УДК 81'32

В. А. Долинский

доктор филологических наук, профессор кафедры прикладной и экспериментальной лингвистики Института прикладной и математической лингвистики факультета английского языка МГЛУ; e-mail: vdolinsky@yandex.ru

СЕМЕЙСТВО РАНГОВЫХ РАСПРЕДЕЛЕНИЙ В КВАНТИТАТИВНОЙ ЛИНГВИСТИКЕ

В статье рассматривается семейство ранговых распределений в квантитативной лингвистике: частотные характеристики слов, полисемия, ассоциативные поля слов, стилеметрия, языковые различия, распределения языков и другие объекты изучения в квантитативном аспекте.

Анализируются языковые модели, развиваемые в русле лингвистического ассоцианизма, синергетики и квантитативной лингвистики.

Структурообразующим компонентом значения слова является частотность вызываемых им ассоциативных реакций. Ассоциативные поля слов раскрывают психологически адекватную картину мира носителей языка.

Эффект концентрации и рассеяния, характерный для частотной структуры текстов, имеет своим источником, по-видимому, такой механизм порождения текста, который обусловлен ассоциативной природой языкового сознания.

В качестве иллюстраций к статье приведены таблицы и графики различных ранговых распределений в квантитативной лингвистике.

Ключевые слова: квантитативная лингвистика; корпусная лингвистика; частотно-ранговые распределения; закон Ципфа; ассоциативные поля слов; лексика; семантика; стилеметрия; языковые различия; распределения языков.

V. A. Dolinsky

Doctor of Philology, Professor, Department of Applied and Experimental Linguistics, Institute of Applied and Mathematical Linguistics, Faculty of the English Language, MSLU; e-mail: vdolinsky@yandex.ru

FAMILY OF RANK DISTRIBUTIONS IN QUANTITATIVE LINGUISTICS

The article considers the family of rank distributions in quantitative linguistics: word frequency characteristics, polysemy, word association fields, stylometrics, language distinctions, language distributions and other objects of study in quantitative aspect.

The language models being developed within the framework of linguistic associationism, synergetics, and quantitative linguistics are analyzed.

The frequency of association responses evoked by a word is a structural component of the word's meaning. Word association fields reveal the psychologically relevant worldview of native speakers.

The concentration and diffusion effect that is inherent in the frequency structure of texts appears to derive from a text generation mechanism that is driven by the associative nature of language consciousness.

Tables and graphs of different rank distributions in quantitative linguistics are given as illustrations to the article.

Key words: quantitative linguistics; corpus linguistics; rank frequency distributions; Zipf's law; word association fields; vocabulary; semantics; stylometrics; language distinctions; language distributions.

Распределение в статистике определяется как перечисление значений случайной величины и их вероятностей. Распределение – это разбиение совокупности лингвистических явлений на классы. В широком смысле распределение можно понимать как: «упорядоченную совокупность количественно выраженных значений, т. е. результатов измерения объекта, обычно с указанием значимости (частоты, вероятности, ранга) этих значений в данной совокупности» [Тулдава 1987, с. 41].

> Распределение является и процессом, и результатом такого разбиения, группировки, упорядочения. Распределение – это ряд значений признака лингвистического явления и численностей, частот этих явлений [Алексеев 2001, с. 26–27].

Ранговым называется распределение, в котором значения случайной величины расположены в порядке их убывания.

Лексическая система языка, состоящая из необозримого множества слов, распределяет в речи свои единицы по принципу «согласованного оптимума» (*илл. 1*). Первым, кто обратил внимание на удивительную гармоничность (parsimony, equilibrium – упорядоченность в частотном распределении) в речевой реализации лексического «инвентаря» разных языков, казавшегося ранее «свалкой» неупорядоченного словесного материала, был выдающийся американский филолог, психолог, математик Джордж Кингсли Ципф (1902–1950). Основоположник квантитативной лингвистики и пионер естественнонаучного подхода к языковым явлениям и процессам языковой саморегуляции, Дж. К. Ципф писал:

> Речь не может и не должна быть перманентно отвлечена от живых организмов и их поведения. Речевые явления не могут быть изолированы ни от содержания речи, ни от индивидуального, социального и культурного фона говорящего. Продукт речи должен рассматриваться как естественное психологическое и биологическое явление и изучаться

в объективном духе точных наук и их методами. Наш главный метод исследования – это применение статистических принципов к наблюдаемым явлениям потока речи. <…> Задача, не менее стоящая того, чтобы за нее взяться, так как … ее решение может пролить желанный свет на многие философские вопросы о природе жизни и смерти вообще. <…> Тенденция, которую отражают наши формулы, – на мой взгляд, результат «привычки» или «накопленной силы случая», – может рассматриваться как «стремление к устойчивости», понимая под устойчивостью (perseveration), следуя Скиннеру, психологический термин, описывающий часто наблюдаемый феномен. Я уверен, что уравновешивающие силы, подсказанные нашими формулами, – шире и глубже, и что они действуют, чтобы вызвать равновесие ради равновесия. Ни в каком отношении нельзя абстрагировать поток речи и всё то, что по традиции и фактически является частью речевого поведения, от всего остального нашего бытия. В самом деле, если кто-нибудь склонен верить, что его поток речи не связан с любым данным органом его анатомии, дайте ему только уколоться иголкой, раскаленной докрасна, в данный орган и наблюдайте возможный отзвук оттуда в его речепроизводстве <…>. Законы жизненного процесса хорошо могут быть обоснованы как изоморфные законам речи. Это убеждение приобретает прочность, если посмотреть на поток речи, как на равнодействующую множества сил, на которые мы реагируем. Линейность речевых привычек поразительна, даже в латентной речи и ассоциативно связанных словах» [Zipf 1949].

Ни одна квантитативная модель в лингвистике не вызывает столь разнообразных интерпретаций и оценок, порой противоположных, как модель Ципфа [Арапов 1988; Altmann 1997; Herdan 1966; Köhler, Altmann, Piotrowski 2005; Долинский 2012]. Ряд аналогий из смежных наук (закон Виллиса в биологии, закон Парето в экономике и др.) позволяет предположить, что закон Ципфа является одним из первых, математически сформулированных, законов в гуманитарном знании: обратно пропорциональная связь между частотой и рангом слова. Аналогично ранговому распределению слов в тексте, характеризуются встречаемость химических элементов в земной коре, распределение обитателей мирового океана по весу, космических тел по массе, распределения городов по численности населения, фирм по количеству служащих, что заставляет рассматривать этот тип распределения как универсальный «закон предпочтения», «закон иерархии значимостей» или (в терминах Ципфа) – «закон наименьшего усилия». В природе, обществе и языке действует некий универсальный механизм,

который абсолютно разнородные объекты выстраивает по иерархии (или ранжирует) сходным образом. Причины этого не вполне ясны, несмотря на обилие гипотез [Орлов 1980; Тулдава 1987; Чайковский 2001]. Представляется обоснованным вывод о том, что закон Ципфа является не одним из многих эмпирических распределений, но теоретическим законом, имеющим математическую базу в теории устойчивых негауссовых распределений.

Одним из реальных оснований применения квантитативных методов в изучении языка и речи (текста) следует признать объективную присущность языку количественных признаков, квантитативных характеристик [Алексеев 2001; Арапов 1988; Долинский 1988; Мартыненко 2009; Пиотровский 2006; Köhler, Altmann, Piotrowski 2005]. Повторяемость (рекуррентность, периодичность) языковых, в том числе лексических единиц, их воспроизведение в различных текстах является наиболее важным условием квантификации языкового материала и применения методов количественной математики для его анализа. Квантитативный подход способен охватить лишь определенный аспект языка и речи. Но это важный аспект, отражающий многие стороны речевой деятельности, которые невозможно обнаружить чисто качественным анализом. Квантитативный анализ иногда упрощает языковую реальность (например, когда не учитываются языковой полиморфизм, поливалентность и многообразие оттенков значений). Но при этом подходе возможен и более дифференцированный анализ полисемичности и других свойств языковых единиц. Часто оказывается, однако, что «переплетение» ассоциаций элементов языка настолько сложно и бесконечно, что оно не может поддаться в полной мере не только количественному, но и качественному анализу. Кроме того, следует указать на неизбежный недостаток качественного анализа, который нередко остается на уровне субъективных, произвольных, интерпретаций.

