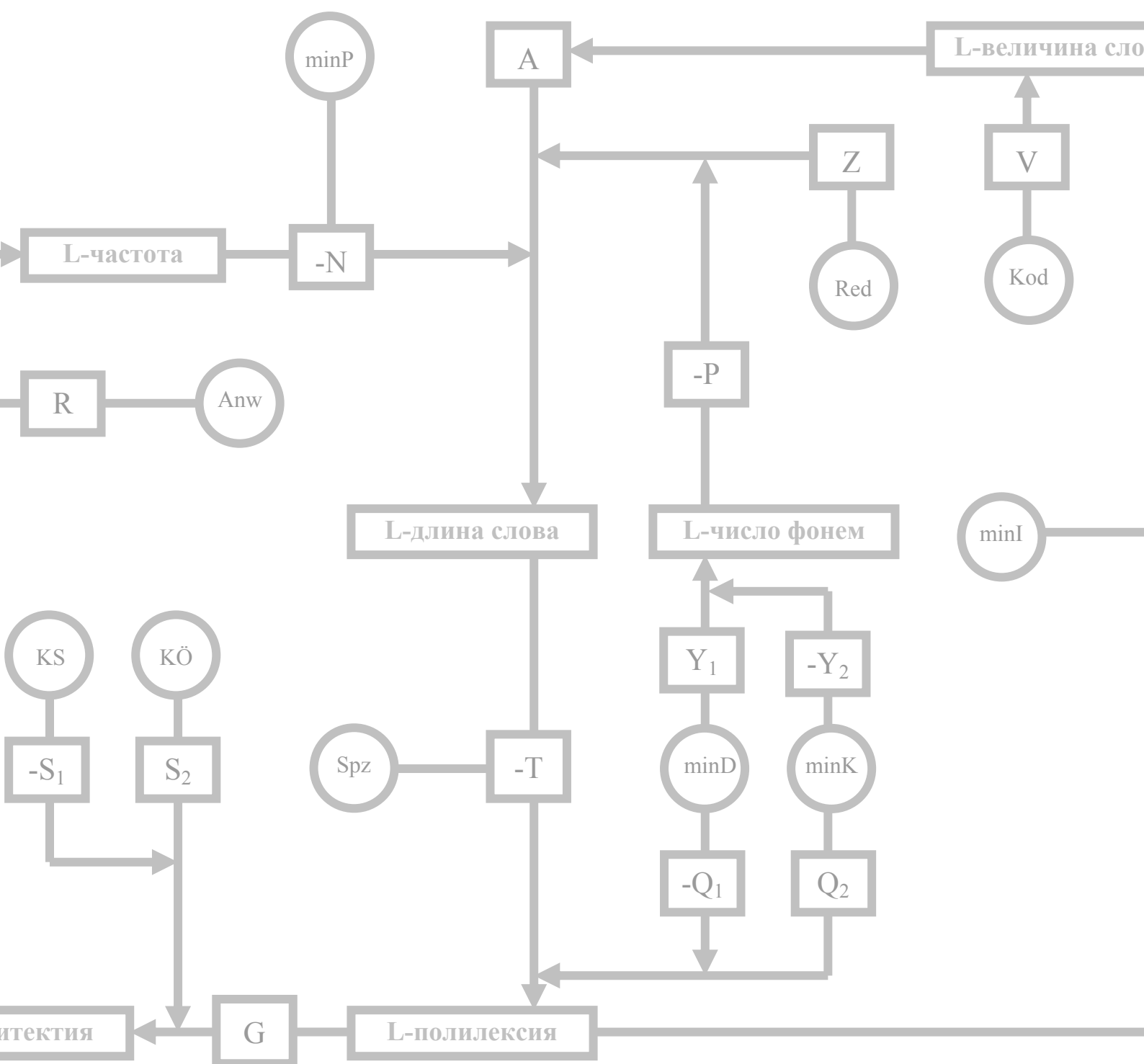


Райнхард Келер

Синергетическая лингвистика: структура и динамика лексики



ВВЕДЕНИЕ

Квантитативная лингвистика совместно с другими лингвистическими дисциплинами участвует в решении задачи построения теории языка. Специфика квантитативной лингвистики заключается в этом отношении не в целях познания, а в первую очередь в используемых для этого средствах и методах. Математическая лингвистика возникла в результате разработки категориальных (качественных) понятий и формулирования связей между ними. В соответствии с их назначением в языкознание были введены в первую очередь качественные разделы математики: логика, теория множеств, теория групп, теория автоматов, формальные грамматики, линейная алгебра, решетчато-упорядоченные множества, теория графов, топология и геометрия. В результате этого были созданы предпосылки для точных и непротиворечивых лингвистических высказываний. Однако, как показывает опыт других наук, для более глубокого проникновения в исследуемый предмет таких понятий, лежащих на уровне категориальной шкалы, бывает недостаточно. Так, открытие большого числа новых веществ в химии потребовало введения все новых и новых категорий и обозначений, методов синтеза одних веществ из других. Но для создания первого представления о структуре веществ – открытия молекулы был использован количественный анализ, с помощью которого были установлены точные массовые пропорции в таких соединениях. В лингвистике со временем тоже потребовались порядковые метрические, т.е. количественные понятия, а вместе с тем модели и методы, которые, естественно, основывались на качественных понятиях. Потребность лингвистики в понятиях более высокого уровня можно обосновать помимо общих эпистемологических соображений (ср., напр., Stegmüller 1970: 98-109) – следующими четырьмя аргументами:

1. Детерминистский подход позволяет описать среди языковых явлений только "крайне закономерные" специальных случаев которых, однако, очень мало. Для преодоления этого ограничения требуется стохастическое моделирование языка, а вместе с ним и количественные методы. Одним из оснований для такого утверждения является факт изменения языка с течением времени. Каждая языковая система подвергается очень большому числу слабых и малому числу сильных влияний, изменяющихся скачкообразно. В синхронной детерминистской системе правил отсутствуют какие-либо средства для описания соответствующих отклонений. Изменения в языковых системах не только поддаются измерению, они сами лежат в основе определенных (стохастических) закономерностей. Еще одним, общеизвестным примером недостаточности детерминистского описания является размытость значения; в связи с этим в современной семантике все в большей степени используются количественные модели (например, основанные на теории нечетких множеств).
2. При изучении языка нам приходится иметь дело с дефицитом информации, характерным особенно для общественных наук: мы можем охватить лишь небольшой фрагмент исследуемого объекта: он или бесконечен (как, например, множество текстов на немецком языке), или меняется быстрее, чем его можно описать (как, например, лексика языка). Однако такое положение дел представляет собой классический случай необходимости применения статистических методов, которые позволяют делать выводы при недостатке информации.
3. Построение теорий и гипотез является дедуктивным процессом: новые гипотезы выводятся из известных высказываний или из предполагаемых связей, а не из данных. Дедукция – "качественный" процесс, независимо от того, составляется или

решается дифференциальное уравнение, моделируется ли стохастический процесс или делаются выводы на основе классической логики. Различие между качественным и количественным методом заключается в получаемом результате: путем подстановки количественных значений в параметры дифференциального уравнения получаемого вида достигается лучшая аппроксимация, большая точность отображения по сравнению с другими видами дедукции, получаемый качественный вид далее не уточняется, и предмет отображается лишь нечетко. Этот недостаток качественных конечных результатов вывода привел в последнее время к разработке теории нечетких множеств, нечеткой и вероятностной логики.

4. Поскольку, по-видимому, все процессы и отношения в языке имеют стохастический характер ("в языке" – не значит обязательно "в говорящем"), соответствующие модели можно проверять только стохастически. Этот индуктивный процесс осуществляется в квантитативной лингвистике средствами теории выборов, оценок и испытаний.

Обычно применяемые математические средства – дифференциальные уравнения, разностные уравнения, стохастические процессы, теория вероятностей, шкалирование, теория графов, теория информации и математическая статистика позволяют квантитативной лингвистике с помощью своих более дифференцированных понятий и соответствующего исследовательского аппарата добиться более точного описания языковой действительности, чем качественно ориентированные поддисциплины. Между тем в лингвистике наблюдаются первые шаги объяснительного уровня этой науки. Поиск объяснений, поиск общих законов осуществляется в ней в соответствии с уровнем развития понятий, главным образом, представителями квантитативной лингвистики. У предшественников этого направления, например, у младограмматиков (общие звуковые законы) или у Дж. К. Ципфа (принцип наименьших усилий) для построения теории не было необходимых предпосылок. Только с началом использования математических методов в лингвистике был сделан решительный шаг в развитии объяснительного подхода, согласно которому под объяснением понимается вывод объясняемого явления из начальных условий с помощью законов (ср. Hempel 1965).

Здесь следует указать на столь же широко распространенное сколь и ошибочное убеждение, что лингвистика давно прошла до теоретическую стадию. Возможно, это недоразумение вызвано сходством между аксиоматическими системами (в том виде, как они известны в формальных разделах математики и логики) и правилами их вывода, с одной стороны, и дедуктивным принципом научного объяснения, с другой стороны. Применение дедуктивных описательных моделей, например, формальных грамматик, позволяет получить высказывания типа "Предложение S объясняется правилом R" и т.п.; часто при использовании таких высказываний стираются различия между описательными системами правил (например, грамматиками) и объяснительными системами законов (например, научными теориями). Так, например, часто не учитывается, что наблюдаемые и описываемые с помощью правил (более или менее сильные) закономерности требуют объяснения. В свою очередь объяснения возможны лишь с помощью законов; понятийные системы, аксиоматики и систематизирующие описания являются лишь необходимыми, но недостаточными предпосылками. Общие цели квантитативной лингвистики – в той мере, в какой они выходят за пределы качественной лингвистики – можно сформулировать следующим образом:

- введение понятий на уровне ординальной, интервальной и пропорциональной шкалы и разработка или адаптация соответствующих методов для уточнения описания;
- поиск общих законов структуры и динамики языков и текстов;
- определение языковых и внеязыковых параметров и граничных условий для этих законов;
- введение этих законов в теорию языка или текста.

Поскольку используемые для объяснения законы в свою очередь тоже требуют пояснений, поиск объяснений не имеет ни логического, ни эмпирического завершения (ср. Popper 1973, 216 и далее). Для лингвистики пределы объяснительных цепей определяются границами ее предметной области; таким образом, "последние" объяснения в рамках будущей лингвистической теории находятся на стыках со смежными науками: биологией, физикой, психологией и социологией¹. С построением лингвистических теорий связан исключительно важный вопрос о научно-теоретических нормах и методологических стандартах, которые должны при этом соблюдаться.

