, предложенная здесь диссипативная стохастическая динамическая модель развития языковых знаков позволяет моделировать и исследовать потоки языковых знаков в предположении статистической независимости их появления в языке, развития многозначности, постепенной утраты многозначности и, наконец, выхода знаков из употребления. При этом моделируются знаки различной активности в образовании новых, как правило, все более абстрактных по смыслу значений, различной стабильности, а следовательно, длительности жизни и возраста в каждый текущий момент времени, различной полисемии, различного смыслового объема и, следовательно, различной частоты использования в языке. Все эти характеристики знаков случайно изменяются от знака к знаку по определенным устойчивым статистическим законам распределения, производимым, по нашему предположению, от экспоненциальных законов распределения как одних из устойчивых статистических законов природы. Справедливость предложенной модели подтверждается хорошим совпадением модельных распределений полисемии

знаков, а также модельных частотно-ранговых распределений с соответствующими распределениями, полученными из представительных частотных, хронологических и толковых словарей русского и английского языков.

«За бортом» предлагаемой модели остаются пока длины знаков и их распределения, связь длин знаков с частотами их употребления в языке. Исследование этих вопросов является ближайшей целью авторов.

Список литературы

- Айзерман М. А.* Классическая механика. — М.: Наука, Гл. ред. физ.-мат. лит., 1980. — 368 с.
- Вазан М.* Стохастическая аппроксимация. — М.: Мир, 1972. — 296 с.
- Василенко Т.* Миф о 80/20.
<http://www.improvement.ru/zametki/pareto/>
- Земская Е. А.* Словообразование как деятельность. — М.: Наука, 1992. — 221 с.
- Крылов Ю. К.* Вероятностно-статистические модели синергетизма // Математическое описание ценозов и закономерности технетики. Философия и становление технетики. Вып. 1. — Абакан: Центр системных исследований, 1996. — С. 110–142.
- Крылов Ю. К., Кудрин Б. И.* Целочисленное аппроксимирование ранговых распределений и идентификация техноценозов // Ценологические исследования. Вып. 11. — М.: Центр системных исследований, 1999. — С. 7–78.
- Крылов Ю. К., Якубовская М. Д.* Статистический анализ полисемии как языковой универсалии и проблема семантического тождества слов // НТИ. — 1977. — Сер. 2, № 3. — С. 1–6.
- Лоскутов А. Ю., Михайлов А. С.* Основы теории сложных систем. — М.–Ижевск: Институт компьютерных исследований, 2007. — 620 с.
- Мандельброт Б.* Теория информации и психологическая теория частот слов // Математические методы в социальных науках. — М.: Прогресс, 1973. — С. 326–337.
- Маркеев А. П.* Теоретическая механика: учеб. для университетов. — М.: ЧеРо, 1999. — 572 с.
- Орлов Ю. К.* Невидимая гармония // Число и мысль. Вып. 3. — М.: Знание, 1980. — С. 70–105.
- Орлов Ю. К.* Динамика частотной структуры (Компьютерная графическая модель зависимостей вида «ранг – частота»). Анализ. Прогноз. Сравнение. — Тбилиси, 2001. — 32 с.
- Поддубный В. В., Поликарпов А. А.* О диссипативной стохастической динамической модели развития языковых знаков // Информационные технологии и математическое моделирование (ИТММ-2010): Материалы IX Всероссийской науч.-практ. конф. с междунар. участием (19–20 ноября 2010 г.). — Томск: Изд-во Том. ун-та, 2010. — Ч. 2. — С. 110–115.
- Поддубный В. В., Поликарпов А. А. (а).* Диссипативная стохастическая динамическая модель эволюции языковых знаков // Труды XIV Междунар. конф. по эвентологической математике и смежным вопросам (ЭМ'2010). / Под ред. О. Воробьева. — Красноярск: Крас. гос. торг.-эконом. ин-т, Сиб. фед. ун-т, 2010. — С. 182–186.
- Поддубный В. В., Поликарпов А. А.* Вывод закона синхронного полисемического распределения языковых знаков на основе диссипативной стохастической динамической модели эволюции знаковых ансамблей // Синхронное и диахронное в сравнительно-историческом языкознании: Материалы VII Междунар. науч. конф. — М.: «Добросвет», Изд-во «КДУ», 2011. — С. 182–190.
- Поликарпов А. А.* К теории жизненного цикла лексических единиц // Прикладная лингвистика и автоматический анализ текстов: Тез. докл. науч. конф. (28–30 января 1988 г.). — Тарту: Изд-во Тартуского университета, 1988. — С. 66.