Выявление частотных свойств текстов и построенных на них словарей, установление связей между словами в парадигматике и синтагматике, изучение факторов порождения текста – должны вести к обобщению, упорядочению и осмыслению эмпирического материала на более высоком теоретическом уровне. Конечной целью является синтетический, интегральный, подход к изучению (описанию и объяснению) квантитативных свойств системы речевой деятельности и языкового механизма в целом – в неразрывном единстве квантитативного анализа с качественной интерпретацией.

Квантитативная лингвистика придает математическую форму языковым механизмам, порождающим структуры и зависимости. Поскольку язык принадлежит к классу самоорганизующихся систем, в нем всё взаимосвязано, но не всегда напрямую. Связи, обнаруживаемые в статистических распределениях, формируют цепи зависимостей. Чем больше расстояние между двумя элементами цепи, тем слабее зависимость. Изучение зависимостей (часто взаимозависимостей), установление закономерностей – важный аспект квантитативных исследований.

Одной из задач квантитативной лингвистики является составление *частотных словарей*, необходимость использования которых для решения прикладных и теоретических задач постоянно возрастает. Частотный словарь – это модель особым образом преобразованного текста, модель рангового распределения частот употребления единиц в тексте. Частотный словарь включает в себя упорядоченный список слов или других языковых единиц (словоформы, словосочетания), которые зарегистрированы составителем в обследованном им тексте или корпусе текстов и снабжены данными о частоте их употребления. (В конкордансе к этим данным добавляются и данные об их *адресе* и *контекстах*). С помощью частотных словарей можно ответить на вопросы: каков вокабуляр (лексический состав) языка (текста), с какой интенсивностью слова используются в речи, какие из них употребительнее (предпочтительнее) в той или иной сфере коммуникации. Частотные словари находят применение в лингводидактике, психолингвистике, терминографии, когнитивистике, культурологии, медицине, юриспруденции, криминологии и т. д. [Алексеев 2001; Частотный 1977; Ляшевская. Шаров 2009].

Хорошо известно, что слова по частоте их употребления (хотя и весьма неустойчивой для каждого отдельного слова) распределяются в любом крупном наборе (корпусе) текстов в виде «гиперболической лестницы». Например, в «Частотном словаре русского языка» [Частотный словарь 1977] почти 10 % словоупотреблений приходится на первые три слова (*в*, *и*, *не*); еще 10 % – на следующие 8 слов (*на*, *я*, *быть*, *что*, *он*, *с*, *а*, *как*); а 50 % всех словоупотреблений составляют всего 213 лексем (в этом словаре, построенном на текстах объемом 1 056 тыс. словоупотреблений, учтено более 39 тыс. разных слов). При этом более 13 тыс. слов (33,7 %) употреблено по одному разу.

Сравнение частотных словарей указывает, что нет никакой сходимости частот даже для самых употребительных слов, и их дисперсия превышает нормальную в сотни раз. На илл. 2–4 приведены данные разных частотных словарей и подвыборок одного словаря.

Если из подсчетов, выполненных на языковом материале, удается сделать качественные выводы, то в них сами числа не играют существенной роли, важны исключительно определяющиеся этими числами отношения *порядка*. На шкале порядка различаются степени интенсивности признака или свойства, происходит ранжировка единиц по степени проявленности. Порядковая шкала допускает операции больше / меньше, равенство / неравенство. Содержательный смысл имеет не сама по себе частота употребления слова, число производных от него или число значений, а найденное по этим величинам место данного слова среди других слов, его относительная *ценность*.

Количественным измерениям подвергаются и такие характеристики языковых единиц, как их «объем» в плане выражения и в плане содержания. Это длина слова (морфемы, словосочетания, предложения) и число его значений (выраженное, например, в словарных ЛСВ). Положение слова на ценностной шкале связано с его количественными характеристиками. Ранг слова, отражающий его употребительность в тексте (корпусе текстов), содержит информацию о положении слова в линейно упорядоченном словаре. Все данные о связи между рангом и другими количественными характеристиками единиц имеют вид нелинейных уравнений регресии. Эти зависимости как общие тенденции следует рассматривать не как точно установленные факты, а, скорее, как «полезные метафоры». Таковы зависимости между рангом *i* слова (определенном на основании его частоты F_i), длиной l_i, числом значений k_i. В общем случае, например, чем выше ранг слова, тем оно, в среднем, короче и многозначней [Арапов 1988; Altmann 1997].

Квантитативные параметры *полисемии*, измеренные по данным толковых словарей, варьируют в зависимости от объема словаря (большой, средний, малый). Как правило (но далеко не всегда) увеличение объема толковательной статьи сопровождается более «дробным» членением семантики слова (*илл. 5*).

Толкование дает лишь приближенное представление о значении слова. Число частей в нем – только косвенная оценка сложности значения, как число ломтей, на которые можно разрезать пирог, – косвенная характеристика его размеров [Арапов 1988, с. 137].

Реализацией свойства слова иметь некоторое число значений является *семантический объем* слова, который подлежит количественному измерению. Многозначное слово (лексема) как единица системы языка представляет собой единое смысловое целое, объединяющее в плане содержания ряд лексико-семантических вариантов, или отдельных «значений» слова. Опыт изучения полисемии показывает, что в рамках одного толкового словаря разграничение значений слов проводится достаточно последовательно, о чем свидетельствуют удачные попытки изучения общих квантитативных закономерностей полисемии на основе таких словарей [Поликарпов 1987; Тулдава 1987]. Данные различных языков показали, что семантический объем слова и количество слов с данным семантическим объемом в словаре статистически связаны: наибольшую долю в словаре составляют однозначные слова, затем следуют (в порядке убывания численности) группы слов с двумя, тремя и т. д. значениями (*илл. 6*). Такой вид распределения представляет собой, по-видимому, универсальную квантитативно-системную характеристику полисемии естественных языков. Здесь, как и во многих других ранговых распределениях, проявляется известный принцип *концентрации и рассеяния* языковых единиц.

Ряд актуальных проблем стилистики выделяется с позиции квантитативного анализа текстов (*стилеметрия*): лексическое «богатство» автора, классификация текстов на основе квантитативных параметров, лексическая близость и атрибуция текстов. Частота слова в авторских текстах – важная характеристика значимости этого слова. В первом приближении можно считать, что наиболее частые слова являются и наиболее важными. Однако при оценке авторских текстов, из принципа «значимое – часто» не следует, что «частое – значимо». В противном случае самым значимым словом для большинства языков служил бы союз *И* (или предлог *В*). Для того чтобы оценить ценность, значимость того или иного слова, следует сравнить его ранг в частотном словаре данного автора с рангом в частотном словаре, составленном по соответствующему классу текстов исследуемого периода.

На *илл. 7* приведено распределение персонажей романа М. А. Булгакова «Мастер и Маргарита» по частоте их упоминаний [Чайковский 2001]. На *илл. 8* дано распределение лексем с семантикой цветообозначения (цветовой спектр, в %) в текстах Н. В. Гоголя [Белый 1934].

Наряду со словарным толкованием слова, дающим логически непротиворечивое представление о его семантике в виде упорядоченного и структурированного текста, *ассоциативное поле* слова, полученное в эксперименте с группой носителей языка, также является своего рода текстом, манифестирующим смысл стимульного слова, раскрывающим содержание языкового сознания во всем его явном и имплицитном преломлении [Долинский 1988; 2012].

Важной областью использования данных психолингвистических экспериментов стало изучение языкового сознания и культурноязыковых архетипов. Слово в сознании человека всегда связано бесчисленными нитями с языковой картиной мира и, соответственно, осмысление каждой языковой единицы есть установление связи с лексической (и, шире, – культурноязыковой) системой в целом. В разных языках не существует «полных» лексических эквивалентов, как не существует их у носителей одного языка разных возрастов, профессий и т. д. [Norms 1970].

Слова – «черпаки смысла». Они пригодны для «зачерпывания» смыслового содержания, единого для всех, однако разные люди черпают далеко не одинаково. Так, для дальтоников *красный* и *зеленый* означают одно и то же, а для героев Дж. Оруэлла «Мир есть война». Зачерпывание глубокого и емкого содержания, проникновение в смысл – есть выход в пространство, полное индивидуальных ассоциаций, связываемых в узел психическим единством слова. Представление об атомах смысла, восходящее к Декарту, Лейбницу и авторам грамматики «Пор-Рояль», столь необходимое для построения логической семантики, в психологическом плане не более, чем дискретная составляющая континуального в своей основе сознания человека. Отнесение к слову фиксированного значения, описываемого в терминах признаков и компонентов, эквивалентно представлению о некоем усредненном и редуцированном его смысле. Такое значение, «очищенное» от спонтанных и непредвиденных ассоциаций, неизбежно оказывается бедным, жестким, неглубоким.