В настоящей работе представлена разделяемая большинством количественных лингвистов точка зрения, что положение подобающее этой науке, определяется критическим рационализмом (ср. Popper 1972, 1973; Bunge 1967). Таким образом, под теорией понимается система предложений (законов), которая делает возможным объяснение других предложений (отдельных высказываний). Процесс объяснения носит дедуктивно-законоведческий характер и должен удовлетворять условиям схемы Хемпеля-Оппенгейма (Hempel, Oppenheim 1948, Stegmüller 1969). Кроме того, с данной точкой зрения связано требование об эмпирической фальсифицируемости теории и каждой гипотезы².

2. ИЗМЕРЕНИЕ И МОДЕЛИРОВАНИЕ В ЛИНГВИСТИКЕ

Структурализм позволил языковедению обрести фундамент точных научных понятий и создал тем самым предпосылки для ее математизации. В качестве центрального было введено понятие системы, в рамках нового подхода язык стал знаковой системой, элементы которой – отдельные языковые знаки – по своей форме являются результатом договоренности. Успехи, достигнутые с помощью структурного подхода, с самого начала характеризовались значительной односторонностью. В частности, (необходимый в то время) акцент на разграничение синхронического и диахронического аспектов рассмотрения языка позволял нарисовать лишь статическую картину языковой системы. Процессы и взаимодействия между элементами системы выпадали из поля зрения исследователей. С другой стороны, лингвистика интересовалась только отношениями между знаками в системе, причем они рассматривались как (парадигматические и синтагматические) отношения исключительно категориальной природы. Самое позднее с начала этого столетия все большую роль в лингвистике стали играть количественные понятия и описания. Практические применения потребовали разработки методов измерения для подсчета частот комбинаций букв и частот слов. Потребность в частотных списках языковых элементов при разработке скорописи (Estoup 1916), построении языковой теории в сравнительных философских исследованиях (Condon 1928, Zipf 1949) явилась

¹ Естественно, следует иметь в виду, что предметные области наук и самоосознание дисциплин подвержены изменениям

² ср., однако, современные взгляды теоретического плюрализма, согласно которым теория отвергается только при наличии другой, лучше подтверждаемой теории (Lakatos 1974, 1975; Spinner 1974).

побудительным мотивом для получения эмпирических данных, имевших огромное значение в истории количественной лингвистики. Эти результаты свидетельствовали, в частности, о том, что произведение частот слов в текстах и их ранга (в порядке убывания частоты) для всех рангов в каком-либо тексте примерно постоянно. Дж. К. Ципф, открывший подобную форму рангово-частотного распределения для большого числа других явлений, попытался объяснить эту количественную связь на основе всеобщего принципа наименьших усилий. Он был убежден, что открыл новый всеобщий закон, согласно которому все действия людей находятся под воздействием факторов унификации и диверсификации и подчиняются требованию экономии затрат.

Под экономией в данном случае отнюдь не следует понимать осознанную стратегию действий типа оптимизации затрат в хозяйственных связях; здесь речь должна идти скорее о независимом от возможности осознанного управления принципе человеческого поведения в виде антропологической константы.

Понятия, введенные Ципфом, были уточнены Б. Мандельбротом. В частности, он ввел рабочее определение понятия затрат, которое основывается на длине слов, что позволило ему создать математическую модель минимизации средних затрат путем оптимизации длины слов. Данная модель, позволившая лучше, чем формула Ципфа, спрогнозировать численную связь между частотой и рангом, известна под названием закона Ципфа-Мандельброта. Для описания того же эмпирического материала Саймон (Simon, 1955; 1960) предложил проект альтернативной модели, в которой порождение текста рассматривается как стохастический процесс. Несмотря на убедительный вывод и обоснование модель не нашла подтверждения, поскольку не совсем согласуется с эмпирическими данными частотно-рангового распределения.

В противовес Саймону гипотеза Ципфа-Мандельброта до сих пор подтверждалась. На основе ципфовой модели удалось также выявить другие интересные свойства частотной структуры текстов. К важнейшему из них относится, несомненно, вывод так называемого объема Ципфа Z , с помощью которого достигается сопоставимость текстов различной длины с т.з. их частотной структуры.

Объем Ципфа определяется как оптимальная длина текста (которая в зависимости от граничных условий и цели текста может, естественно, быть очень различной) и дает прогноз информационного потока в тексте (см. Орлов, Борода, Надарейшвили 1982). Другие более совершенные варианты этой модели, например, предложенные Хейтом и Джонсом, рассматриваются в работе Lanksy/Rodil-Weiss (1980, 53-65). Кроме языковых законов справедливость закона Ципфа-Мандельброта была доказана также на примере музыкальных текстов (единица музыкального текста, соответствующая слову, получила название F-мотива (ср. Борода 1982a, 1982b, 1982c) и для распределения площади цветов на картинах (ср. Волошин, Орлов 1982). Соответствующие частотно-ранговые распределения известны также в науках о документации, в социологии, в географии, математике и др.

Огромное значение закона Ципфа-Мандельброта заключается, вероятно, не столько в рассматриваемом предмете исследований, сколько в том, что с его появлением в языкознании впервые была сформулирована гипотеза, которая

- оперирует количественными понятиями,
- вытекает из теоретических соображений,
- является составной частью системы высказываний (о человеческом поведении),

- хорошо подтверждается эмпирически и поэтому по праву носит название закона.