- Поликарпов А. А.* Лексическая полисемия в эволюционном аспекте // *Linguistica-1990 (Acta et Commentationes Universitatis Tartuensis, No 911)*. — Тарту: Изд-во Тартуского университета, 1990. — С. 76–86.
- Поликарпов А. А.* Закономерности жизненного цикла слова и эволюция языка. Статья 1. Моделирование основных системных соотношений // *Русский филологический вестник*. — 1994. — Том 79, вып. 1. — С. 85–100.
- Поликарпов А. А.* Закономерности жизненного цикла слова и эволюция языка. Статья 2. Теория и эксперимент // *Русский филологический вестник*. — 1995. — Том 80, вып. 1. — С. 77–92.
- Поликарпов А. А.* Циклические процессы в становлении лексической системы языка: Моделирование и эксперимент: автореф. дис. д-ра филол. наук. — М., 1998. — 55 с.
- Поликарпов А. А.* Системно-квантитативный подход в лингвистике // *Филологические школы и их роль в систематизации научных исследований*. — Смоленск: Маджента, 2007. — С. 35–59.
- Поликарпов А. А.* Системная зависимость степени сохранности древнерусских слов в современном русском языке от их возраста, категориальной принадлежности, частоты и полисемии // *Лингвистическая компаративистика в культурном и историческом аспекте / Под общ. ред. В. А. Кочергиной*. — М.: Изд-во Моск. ун-та, 2007а. — С. 232–260.
- Поликарпов А. А.* Системно-квантитативный анализ многозначности лексики китайского языка в ее взаимосвязи со структурными и стилистическими характеристиками // *Вестник Смоленского государственного университета*. — 2009. — №4 (8). — С. 133–152.
- Поликарпов А. А.* Универсальные, типологические, локально-смысловые и частно-лингвистические факторы исторической сменяемости знаковых единиц языка // VIII Международная конференция по языкам Дальнего Востока, Юго-Восточной Азии и Западной Африки (Москва, 23–24 сентября 2009 г.): тезисы и доклады. — М.: Ключ-С, 2009а. — С. 125–141.
- Поликарпов А. А., Кукушкина О. В., Токтонов А. Г.* Проверка теоретически предсказанных неодериватологических закономерностей данными русской корпусной неодериватографии // *Теория и история славянской лексикографии*. / Отв. ред. М. И. Чернышёва: Сб. М., 2008. — С. 392–427.
- Поликарпов А. А., Селезнева-Елецкая Л. А.* Степень абстрактности и субъективности смысла — факторы варьирования степени сохранности во времени его лексических обозначений // *Сравнительно-историческое исследование языков: Современное состояние и перспективы: Сб. статей по материалам Междунар. науч. конф. (Москва, 22–24 января 2003 г.)*. / Составитель В. А. Кочергина. — М.: Изд-во Моск. ун-та, 2004. — С. 327–376.
- Ahlswede R., Arikan E., Bäumer L., Deppe C.* Information Theoretic Models in Language // *Information Transfer and Combinatorics, LNCS 4123*. — Berlin–Heidelberg: Springer-Verlag, 2006. — P. 769–787.
- Ferrer i Cancho R, Sole R. V.* Least Effort and the Origins of Scaling in Human Language // *PNAS*. — 2003. — Vol. 100, No. 3. — P. 788–791.
- Khmelev D. V., Polikarpov A. A.* Basic Assumptions about a Sign's Life Cycle for Mathematical Modelling of Language System Evolution // *Abstracts of Papers from Qualico'2000 — 4th Quantitative Linguistics Conference (Prague, 24–26 August 2000)*. — Prague, 2000.
- Krylov Yu. K.* A Markov Model for the Evolution of Lexical Ambiguity // *Journal of Quantitative Linguistics*. — 1995. — Vol. 2, No. 1. — P. 19–27.
- Li W.* Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution // *IEEE Transactions on Information Theory*. — 1992. — Vol. 38(6). — P. 1842–1845.
- Mandelbrot B.* Information Theory and Psycholinguistics // *Wolman B. B., Nagel E.* Scientific psychology. — Basic Books, 1965. Reprinted as: *Mandelbrot B.* Information Theory and Psycholinguistics // *Oldfield R. C., Marchall J. C.* Language. — Penguin Books, 1968.
- Mitchener W. G., Nowak M. A.* Chaos and Language // *Proc. R. Soc. Lond. B*. — 2004. — Vol. 271. — P. 701–704.

- Nowak M. A., Komarova N. L., Niyogi P.* Computational and Evolutionary Aspects of Language // *Nature*. — 2002. — Vol. 417. — P. 611–617.
- Nowak M. A., Krakauer D. C.* The Evolution of Language // *PNAS*. — 1999. — Vol. 96, No. 14. — P. 8028–8033.
- Plotkin J. B., Nowak M. A.* Language Evolution and Information Theory // *J. Theor. Biol.* — 2000. — Vol. 205. — P. 147–159.
- Polikarpov A. A.* On the Hypothesis of Word Life Cycle // *Qualico-91. First Conference on Quantitative Linguistics: Abstracts*. — Trier: Trier Univ.Press, 1991. — P. 60–63.
- Polikarpov A. A.* A Model of the Word Life Cycle // *Contributions to Quantitative Linguistics* / Ed. by R. Koehler, B. B. Rieger. — Dordrecht: Kluwer, 1993. — P. 53–66.
- Vogt P.* Modeling Interactions Between Language Evolution and Demography // *Human Biology*. — 2009. — Vol. 81, Nos. 2–3. — P. 237–258.
- Zipf G. K.* Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology. — Cambridge, MA: Addison–Wesley, 1949. Reprinted as: *Zipf G. K.* Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology, 1st ed. — New York: Hafner, 1972.