Представить психологически адекватный смысловой облик слова как конечную величину или как совокупность дискретных значений или признаков – то же, что представить себе трехмерный образ шара, построенного из кубиков. Какие образы вызывают в памяти такие, например, словосочетания и обороты речи: *глубокий смысл*, *вместить смысл*, *неисчерпаемый смысл*; *смысл*, *не лежащий на поверхности*,

углубиться в смысл, *сокровенный смысл*, *уловить смысл?* Не является ли ассоциативное поле слова ответом на вопрос: «Что *стоит за словом* X?» и «Какой смысл вы *вкладываете в слово* X?»

Наличие ассоциативного поля проявляется в широкоизвестных явлениях трудностей припоминания слова, состояниях, при которых искомое слово как бы находится «на кончике языка» («tip of the tongue phenomenon»), когда искомое слово замещается другим, взятым из общего ассоциативного смыслового поля.

> Когда человек не может вспомнить какое-то слово, хотя оно «вертится на кончике языка», или когда опознание ранее предъявленных знаков более успешно, чем их воспроизведение, – это означает, что человек помнит несколько больше, чем может вспомнить, а значит, что-то хранится в памяти в каком-то туманном виде [Аллахвердов 2003, с. 97–98].

Два главнейших свойства ассоциативного ряда – неопределенность порядка и безграничность количества. Отличительной чертой ассоциативного эксперимента является эмпирический и процедурный характер лингвистической информации, содержащейся в его данных, возможность их квантификации, а также возможность повторения эксперимента, оценки и верификации (фальсификации) его результатов. Это выгодно отличает данный подход, близкий к естественнонаучному, освобождая его от ограничений, присущих методу интроспекции.

Главными квантитативными характеристиками ассоциативных полей слов являются частота реакций (ассоциатов) и их репертуар (ассоциативный словарь). Частота F появления данной ассоциации в групповом эксперименте отражает общеизвестность, распространенность, стереотипность, «обкатанность» (или уникальность) этой ассоциации, свидетельствующие о наличии (отсутствии) таких же связей между словом-стимулом и словом-реакцией у данной группы (или свойственные языковому сообществу в целом). Репертуар (инвентарь) L ассоциаций, полученных от группы испытуемых, свидетельствует о широте или узости ассоциативных связей слова-стимула с другими единицами языка. Качественный состав ассоциативного поля неизмерим с помощью квантитативных параметров без привлечения дополнительных логико-дедуктивных гипотез или аксиоматических построений.

В процессе ассоциирования акцентирование отдельных аспектов содержания, отдельных областей смыслового поля не исключает одновременный учет всего ассоциируемого с данным словом. В условиях речепроизводства адекватность выбранного слова замыслу говорящего и ожиданиям слушающего обеспечивается подсознательным учетом всего так или иначе соотносимого с данным словом, всего того, что, оставаясь «за кадром», не получая выхода в «окно сознания», в то же время всегда может быть актуализовано через смену «фокусного расстояния», изменение «ракурса», редупликацию, «размывание» деталей и т. п. Прежде чем стать абстрактным элементом, знаком, подчиненным логике высказывания, слово должно быть проинтерпретировано на уровне неосознанной языковой символики. Дж. К. Ципф подчеркивал

> Особые возможности количественного изучения вербальных ассоциаций личности, которые более чего-либо другого характеризуют специфику индивидуальности [Zipf 1949].

Ассоциативные поля слов раскрывают психологически адекватную картину мира носителей языка. Структурообразующим компонентом значения слова является частотность вызываемых им ассоциативных реакций. График функции распределения вербальных ассоциаций слова оказывается тем зримым образом, который дает нам возможность представить семантический мир языковой личности и лингвокультурного сообщества.

> На илл. 9 приводится график ранговых распределений ассоциативных полей русских существительных СМЫСЛ, РАДОСТЬ, ЛЮБОВЬ [Долинский 2012]. На илл. 10 даны графики и полные списки ассоциаций американцев и русских на слова-стимулы PREFIX [Norms 1970] и ПРЕФИКС [Долинский 2012].

В последние десятилетия исследование количественных характеристик языковых единиц и структур привлекло внимание представителей *синергетической лингвистики* (G. Altmann, R. Köhler, Р. Г. Пиотровский). Синергетический подход опирается на идеи «сил» диверсификации и унификации, коммуникативные «потребности» носителей языка и ориентирован на изучение языковых процессов и языковой саморегуляции [Пиотровский 2006; Altmann 1992; 1997; Köhler, Altmann, Piotrowski 2005]. Р. Г. Пиотровский отмечал:

> Своеобразие общей методологии синергетики состоит также в том, что она пытается описать диалектику конфликта между неподвижным бытием и становлением, статичным законом и динамичной креативностью, а что касается творческих личностей, то борьбу рутинных правил и новаторства. <…> Если обратиться к системной лингвистике, использующей теоретико-множественный и логико-математический аппарат, то она ориентирована на описание статического (синхронного) строения языка, но не на его динамику (диахронию). И всё же, несмотря на все своеобразие путей развития современной науки о языке, идеи синергетики в связке с концепциями саморегуляции систем и теории катастроф стали проникать в лингвистический обиход [Пиотровский 2006, с. 8–9].

Согласно модели синергетической лингвистики, в сфере семантики слушающий старается исключить полисемию, не допуская существования у каждого слова более, чем одного значения. Подобная мера дает ему возможность получать сообщения с наименьшими усилиями (least efforts) по их декодированию. Говорящий, напротив, старается придать каждому слову максимум значений, что потребует наименьших усилий с его стороны по кодированию сообщений. Победа в коммуникации говорящего привела бы к унификации: исключению всех слов, кроме одного, передающего все значения. Победа слушающего привела бы к диверсификации: превращению всех слов в однозначные. Эмпирический результат этих противоположных тенденций – компромисс, направленный к определенному равновесному вероятностному распределению значений. По отношению к ассоциациям сло́ва тенденция та же, что и со значениями: слушающий стремится игнорировать скрытые (неявные, имплицитные) ассоциации слова, непривычные смыслы; говорящий постоянно порождает новые ассоциации на слово, новые коннотации. Их число и частота могут различаться для каждого слова. Это означает, что семантика слова постоянно подвергается процессам диверсификации и унификации. Результатом этого будет компромисс между коммуникативными потребностями (communication needs) говорящего и слушающего (эти роли попеременно играет каждый коммуникант), компромисс, порождающий асимметричное распределение, в котором одна или несколько ассоциаций (значений) появляются очень часто, а остальные многочисленные ассоциации образуют периферию частотного распределения.

Синергетическая метафора построена на принципе дискретности. Однако в акте коммуникации можно наблюдать не только борьбу

говорящего и слушающего за «победу», но и стремление к универсальной «ранговой шкале», общему смыслу, требующему не войны «интересов» и «коммуникативных потребностей», а сотрудничества и компромисса.

Корпусный подход в лингвистике основан на анализе больших баз данных языковых массивов, хранящихся на компьютере. Во всех случаях, когда мы имеем дело с корпусными исследованиями (текстами естественного языка, данными психолингвистических экспериментов и т. п.), проявляется так называемый эффект *концентрации и рассеяния*, который состоит в том, что имеется небольшая группа очень частых единиц («ядро», или «голова» рангового распределения), и большая группа редких единиц («периферия», или «хвост», распределения).

Для аналитического выражения зависимости между частотой и рангом слова (квазигипербола), предлагалось множество формул, которые представляют собой разновидности классического «закона Ципфа»: самая простая из них ($F_i \cdot i = const$) не всегда надежно аппроксимирует эмпирические частоты, а самая изощренная – страдает избытком плохо интерпретируемых параметров и отсутствием универсальности применения. Неоднократно отмечалось, что распределение имеет слишком длинный «хвост» и, как результат, слишком медленную сходимость из-за бесконечной дисперсии, а также «патологическую» концентрацию в начале и резкую асимметрию, что позволяет сделать вывод о негауссовости данных распределений.