2.1 РАСПРЕДЕЛЕНИЯ ФУНДАМЕНТАЛЬНЫХ ЛИНГВИСТИЧЕСКИХ ХАРАКТЕРИСТИК

Еще одна математическая модель, описывающая частоту слов, была предложена Дж. Б. Кэрроллом, которого особенно интересовала связь между объемом словаря и объемом текста (TTR). Он показал (J. V. Carroll, 1967, 1968, 1969, 1970/72), что логнормальное распределение дает эмпирически значительно лучшие результаты, чем распределение Ципфа-Мандельброта. Кроме того, модель позволяет получать текстовую частоту из словарной частоты. Однако неизвестно ни одного способа, позволяющего теоретически обосновать логнормальное распределение в качестве модели для частоты слов.

Для других фундаментальных величин³ также были найдены статистические распределения, описывающие встречаемость этих лингвистических характеристик в инвентаре языка или даже их появление в текстах. Уже достаточно давно (см. Fuchs 1955) длина языковых единиц была исследована и использована в качестве индикатора в целях типологии и филологии (определение авторства и классификация текстов; ср. особ. Rieger 1979). Отрицательное биномиальное распределение в качестве модели, описывающей распределение длин слов удалось получить Гротьяну (Grotjahn 1982).

Распределение этой величины в словаре изучал Ю. К. Крылов (1982a, 1982b), разработав комбинаторную модель и проверив ее на материале русского языка.

2.2 ЗАВИСИМОСТИ МЕЖДУ ЛИНГВИСТИЧЕСКИМИ ВЕЛИЧИНАМИ

Ципф первым также обратил внимание на связь между длиной слова и его частотой. Для объяснения этой связи он выдвинул гипотезу, что при частом употреблении слова подвергаются действию процесса сокращения. Другими авторами тоже отмечена зависимость данных двух величин друг от друга; для ряда романских языков проведены количественные исследования (ср. напр. Guiter 1974). Что касается вида этой зависимости, исследователи не могут прийти к единому мнению. Оспаривается прежде всего направление зависимости. Одной из задач настоящей работы является попытка дать удовлетворительный ответ на этот вопрос в рамках предлагаемой здесь модели.

Прогноз зависимости числа значений слова от его длины получил Г. Альтман (Altmann, 1980) как следствие из закона Менцерата (см. ниже). В работе Altmann, Beóthy, Best (1982) это допущение было использовано и проверено; оно нашло подтверждение при исследованиях текстов для десяти типологически различных языков (см. Fickermann, Marker-Jaeger, Rothe 1984; Sambor 1984).

Семантическая сложность, как называет эту величину Альтман, была операционализирована им как число дефиниций в словаре⁴, длина – как число слогов или графем. В названных работах по данному вопросу связь между этими величинами

³ Под фундаментальной величиной мы понимаем величину, которую нельзя составить или вывести из других величин (ср. Bunge 1967).

⁴ Такая операционализация, естественно имеет свои недостатки (ср. для этого анализ полилексии в разделах 464 и 61); однако ее можно обосновать и она дает стабильные и убедительные результаты.

подводится под закон Менцерата путем придания семантической сложности роли компонента. Эта точка зрения из-за несколько слабой аналогии между числом значений и величиной, используемой в качестве компонента, не будет здесь излагаться, хотя возможны интерпретации закона Менцерата, позволяющие включить соотношение одних длин с другими и соотношение длин и числа значений в одну модель (см. Altmann, Schwibbe 1985). Здесь, однако, мы считаем нецелесообразным подробнее рассматривать эту проблему, поскольку данные соотношения будут получены ниже другим способом.

Согласно Ципфу, число значений слова зависит от его частоты. Вид связи, которую Ципф выдвинул в виде гипотезы, можно рассматривать как особый случай закона Менцерата. Ципф сумел эмпирически доказать связь этих двух величин, однако на теоретический вывод данного соотношения он оказался неспособен. Поэтому его формула имеет лишь описательный статус.

Исследование отношений между тремя фундаментальными величинами – длиной, частотой и числом значений – провел Х. Гитэ (Guiter 1974). В данном случае снова оказалось, что (за недостатком теоретических модельных представлений) ясность отсутствует даже в вопросе о направлениях зависимостей, а также о том, какие математические функции адекватно отображают языковые отношения.

3. СТРУКТУРНЫЕ ЗАКОНЫ

В области структурного описания языков и текстов использование таких качественных разделов математики как алгебра и логика при построении моделей приводит к необходимости ограничить рассмотрение крайне закономерных явлений. Все грамматические явления и большое число стилистических поддаются, однако, описанию средствами комбинаторики и теории вероятностей при их рассмотрении в терминах непосредственной или опосредованной сочетаемости или одновременной реализации их компонентов. Вероятностные модели позволяют покрыть весь спектр между жестким детерминизмом и слабой связью, традиционные филологические (при их достаточной точности) и формально-лингвистические представления оказываются очень ограниченными специальными случаям и более дифференцированных количественных возможных представлений⁵.

Хотя лингвисты, изучающие структуру языка или текста, почти все без исключения – сознательно или бессознательно – ограничиваются описанием структуры, нет никаких оснований отказываться от попыток создать в этой области модели, обладающие объяснительной силой. Напротив, именно поиск законов строения языка и текста должен стать центральной задачей языкознания в части построения теории. Грамматические, семантические и фонологические структурные свойства всех языков и закономерности их употребления требуют объяснения и могут быть объяснены только на основе такой теории.