Известный лингвостатистик Густав Хердан писал:

> «Математики верят в него [закон Ципфа], потому что лингвисты постановили считать его лингвистическим законом, а лингвисты со своей стороны верят в него потому, что математики постановили считать его математическим законом» [Herdan 1966].

Попыткам аналитического выражения данной функции посвящен ряд работ, так или иначе модифицирующих формулы Ципфа.

Для ранговых распределений ассоциатов, полученных в эксперименте, наилучшие результаты дает формула Ципфа-Долинского [Долинский 1988; 2012; Altmann 1992]:

$$\mathbf{F_i = F_1 \cdot i^{-(a + b \cdot \ln i)}}$$

или после логарифмирования:

$$\mathbf{\ln F_i = \ln F_1 - (a + b \cdot \ln i) \ln i}$$

где a и b – параметры.

Язык служит как раскрытию и сохранению смыслов, так и их сокрытию и утрате. Противоречивый «скрывающе-раскрывающий» статус слова по отношению к смыслам определяет квантитативную структуру ассоциативных полей – совмещение гомогенности и гетерогенности распределений. Эффект концентрации и рассеяния, характерный для частотной структуры текстов, имеет своим источником, по-видимому, такой *механизм порождения текста, который обусловлен ассоциативной природой языкового сознания*.

С одной стороны, одна или несколько особо прочных ассоциаций играют роль социальных «сторожей» и «ограничителей» смыслов; с другой – множество уникальных непрочных ассоциаций выступают в роли индивидуальных «первопроходцев» и «расшатывателей» смыслов. Фактами языка являются и «голова», и «хвост» распределения, независимо от их употребительности и отношения к «норме».

Первая тенденция *сохранения смыслов* проявляется в гомогенности распределения ассоциаций, сужении их поля, доминировании принципа наименьших усилий; она служит сохранению языка от естественных шумов, разрушающих его строй. Вторая тенденция *раскрытия смыслов* проявляется в гетерогенности распределения, расширении поля ассоциирования, доминировании принципа наибольших усилий; она служит сохранению языка от преднамеренных попыток людей выхолостить его смысл.

В сфере *социального функционирования* языки также распределены крайне неравномерно – от нескольких самых распространенных до сотен находящихся на грани вымирания [Потапов 1997]. Из 6–7 тыс. языков мира (социально признанных – от 3 до 4 тыс.) только 23 языка (от 50 млн носителей каждый) являются родными для 4,1 млрд человек (из 7,2 млрд жителей Земли). 83 % языков распространены в пределах одной страны. 90 % языков имеют менее 10 тыс носителей каждый. По прогнозам, к концу XXI в. исчезнет от 65 до 90 % языков мира.

На илл. 11 дано многомерное ранговое распределение языков по числу стран, в которых на них говорят, в сопоставлении с числом носителей, числом изучающих, числом интернет-пользователей и долей вебсайтов.

Асимметричность рангового распределения языков приводит (в диахронии) стихийно развивающиеся языки к форме литературных языков, характеризующихся социальной регламентированностью и нормированностью, ведущей к консервации языкового состояния. Общая лингвистическая теория, в которой по-прежнему нет места для «малых» языков, до сих пор страдает крайним «англо- и европоцентризмом» [Кибрик 2011].

На илл. 12 представлено распределение литературных языков по числу лауреатов Нобелевской премии по литературе [Лауреаты 2016]. В 1901–2016 гг. премия вручалась 109 раз (в 1904, 1917, 1966, 1974 по 2 лауреата). Премия не вручалась в 1914, 1918, 1935, 1940–1943 гг. 113 Нобелевских лауреатов по литературе писали на 25 языках: Рабиндранат Тагор (Нобелевская премия по литературе 1913) писал на бенгальском и английском языке, Сэмюэль Беккет (1969) писал на французском и английском языке, Иосиф Бродский (1987) писал поэзию на русском языке и прозу на английском языке. Эти три лауреата были маркированы бенгальским, французским и русским языком, соответственно.

Альфред Нобель в своем завещании отклонял любые соображения о национальности кандидатов: должен быть выбран самый достойный, «скандинав он или нет». Проблема представленности языков всего мира стала ведущей, и в течение долгого времени шведская Академия, выбиравшая Нобелевских лауреатов, справедливо подвергалась критике за то, что она делала присуждение награды делом Европы. В 1984 г. постоянный секретарь Академии объявил, что внимание к неевропейским авторам «постепенно возрастало», и были сделаны усилия, чтобы «достичь глобального распределения».

Экспериментальное изучение ранговых распределений методами, разрабатываемыми в рамках квантитативной лингвистики, позволяет выявить многообразные связи качественного и количественного характера между элементами языковой системы, ее многоуровневое иерархическое устройство, пружины ее функционирования и эволюции. Многие аспекты языка, связанные с количественными параметрами распределений его единиц, их ассоциативным потенциалом, не позволяют нам теперь разделить убеждение Н. С. Трубецкого в том, что «язык лежит вне меры и числа».

Приведем ряд высказываний лингвистов и математиков.

Г. Хердан: «Стабильность относительных частот символов является общей характеристикой лингвистических форм <…>. Это выражение

того факта, что даже здесь, где человеческая воля и выбор имеют широчайшую сферу, где сознательно избираемый замысел (intention) и бесконечная игривость быстро сменяют друг друга, – даже здесь выявляется в большем целом регулярность главного значения» [Herdan 1966, p. 36].

Ю. К. Орлов: «Ни один биолог никогда не исследовал фарша из множества лягушек, чтобы установить "среднелягушечные" характеристики. Ни один филолог не отнесется всерьез к фразе: "Все счастливые семьи похожи … на перекладных из Тифлиса". А ведь это – модель репрезентативной выборки… Нужно отказаться от привычной трактовки частоты как оценки вероятности, так как в условиях недостаточных выборок эта трактовка приводит к слишком большим ошибкам» [Орлов 1980, с. 39].

В. В. Налимов: «Усредненные характеристики здесь, в отличие от физики, не имеют значения. Важны отдельные явления в своем индивидуальном проявлении вне зависимости от того, какова вероятность их появления <…>. К нашей концепции, может быть, ближе стоит невероятностный подход к построению количественной семантики расплывчатого смысла слов» [Налимов 2003, с. 96].

Ю. В. Чайковский: «Вероятностный язык здесь заведомо неприемлем, но если все-таки пользоваться языком частот, то придется признать модели, в которых дисперсии неограниченны, наиболее удобными <…>. Дальнейшее продвижение в статистической лингвистике будет, как мне видится, достигнуто после решения более общего вопроса – почему столь общ феномен квазигипербол» [Чайковский 2001, с. 207].

Исследования в квантитативной лингвистике – статистике речи и текста, словаря и ассоциативных полей, стилеметрии и лингвогеографии – позволяют по-новому оценить природу важнейших системных закономерностей, носящих существенно «недискретный», «нежесткий» характер и плохо поддающихся описанию на неадекватном языке «алгебраической» лингвистики. В то же время методы теории вероятностей и математической статистики, применяемые в лингвистике, имеют свои границы. Распределения частот в текстах и в ассоциативных полях, полисемия, стилеметрия, языковые различия и многие другие объекты изучения как в лингвистике, так и в биологии, экономике, социологии, технетике, – слишком сложны и нестабильны для построения их моделей, исходя из удаленных от эксперимента общих принципов, характерных для фундаментальной физики. Следует настороженно относиться к привлечению сложной математики, отвлеченным построениям, легковесному теоретизированию. Здесь,

по-видимому, более уместны феноменологические модели, опирающиеся на реально измеряемые характеристики, чем опора на принцип максимума статистической энтропии.

> Так, в настоящее время эксперимент позволяет ручаться за относительную точность выполнения закона сохранения энергии в ядерных реакциях, равную 10–6, тогда как законы Менделя выполняются в эксперименте гораздо менее точно.

В литературе встречаются настоятельные рекомендации измерять функции распределения на том основании, что стабилизация функций $F_m(x)$ с ростом m всегда более заметна, нежели стабилизация гистограмм – эмпирических плотностей распределения. Между тем это достоинство, по мнению Ю. И. Алимова, «<подобно> преимуществу нечувствительного прибора, дающего малое рассеяние результатов измерений, перед прибором чувствительным, у которого рассеяние в тех же условиях оказывается значительно большим, как раз из-за чувствительности» [Алимов 1980, с. 24].

Следует отметить, что многие ранговые распределения не являются *вероятностными* в принятом смысле слова [Налимов 2003; Чайковский 2001]. Для ассоциативных полей слов, как и для квазитекста неограниченной длины, неопределенным остается понятие «генеральная совокупность», и дело здесь не только в бесконечной дисперсии. Однородность культурноязыкового сообщества и однородность ассоциативного поля находятся в тесной взаимосвязи и взаимозависимости. К неформальной должна быть отнесена и проблема исходных посылок при определении представительности выборок. С ростом выборки доля вновь появляющихся слов не обнаруживает заметного падения. Вспомним афоризм: статистике часто принадлежит первое слово, но последнее – никогда.

Вот что писал об этом Джордж Кингсли Ципф:

> Демонстрируя наши формулы, мы не должны считать вопрос решенным статистически; наши формулы – не «статистические артефакты», в обычном понимании этого термина. Кроме того, имея в виду число объектов, к которым формулы применялись эмпирически, нельзя сказать, что соответствия – всего лишь дело случая, как обыкновенно интерпретируется слово «случай»; на самом деле, то, что мы обнаружили такое близкое соответствие нашим формулам, проявившееся даже в двух произвольных языковых выборках, *чрезвычайно противоположно вероятностям* [Zipf 1938, с. 360]. (Курсив наш. – *В. Д.*).

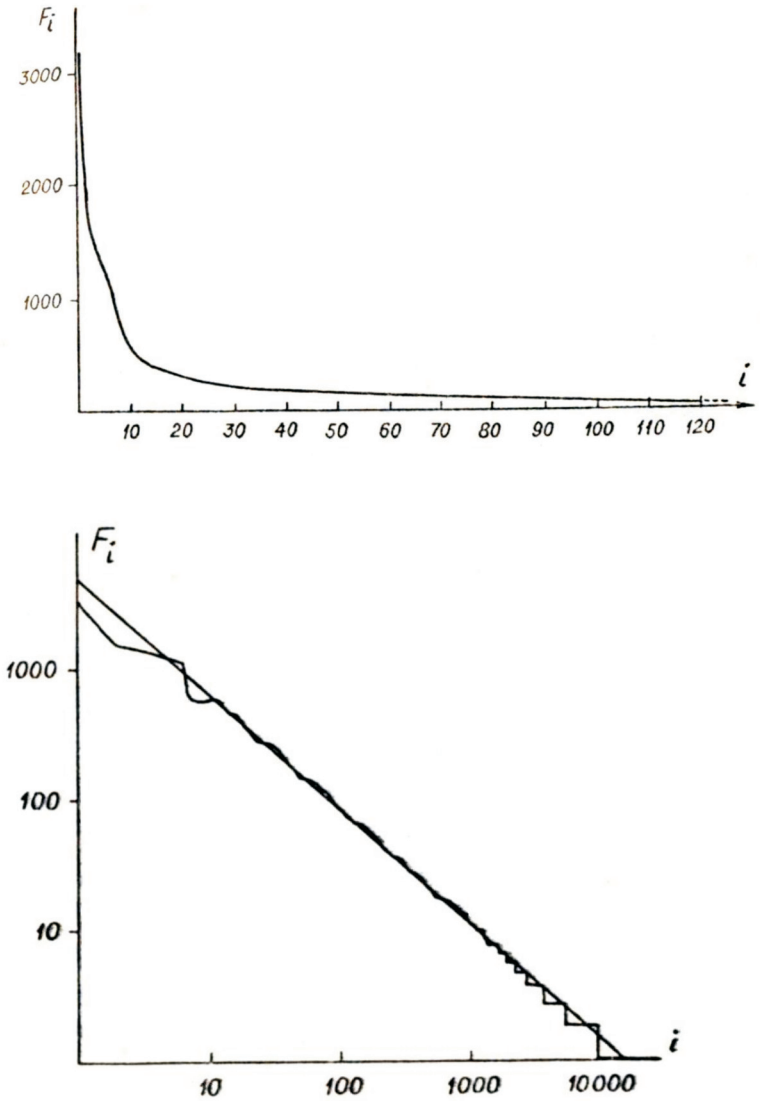
Основатель кибернетики Норберт Винер подчеркивал:

> Как ни труден отбор надежных данных в физике, гораздо сложнее собрать обширную информацию социологического характера, состоящую из многочисленных серий однородных данных. <…> В этих обстоятельствах безнадежно добиваться слишком точных определений величин, вступающих в игру. Приписывать таким неопределенным по самой своей сути величинам какую-то особую точность бесполезно, и, каков бы ни был предлог, применение точных формул к этим слишком вольно определяемым величинам есть не что иное, как обман и пустая трата времени [Винер 2003, с. 220–221].

Лингвистические модели являются наиболее «узким местом» в информационных технологиях, что связано с противоречием, существующим между смысловым характером информации, выражаемой средствами естественного языка, и необходимостью формальной ее обработки с помощью компьютерных программ. Преодоление этого противоречия, или, скорее, его «смягчение», видится, в частности, в разработке языковых моделей, развиваемых в русле лингвистического ассоцианизма, синергетики и, в целом, квантитативной лингвистики.

Илл. 1

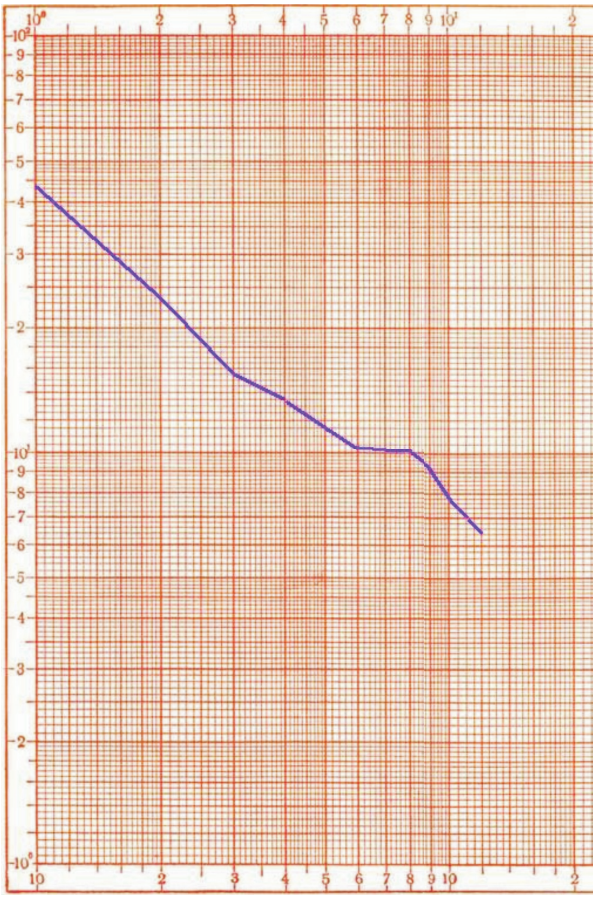
Связь между частотой F_i и рангом слова *i*
(а – обычный масштаб; б – билогарифмическая система координат)
[Тулдава 1987]



Илл. 2

Частотный словарь Нового Завета (Четвероевангелие)
Фрагмент частотного списка [Алексеев 2001]

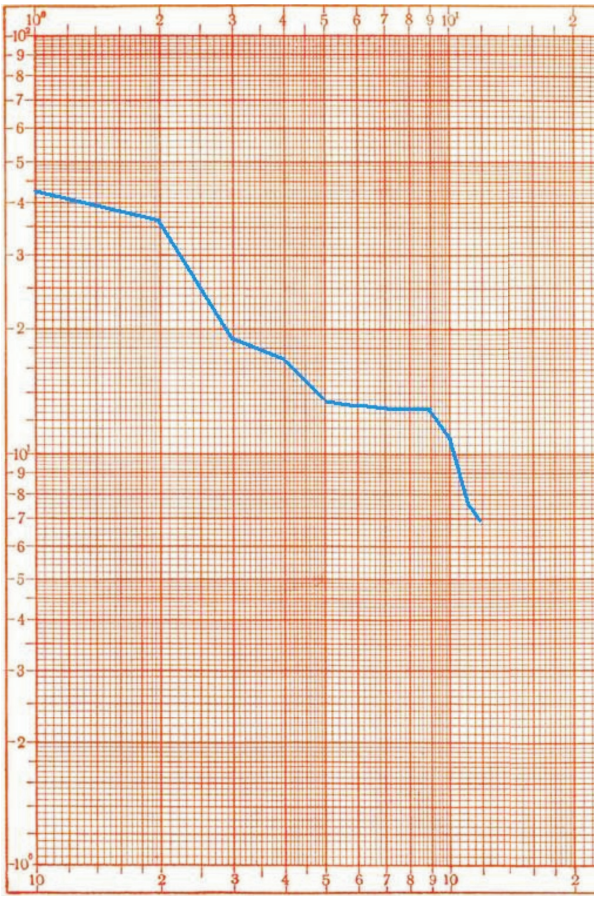
Ранг	Слово	Частота	Ранг	Слово	Частота
1	*и*	4400	7	*я (Я)*	1046
2	*он (Он)*	2352	8	*сказать*	1023
3	*в (во)*	1591	9	*вы*	918
4	*не*	1383	10	*говорить*	776
5	*они*	1151	11	*ты (Ты)*	706
6	*быть*	1049	12	*Иисус*	649