⁵ Постоянно повторяющееся возражение против качественных синтаксических моделей (связанные в первую очередь с использованием цепей Маркова), что из-за повышенной обобщенности целесообразное допущение о бесконечной синтаксической глубине предложения (ср. Chomsky 1957; Osgood 1963; Levelt 1974) нельзя передать стохастическими средствами, легко опровергается применением рекурсивных вероятностных грамматик (см. напр. Suppes; ср. тж. Levelt 1974).

На ряду с законом Ципфа-Мандельброта, предметом которого является частотная структура текстов, следует также остановиться в первую очередь на законе Менцерата.

Такое, распространенное в настоящее время в компьютерной лингвистике применение стохастических моделей встречается, например, во многих вариантах систем машинного анализа (парсеров), где синтаксическим альтернативам без особых рассуждений приписываются вероятности.

Исходя из наблюдений Менцерата (1954), который выявил на материале немецких слов корреляцию между длиной слова (измеренной в слогах) и длиной слогов (измеренной в фонемах), Альтман (1980) вывел из теоретических соображений дифференциальное уравнение, решение которого в общем виде выражает зависимость длины компонента (напр., слога) от длины конструкта (напр., слова). Он назвал эту связь, которая претендует на справедливость для всех языков и всех уровней лингвистического описания, законом Менцерата.

Все проведенные до сих пор эмпирические исследования (ср. Gerlach 1980; Neups 1983; Kцhler 1982; Schwibbe 1984; Teupenhayn, Altmann 1984) достаточно подтверждают эту гипотезу, которую в самом обобщенном виде можно сформулировать следующим образом:

”Длина компонентов является функцией длины состоящего из них языкового конструкта” (Altmann 1980).

Помимо всего прочего, закон Менцерата допускает ряд интересных психолингвистических интерпретаций и имеет некоторые последствия, значимые для замысла настоящей работы (в первую очередь в области лексики): как будет показано ниже, дифференциальное уравнение, лежащее в основе закона, имеет фундаментальное значение и для других связей между лингвистическими величинами.

2.4 ЗАКОНЫ ЯЗЫКОВОЙ ДИНАМИКИ

Еще младограмматики пытались объяснить определенные изменения языковой системы, в частности изменения свойств ее элементов путем простого описания с помощью (звуковых) законов.

Основной причиной того, что их программа не могла осуществиться, явилось предположение о детерминистской закономерности изменения звуков; стохастические модели, отражающие тенденции постепенного развития или затухания явлений, еще не были в их арсенале. Сегодня этого ограничения больше не существует, что позволяет моделировать изменение языка в смысле "точного изучения изменения языка" (ср. Best, Kohlhase 1983) как управляемого стохастическими процессами. Эмпирической основой для этого исследования являются, с одной стороны, соответствующие данные по разным уровням языка, с другой стороны закономерная организация и распределение лексических единиц (рассматриваемых на определенном уровне языка) с т.зр. их возраста и происхождения.

Р. Г. Пиотровский, К. Б. Бектаев и Л. Л. Пиотровская (1977) сформулировали экспоненциальный закон увеличения числа слов любой семантической, терминологической или другой определенной сферы лексики с течением времени. Кроме того, оказывается, что восприятие иноязычных лексических элементов подчиняется другому

закону Р. Г. Пиотровского (ср. например Веóthy, Altmann 1982). Этот закон был выведен на основе некоторых теоретических соображений о распространении языковых инноваций любого рода. Он был уже использован для изучения тенденций развития в русском языке, а также динамики заимствования в немецком и венгерском языках (см. Altmann, V. Buttler, Rott, Strauss 1983; Best, Kohlhase 1982).

Еще один подход, связанный, в принципе, с глоттохронологией, предложен М. В. Араповым и М. М. Херцем (1983). Описано разрушение словаря языка в зависимости от частотного класса и возраста слов.

За последние 30 лет во всех разделах квантитативной лингвистики следует отметить значительные достижения в области методологии. Эти достижения проявляются как в эмпирической области, где введение количественных понятий и все более претенциозных методов привело к появлению многих важных данных и гипотез, так и в теоретической квантитативной лингвистике, где все большую роль играет разработка дедуктивных моделей, и, несомненно, проявляется тенденция к постановке вопросов и задач в объяснительном аспекте.

Несмотря на то, что в последнее время открыто достаточное количество языковых и текстовых законов, выяснить граничные условия или факторы, определяющие их параметры, до сих пор еще не удалось. Кроме того, между различными гипотезами и законами нет никакой связи; в первую очередь, у них еще нет объединяющего их всех теоретического фона.

3. ПОСТАНОВКА ЗАДАЧИ И МЕТОДИКА ИССЛЕДОВАНИЯ

В настоящей работе мы ставим задачу сделать шаг в направлении интеграции различных подходов в общую модель; теоретической платформой для реализации данной лингвистической теории послужили идеи, намеченные Г. Альтманом (Altmann 1978).

Самой актуальной задачей в области построения теории является сопряжение известных и новых законов и гипотез для построения системы⁶ общих высказываний. В настоящей работе будет сделана попытка построить базовую модель, на основе ядра лингвистических связей, в которой для начала используется небольшое число переменных, принятых в качестве центральных. Однако эта модель должна выполнять следующие функции

- выводить некоторые наиболее важные законы квантитативной лингвистики, которые до сих пор были теоретически изолированными;
- выводить другие законы и прогнозировать неизвестные до сих пор связи;
- определять параметры законов, а также их граничные условия;
- осуществлять уточнения двух видов: вводить дополнительные переменные и учитывать дополнительные отношения.