Илл. 3

Частотный словарь русского языка
Фрагмент частотного списка [Частотный 1977]

Ранг	Слово	Частота	Ранг	Слово	Частота
1	*в (во)*	42854	7	*что*	13185
2	*и)*	36266	8	*он*	13143
3	*не*	19228	9	*с (со)*	12975
4	*на*	17262	10	*а*	10719
5	*я*	13839	11	*как*	7425
6	*быть*	13307	12	*это*	6940



Илл. 4

Ранг слова
по данным частотного словаря русского языка
в четырех подмассивах и в корпусе в целом [Частотный 1977]

РАНГ \ ЛЕКСЕМА	*в*	*и*	*не*	*на*	*я*	*быть*
Словарь в целом	**1**	**2**	**3**	**4**	**5**	**6**
По подвыборкам:						
Художественные прозведения	2	1	4	3	5	6
Драма	5	2	1	4	6	3
Научно-публицистические	2	1	4	3	6	5
Газетно-журнальные	1	2	4	3	5	6

Илл. 5

Полисемичность слова
по данным толковых словарей русского языка:
в среднем и в четырех словарях

ПОЛИСЕМИЯ \ ЛЕКСЕМА	*в*	*и*	*не*	*на*	*я*	*быть*
Словарь в целом	**18,25**	**9**	**6,5**	**22,75**	**3**	**5,75**
По словарям:						
ССРЛЯ	**27**	**16**	**9**	**35**	**6**	**6**
МАС	**23**	**13**	**8**	**34**	**2**	**7**
ТСРЯ	**12**	**3**	**7**	**12**	**2**	**6**
СО	**11**	**4**	**2**	**10**	**2**	**4**

ССРЛЯ – Словарь современного русского литературного языка: в 17 т. (1950–1965); 120 тыс. слов
МАС – Словарь русского языка: в 4-х т. (1981–1984); 90 тыс. слов
ТСРЯ – Толковый словарь русского языка под ред. Н. Ю. Шведовой (2007); 82 тыс. слов
СО – Словарь русского языка С. И. Ожегова (14 изд., 1983); 57 тыс. слов

Илл. 6

Распределение лексем по числу значений
по данным толковых и авторских словарей русского языка
[Поликарпов 1987]

Полисемия	Число лексем			
	ССРЛЯ	**МАС**	**СО**	**СЯП**
Всего	**120481**	**82017**	**56998**	**20196**
1	76382	59920	44166	16494
2	26102	13236	8390	3164
3	9316	4680	2546	911
4	3989	2060	960	336
5	1928	996	426	154
6	1057	491	231	55
7	587	282	116	41
8	342	191	42	16
9	222	97	51	8
10	151	79	30	5
11	118	40	14	5
12	83	25	13	3
13	37	17	1	–
14	38	11	6	–
15	35	9	3	2
16	22	11	2	–
17	7	3	–	2
18	12	2	3	–
19	14	1	–	–
20	14	2	1	1
21	9	–	–	–
22	4	–	–	–
23	1	1	–	–
24	2	1	–	–
25	1	3	–	2
26	3	–	2	–
27	1	–	–	1
28	1	–	-	–
29	–	1	–	1
30	1	–	–	1
31	1	–	–	–
32	–	–	–	–
33	1	–	–	–

ССРЛЯ – Словарь современного русского литерат. языка: в 17 т. (1950–1965)
МАС – Словарь русского языка: в 4-х т. (1957–1961)
СО – Словарь русского языка С. И. Ожегова (9 изд., 1972)
СЯП – Словарь языка А. С. Пушкина (1956–1961)

Илл. 7

Распределение персонажей романа М. А. Булгакова «Мастер и Маргарита» по частоте их упоминаний [Чайковский 2001]

Количество персонажей	201	90	34	19	9	4	3	1	3	4	4	3	1	2	3	1	1	2	2
Количество упоминаний	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	более 19

Илл. 8

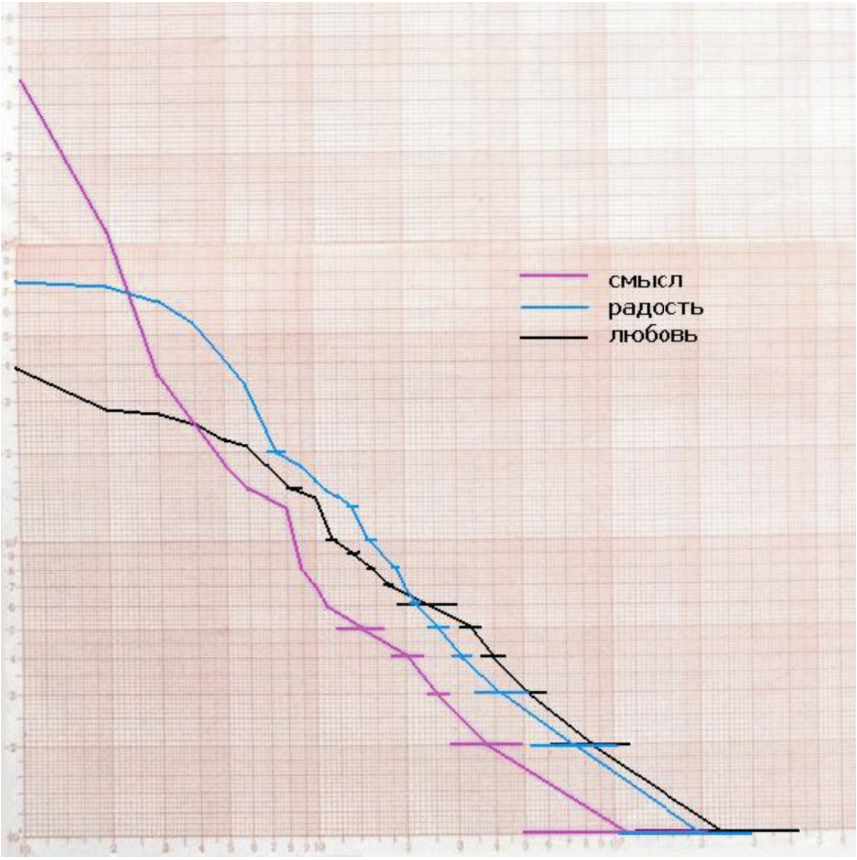
Распределение лексем с семантикой цветообозначения в текстах Н. В. Гоголя (цветовой спектр, в %) [Белый 1934]

Цвет	Средняя	Группа произведений			
		1	2	3	4
Красное	**17,4**	26, 8	12, 5	10, 3	6, 4
Белое	**14,0**	9, 5	9, 0	22, 0	17, 0
Чёрное	**12,0**	11, 0	14, 1	11, 8	4, 8
Зелёное	**9,4**	8, 6	7, 7	9, 6	21, 6
Золотое	**9,2**	11, 6	8, 9	2, 8	12, 8
Синее	**8,7**	10, 7	6, 1	4, 9	6, 4
Жёлтое	**7,0**	3, 5	8, 5	10, 3	12, 8
Серое	**5,8**	2, 6	8, 9	10, 5	6, 4
Голубое	**4,8**	4, 4	5, 7	7, 0	1, 6
Серебряное	**4,8**	7, 1	3, 2	2, 8	4, 8
Коричневое	**4,0**	0, 9	6, 5	8, 4	1, 6
Розовое	**2,3**	3, 8	0, 8	2, 1	3, 2
Оранжевое	**1,2**	0, 3	1, 6	2, 8	0, 0
Лиловое	**0,9**	1, 7	1, 2	0, 0	1, 6

Обозначения групп произведений:
1. «Вечера на хуторе близ Диканьки», «Вий», «Тарас Бульба»
2. Повести и комедии
3. «Мёртвые души», первый том
4. «Мёртвые души», второй том

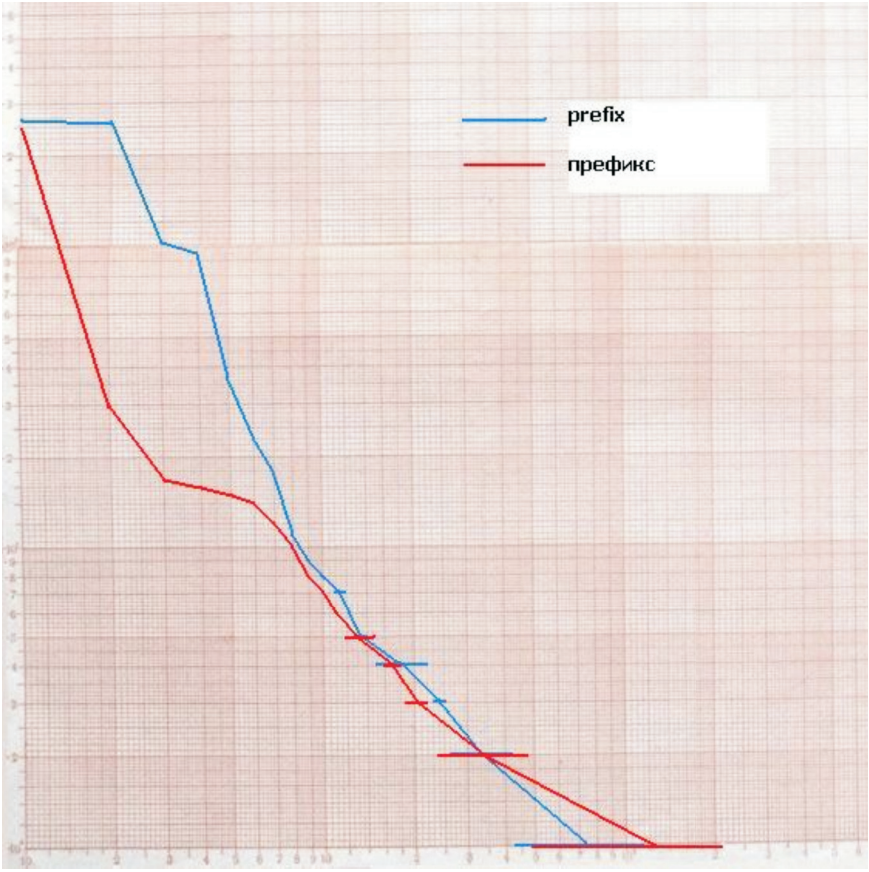
Илл. 9

Частотно-ранговые распределения ассоциативных полей существительных русского языка: СМЫСЛ, РАДОСТЬ, ЛЮБОВЬ
(билогарифмическая система координат)



Илл. 10

Частотно-ранговые распределения ассоциативных полей слов-стимулов PREFIX и ПРЕФИКС
(Языки: американский английский [Norms 1970]; русский [Долинский 2012]; билогарифмическая система координат)



PREFIX (N = 1000; F1 = 264; L = 114; m1 = 72)

suffix 264	adjective 4	after 2
before 260	end 4	ante 2
word(s) 105	index 4	ex 2
beginning 96	pre 4	in 2
first 34	preceding 4	meaning 2
English 23	re 4	name 2
book(s) 18	start 4	noun 2
grammar 11	telephone 4	part 2
begin 9	number 3	precede 2
letter (s) 8	sub 3	reflex 2
ending 7	verb 3	sentence 2
front 7	ad 2	stem 2
spelling 5	addition 2	Subject A 2
syllable (s) 5	affix 2	un 2

43–114 (F=1)
add, additive, adverb, ahead, antifix, are, beginning of, break, came before, change, comes before, coming before, complex, con, dash, de , definition, dictionary, dis, dis/appear, dog, dry, en, first, word, fixe, for, forehand, French, German, glass, head, im, in front of, ing, interesting, into, library, locate, longer, "me"tcalf, modify, mutilation, part of word, preamble, prearrange, precedence, prediction, preface, prefix, preplex, previous, pro, pronominal prefix, rain, ramrod, reading* renew, short, sink, stop, subject, surname, syllable before, telephone number, the, three, verbal, vowel, wood, working, wrench)

ПРЕФИКС (N = 1010; F1 = 246; L = 212; m1 = 164)

суффикс — 246
приставка — 29
аффикс — 17
слово — 16
дефис — 15
русский язык — 14
грамматика — 12
постфикс — 10
язык — 8
школа — 7
часть слова — 6
английский — 5
английский язык — 5
префект — 5
преферанс — 5
глагол — 4
окончание — 4
часть — 4
икс — 3
инфикс — 3
корень — 3
начало — 3
флексия — 3
знак — 2
идея-фикс — 2
иностранный язык — 2
кавычки — 2
компьютер — 2
лингвистика — 2
морфология — 2
немецкий — 2
офис — 2
полиция — 2
сегмент — 2
словообразование — 2
сложно — 2
существительное — 2
сфинкс — 2
тире — 2
урок — 2
учебник русского языка — 2
учитель — 2
фикция — 2
филология — 2
часть речи — 2
чёрточка — 2
экзамены — 2
языкознание — 2
Аникина — 1
Суздальцева — 1
Фикс — 1
Фукс — 1
анекдот — 1
апостериори — 1
апостроф — 1
без ассоциаций — 1
болезнь — 1
бугор — 1
буквы — 1
взлёт — 1
власть — 1
во- — 1
выделение, зелёной ручкой — 1
выключатель — 1
выравнивание — 1
выступающий — 1
грамматический разбор — 1
гримаса — 1
да — 1
деталь — 1
детство — 1
дефект — 1
дефикс (?) — 1
длинный — 1
длинный нос — 1
добавление — 1
дом — 1
дурацкий — 1
дурость — 1
дыра — 1
ерунда — 1
загадка — 1
загадочный — 1
запятая — 1
знание — 1
значение — 1
значок — 1
золото — 1
игра — 1
издание — 1
инспектор — 1
интересно — 1
интересно узнать — 1
какаду — 1
карты — 1
квадрат — 1
кекс — 1
кино — 1
книга — 1
код — 1
колесо — 1
континиус — 1
коронка — 1
круто — 1
крыша — 1
ксерокс — 1
кс-кс-кс — 1
лагерь — 1
ластик — 1
латынь — 1
лексикология — 1
линия — 1
литература — 1
логика — 1
маленький — 1
маска — 1
мафия — 1
мистер — 1
мм! — 1
морфема — 1
морф-анализ — 1
мура — 1
мэр — 1
написан — 1
на- — 1
недоделанный — 1
неизвестно — 1
немецкий язык — 1
неповоротливый — 1
нечто — 1
не въехал — 1
не знаю — 1
ничего — 1
нудистика (!) — 1
облако — 1
округ — 1
орфография — 1
остриё — 1
отделяемый — 1
падеж — 1
папка — 1
пенис — 1
перед — 1

перекись — 1
перо — 1
перфект — 1
печать — 1
правильный — 1
предлог — 1
предложение — 1
префектура — 1
префект и суффикс — 1
префикс — 1
пре- — 1
принтер — 1
при- — 1
программирование — 1
профессия — 1
раз- — 1
рамка — 1
расстрел — 1
ребёнок — 1
речь — 1
рога — 1
розовый — 1
русский — 1
ручей — 1
серьёзно — 1
сила! — 1
синтагма — 1
синтаксис — 1
сифилис — 1
скиферп (!) — 1
скобка — 1
слово, напр., «при-шёл» — 1
собака — 1
странный — 1
строгость — 1
супрефикс — 1
суффикс, флексия — 1
телефак (!) — 1
теория — 1
тоска — 1
точки — 1
тумак — 1
тюфикс (!) — 1
угол — 1
уголок — 1
ужасный — 1
умно — 1
унитаз — 1
уравнение — 1
учебник — 1
ученик — 1
учить — 1
у- — 1
фактура — 1
фиксатор — 1
фикса (!) — 1
фикс — 1
фикус — 1
фи-фи — 1
циркумфлекс — 1
частица — 1
что — 1
чухня (!) — 1
широкий — 1
экзамен — 1
языкознание (МГУ) — 1
«2» по русскому языку — 1
Deutsch — 1
I don't know — 1