⁶ Термин "система" используется в этой работе в нескольких значениях:

1. система законов (теория);
2. (под)система языка (предмет исследования в лингвистике);
3. система моделей (часть лингвистической модели, ее системно-теоретическое изображение).

Наконец, из модели должно вытекать, каким образом она сама выводится из более высоких принципов. Моделирование осуществляется аксиоматически и на уровне абстракции, который лежит выше лингвистического уровня; таким образом, настоящее исследование предполагает уже построение лингвистических понятий.

Здесь следует подчеркнуть, что в работе не преследуется цель полностью, во всех аспектах охватить с помощью развиваемой модели какую-нибудь лингвистическую область, тем более, что поиск подходящего критерия полноты для такой динамической модельной системы может быть успешным только при обращении к теоретическим соображениям с объяснительным содержанием; но они находятся на стадии разработки. Важным прагматическим основанием для самоограничения является опасность чрезмерного усложнения модели. Основной задачей этой работы является разработка подхода или принципа моделирования, что должно осуществляться на основе анализа нетривиального фрагмента предметной области; с другой стороны, работа не должна потонуть в рассмотрении частных проблем.

Поэтому в предлагаемой модели в качестве элементов используется некоторый ограниченный набор величин, которые до сих пор считались особо важными и уже интенсивно исследовались (хотя и изолированно друг от друга). Речь идет о переменных:

- (1) длина
- (2) семантический потенциал (полилексия)
- (3) частота употребления лексических единиц. На основе теоретических соображений было показано, что еще одна величина
- (4) связанность с контекстом (политекстия) тесно связана с тремя вышеназванными и должна быть включена в модель. Кроме того, модель дополнена глобальными величинами, такими как
- (5) объем словаря
- (6) число фонем

Модель использует средства теории графов и линейных операторов в виде набора правил. Предлагаемая модель должна в первую очередь дать возможность объяснить наблюдаемые связи между лингвистическими величинами. Это достигается путем теоретического обоснования отношений между элементами системы в виде дифференциальных уравнений, что позволяет получить гипотезы, претендующие на статус законов.

Основой и предпосылкой для всего предлагаемого подхода является предположение, что человеческий язык (наряду с другими знаковыми системами) является саморегулирующейся системой. В результате становится возможным использовать в качестве основы для объяснения функциональный анализ, не подвергаясь при этом опасности принятия за основу телеологических или других псевдообъяснений. Разумеется, для этого потребуются модифицировать и дополнить рядом элементов традиционную функциональную логику объяснений, известную из научно-теоретической литературы, в первую очередь из общественных наук.

Если модель, как уже отмечалось, внесет свой вклад в построение лингвистической теории, то для доказательства связи с действительностью и правильности настоятельно потребуются эмпирическая проверка теоретически полученных гипотез. С этой целью постулируемые связи и основные следствия, вытекающие из модели в ее теперешнем

состоянии, проверяются на выборках из текстового материала с помощью статистических методов испытаний.

В заключение будет рассмотрено возможное значение предлагаемого подхода для построения языковой теории и для сопряжения лингвистических объяснений в общей теории саморегулирующихся систем.

4. ПОСТРОЕНИЕ МОДЕЛИ

4.1 ВВЕДЕНИЕ: ЭЛЕМЕНТЫ НА УРОВНЕ ФОНЕМ/ГРАФЕМ

Для объяснения частотно-ранговых распределений слов Ципф использовал постулированный им принцип наименьших усилий (Zipf 1949), основанный на оптимальном равновесии противоположно направленных факторов унификации и диверсификации. Аналогичный подход был избран здесь в качестве исходного пункта для моделирования механизмов на уровне фонем и графем. Для начала постулируется наличие у пользователей языка двух потребностей, которые в большей или меньшей степени должна удовлетворять языковая система:

- минимизация необходимых затрат памяти (в дальнейшем minG) и
- минимизация необходимых затрат на порождение речи (в дальнейшем minP).

Эти потребности следует интерпретировать как требования к языковой системе; они лежат в основе "фактора унификации" и сами к системе не принадлежат.

Если с точки зрения этих потребностей рассмотреть свойства фонем или графем, то легко обнаружить, что при фиксированном инвентаре фонем/графем затраты памяти и затраты на порождение тем меньше, чем более сходными между собой они являются. Память нагружается в тем меньшей степени, чем меньше признаков необходимо помнить; оптимум⁷ достигается, когда все элементы⁸ имеют только признаки с различительной функцией.

Порождение требует меньших затрат, когда оно состоит из единообразных артикуляций или манипуляций: при говорении артикуляционные признаки (положение языка, напряженность голоса и т.д.) легче сохранять, чем менять (ср. например Lindblom 1983). То же можно сказать и о письме (ср. тж. Watt 1983). Поэтому при заимствовании иностранных слов новые фонемы, если они вообще сохраняются, адаптируются к признакам воспринимающей их системы; это приводит к дальнейшему развитию фонемной системы и системы письма.

Связь между потребностями и сходством элементов инвентаря схематически представлена на рисунке 1. Сходство знаков между собой рассматривается как глобальное свойство инвентаря знаков. На этом и последующих рисунках знак плюс в квадрате, сквозь который проходит стрелка, выражает положительную зависимость между исходной величиной и величиной, на которую указывает стрелка. Знак минус выражает отрицательную связь: чем больше значение исходной величины, тем более она вызывает уменьшение значения конечной величины.

⁷ Для инвентарей с числом элементов больше 1.

⁸ Инвентарь из одного единственного элемента с одним единственным признаком можно считать теоретическим оптимумом с т.з. minG.

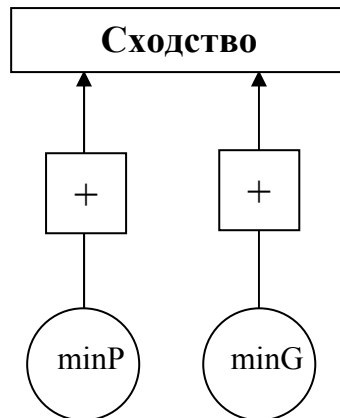


Рис. 1. Системная величина "сходство" зависит от двух потребностей: $\min P$ и $\min G$.

В соответствии с вышеизложенным знаковая система фонем или графем, оптимально удовлетворяющая названные потребности своих пользователей, должна состоять из знаков с максимальным сходством. В крайнем случае максимальное сходство означает совпадение всех знаков в одном единственном. Тогда дополнительно привлекается системная величина "объем инвентаря фонем", которая находится с потребностью $\min G$ в таком же отношении, что и сходство знаков: чем меньше объем инвентаря фонем, тем меньше затраты памяти.

Однако, если добавить еще одну потребность (соответствующую цифровой диверсификации, а именно

- минимизацию затрат на понимание и декодирование ($\min W$),

то возникает другая картина. Для облегчения понимаемости и читаемости знаки должны быть как можно больше непохожими друг на друга (избыточность признаков); таким образом, $\min W$ оказывает воздействие, обратное двум другим потребностям: негативное воздействие на сходство и опосредованно через него – позитивное воздействие на объем инвентаря фонем⁹.

В лингвистику этот принцип был введен структуралистами под названием принципа максимальной дистинктивности и применялся при моделировании фонемных систем. Одно из его применений при количественном анализе фонемных систем описано в работе (Strauss, 1980).

Однако, с другой стороны, изменение величины "объем инвентаря фонем" вызывает одно последствие, которое нужно учитывать. Чем меньше имеется элементов на уровне второго членения, тем длиннее состоящие из них элементы на уровне первого членения, т.е. лексических единиц, если их число постоянно. Так, из фонемного инвентаря, включающего единственный элемент (а), можно построить только слова /а/, /аа/, /ааа/,..., т.е. только одно слово любой длины n . Из инвентаря с объемом x можно построить x слов длины 1, x^2 слов длины 2, вообще x^n слов длины n .

Однако потребности $\min P$ и $\min G$ требуют возможно меньшего значения величины "длина слова". Модель, учитывающая эти отношения между тремя величинами "сходство", "объем инвентаря фонем" и "длина слова", а также между тремя

⁹ Так же, как больше сходство ведет к малому объему инвентаря фонем, уменьшение сходства под влиянием $\min W$ вызывает увеличение объема инвентаря фонем.

потребностями $\min G$, $\min P$, $\min W$, представлена на рисунке 2. В этой конструкции имеются требования, прямо или косвенно противоречащие друг другу. Поэтому оптимальное значение величины не может быть экстремальным. Одновременный учет всех потребностей требует компромисса на основе оптимального равновесия.

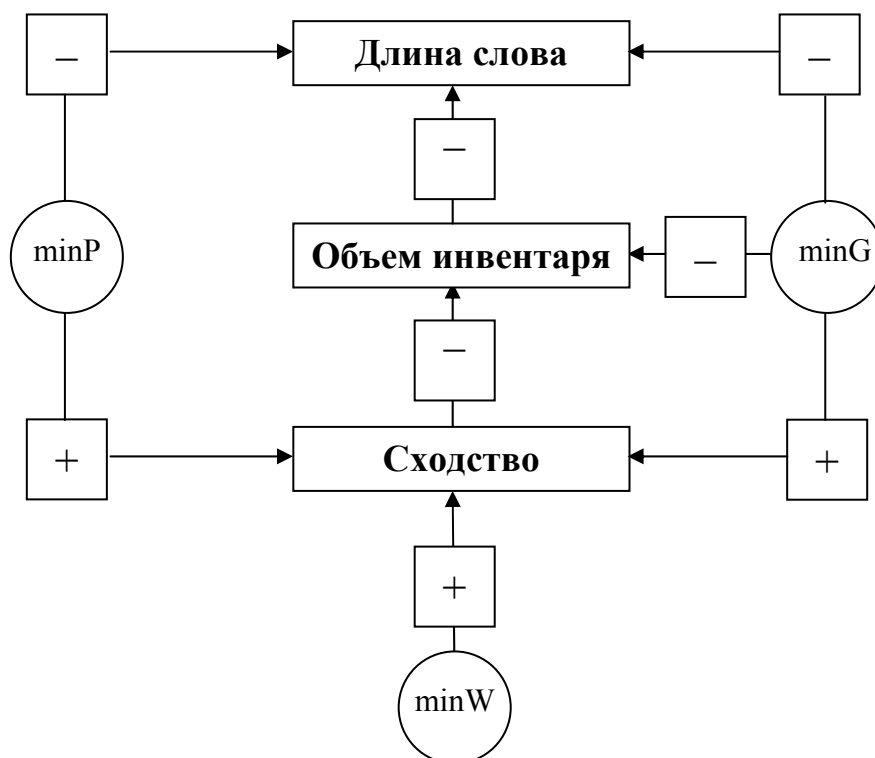


Рис. 2. Подсистема с тремя величинами и тремя системными потребностями.