Илл. 11

Распределение языков
по числу стран, в которых на них говорят,
в сопоставлении с числом носителей, числом изучающих,
числом интернет-пользователей и долей веб-сайтов

[Ethnologue-Languages of the World, CM-US, Unesco, United Nations, University of Düsseldorf, Internetworldstats, Foundation for Endangered Languages]

Ранг	Язык	Число стран	Кол-во носителей (млн)	Число изучающих (млн)	Число интернет-пользователей по языкам (млн)	Доля веб-сайтов (%)
1	Английский	110	341 (4)	1500 (1)	948,6 (1)	54,8 (1)
2	Арабский	60	412 (2)	2,5	168,4	1,2
3	Французский	51	78 (12)	82 (2)	102,1	4,4
4	Китайский	33	1212 (1)	30 (3)	751,9 (2)	4,5
5	Испанский	31	322 (5)	14,5 (5)	277,1 (3)	4,8
6	Персидский	29	31			0,1
7	Немецкий	18	100 (10)	15 (4)	83,8	5,8 (3)
8	Русский	16	167 (8)	1	103,1	6,0 (2)
9	Малайский	13	36		109,4	0,3
10	Португальский	12	176 (7)	2	154,5	2,3
11	Итальянский	11	62 (18)	8 (6)		1,5
12	Украинский	9	47			0,1
13	Польский	9	44			1,1
14	Турецкий	8	61 (19)			1,4
15	Нидерландский	6	22			1,8
16	Тамильский	6	66 (17)			0,1
17	Урду	6	60 (20)			0,1
18	Корейский	5	78 (11)			0,3
19	Хинди	4	366 (3)			0,1
20	Бенгали	4	207 (6)			0,1
21	Яванский	3	75 (13)			
22	Вьетнамский	3	68 (16)			
23	Японский	2	125 (9)	3 (7)	115,1	5,7
24	Телугу	2	69 (14)			
25	Маратхи	1	68 (15)			
26	Панджаби	1	57			
27	Гуджарати	1	46			
28	Тайский	1	46			
…	…	,,,	,,,	…	…	…

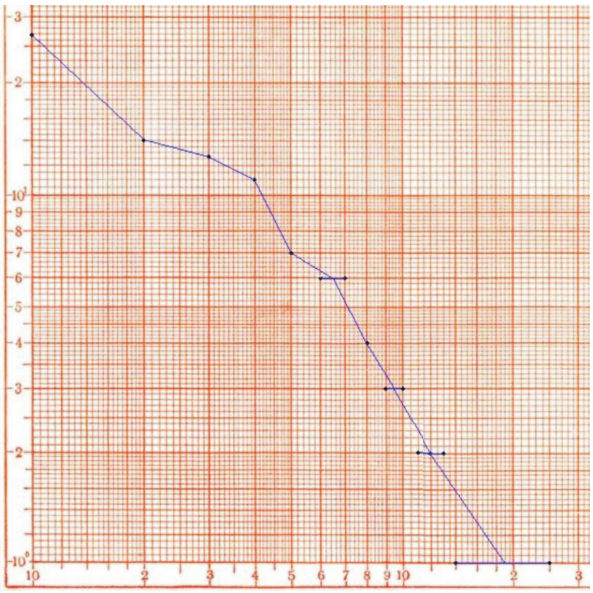
В скобках указаны ранги для распределений по каждому из столбцов таблицы

Илл. 12

Распределение литературных языков
по числу лауреатов Нобелевской премии по литературе
[Лауреаты 2016]

Лауреаты Нобелевской премии по литературе
(113 чел. с 1901 по 2016 гг.)
писали на следующих 25 языках:

Ранг (X)	Язык	Число лауреатов (Y)	Ранг (X)	Язык	Число лауреатов (Y)
1	Английский язык	28	14	Арабский	1
2	Французский	14	15	Бенгальский	1
3	Немецкий	13	16	Венгерский	1
4	Испанский	11	17	Иврит	1
5	Шведский	7	18	Идиш	1
6	Итальянский	6	19	Исландский	1
7	Русский	6	20	Окситанский	1
8	Польский	4	21	Португальский	1
9	Норвежский	3	22	Сербско-хорватский	1
10	Датский	3	23	Турецкий	1
11	Греческий	2	24	Финский	1
12	Китайский	2	25	Чешский	1
13	Японский	2			1



СПИСОК ЛИТЕРАТУРЫ

Алексеев П. М. Частотные словари. СПб. : Изд-во С.-Петерб. ун-та, 2001. 156 с.

Алимов Ю. И. Альтернатива методу математической статистики. М. : Знание, 1980. 64 с.

Аллахвердов В. М. Методологическое путешествие по океану бессознательного к таинственному острову сознания. СПб. : Речь, 2003. 368 с.

Арапов М. В. Квантитативная лингвистика. М. : Наука, 1988. 184 с.

Белый Андрей. Мастерство Гоголя. М.–Л., 1934. 320 с.

Винер Н. Кибернетика и общество. Творец и робот / пер. с англ. М. : Тайдекс Ко., 2003. 248 с.

Долинский В. А. Распределение реакций в экспериментах по вербальным ассоциациям // Уч. записки Тартуского университета. Вып. 827. Тарту, 1988. С. 89–101.

Долинский В. А. Теория ассоциативных полей в квантитативной лингвистике. М. : Тезаурус, 2012. 512 с.

Кибрик А. Е. База естественного человеческого языка и ее основные параметры: доклад на конф. «Диалог-2011». URL: www.dialog-21.ru/digests/dialog2011/

Лауреаты Нобелевской премии по литературе. URL: www.nobelprize.org/nobel_prizes/literature/laureates/index.html/

Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка. М., 2009. 1152 с. Ljashevskaja O. N., Sharov S. A. Chastotnyj slovar' sovremennogo russkogo jazyka. M., 2009. 1152 p.

Мартыненко Г. Я. Введение в теорию числовой гармонии текста. СПб. : Изд-во СПб. ун-та, 2009. 252 с.

Налимов В. В. Вероятностная модель языка. О соотнесенности естественных и искусственных языков. 3-е изд. Томск–М. : Водолей Publishers, 2003. 368 с.

Орлов Ю. К. Невидимая гармония // Число и мысль. Вып. 3. М. : Знание, 1980. С. 70–106.

Пиотровский Р. Г. Лингвистическая синергетика: исходные положения, первые результаты, перспективы. СПб. : Филол. фак-т СПбГУ, 2006. 160 с.

Поликарпов А. А. Полисемия: системно-квантитативные аспекты // Уч. зап. Тартуского университета, вып. 774. Тарту, 1987. С. 135–154.

Потапов В. В. К современному состоянию проблемы вымирающих языков в некоторых регионах мира // Вопросы языкознания. 1997. № 5. С. 3–15.

Тулдава Ю. А. Проблемы и методы квантитативно-системного исследования лексики. Таллин : Валгус, 1987. 204 с.

Чайковский Ю. В. О природе случайности. М. : Центр системных исследований. Институт истории естествознания и техники РАН, 2001. 272 с.

Частотный словарь русского языка / под ред. Л. Н. Засориной. М. : Русский язык, 1977. 936 с.

Altmann G. Two models for word association data // B. Rieger (Ed.), Glottometrika 13. Bochum: Brockmeyer, 1992. P. 105–120.

Altmann G. The art of Quantitative Linguistics // Journal of Quantitative Linguistics, 1997. Vol. 4. No 1–3. P. 13–22.

Herdan G. The Advanced Theory of Language as Choice and Chance. Berlin–Heidelberg–N. Y. : Springer Verlag, 1966. 459 p.

Köhler R., Altmann G., Piotrowski R. G. (Hrsg.): Quantitative Linguistik – Quantitative Linguistics. Ein internationales Handbuch. Berlin–New York : de Gruyter, 2005. 1041 p.

Norms of Word Association / Eds. L. Postman and G. Keppel. N. Y. & L., 1970. 943 p.

Zipf G. K. Homogeneity and heterogeneity in language // The psychological records. 2. 1938. P. 347–367.

Zipf G. K. Human behavior and the principle of least effort. Cambridge (Mass.): Addison-Wesley, 1949. 573 p.