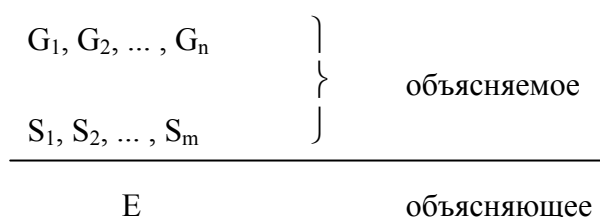
Эта простая модель, однако, не учитывает в частности влияния тонов и ударений на длину слова; тонические языки обходятся более короткими словами, поскольку обладают дополнительным признаком. В таких языках число тонов действует понижающим образом на длину слова. Языки с подвижным ударением тоже обладают дополнительным средством для дифференциации, что, например, часто проявляется во флексии.

Разумеется, в эту систему следовало бы ввести такие другие важные величины, как например, распределение фонем или графем, частоту отдельных элементов, их сложность и параметры их распределений. Однако здесь они не представляют непосредственного интереса, поскольку мы не ставим в качестве первоочередной задачу построения модели для этого уровня фонем или графем. Для наших целей будет достаточно отразить связь между потребностями и величинами "объем инвентаря фонем" и "длина" в качестве связующего звена такой системы с системой на лексическом уровне¹⁰.

¹⁰ Другие влияния на длину слова, например, необходимость в избыточности и экономии затрат мы считаем принадлежащими к лексическому уровню, поэтому они будут рассмотрены несколько позднее.

4.2 ЯЗЫК КАК САМОРЕГУЛИРУЮЩАЯСЯ СИСТЕМА

В предыдущем разделе системные потребности играли роль средств объяснения наблюдаемых или постулируемых явлений. Так, например, мы говорим, что объем инвентаря фонемной системы имеет определенную величину, т.к. он вследствие потребности $\min G$ принимает наименьшее значение, которое возможно при данном значении величины "сходство фонем". Очевидно, здесь речь идет о каузальном объяснении. Согласно Хемпелю и Оппенгейму (см. напр. Hempel 1965) дедуктивно-номологическое объяснение следует общей схеме:



G_1, \dots, G_n – общие, универсальные законы; S_1, \dots, S_m – предложения, описывающие определенные обстоятельства. Законы выполняются только при указанных граничных условиях; когда обстоятельства соответствуют граничным условиям, от них и законов можно дедуктивным способом перейти к *объясняющему*.

Однако для языка законы, каузально связывающие общие потребности человеческой коммуникации (например, минимизацию затрат на память) с каким-нибудь свойством языковой системы (например, величиной фонемного инвентаря) неизвестны и постулировать существование таких законов вряд ли целесообразно.

Лучше исходить из того, что системное свойство объясняется тем, что оно необходимо для функционирования системы. Например, фонемный инвентарь достаточной величины требуется для того, чтобы можно было построить достаточно большое число различных морфем. Такое функциональное объяснение является особым видом дедуктивно-номологического объяснения, поднимающим целый ряд проблем.

Одной из важных проблем является проблема функциональных эквивалентов (Hempel 1965a; Esser, Klenovitz, Zehnpfennig 1977): существование элемента, выполняющего в системе определенную функцию, еще не объясняется этой функцией полностью, если ту же функцию (при определенных обстоятельствах другим способом) могут выполнять и другие элементы. Практически во всех сложных системах (например, языковых, биологических, социологических и т.п.) такая возможность существует. Очень часто в различных родственных или сравниваемых системах существуют различные элементы, выполняющие одну функцию; часто наблюдается даже одновременное существование нескольких функциональных эквивалентов. Так что для объяснения величины фонемного инвентаря ссылка на необходимое число морфем недостаточна, т.к. построение более длинных морфем из меньшего числа различных фонем и/или применение тонов является эквивалентным с т.з. числа морфем. Такая же проблематика известна в других областях знания: существование сердца в организме млекопитающего нельзя объяснить выполняемой им функцией кровяного насоса, т.к. другие варианты кровоснабжения организма не только представимы но и реально существуют в природе (например, открытая система у червей). Деятельность групп людей можно соотнести с помощью принципа уравнения и делегирования, а также через дискуссию и договоренность, через унаследование ролей при разделении работы и многое другое.

Удовлетворительное объяснение существования эквивалентов возможно лишь тогда, когда отношения между всеми функциональными эквивалентами можно задать функцией¹¹. Альтман (1981) требует вывода этого отношения из аксиом, теоретических допущений или законов. В предлагаемой работе делается попытка удовлетворить это требование: отношения элементов системы между собой вытекают из теоретических допущений и эксплицитно моделируются¹². Второй фундаментальный вопрос в связи с объяснением в функциональном анализе – логическая связь между существованием потребности и ее удовлетворением с помощью элемента или свойства системы. Мы не можем сослаться на общие законы, из которых при наличии потребностей их соблюдения можно было бы вынести каузальные связи. Значительно целесообразнее рассматривать язык как саморегулирующуюся систему (ср. Альтман 1981). Язык как таковой имеет механизмы, которые постоянно стремятся оптимально приблизить состояние системы к требуемому состоянию; при этом требуемое состояние можно интерпретировать как вектор потребностей системы. Существование системного свойства, которое обслуживает потребность, может быть объяснено ссылкой на приспособительный механизм такого рода.

Полноценное функционально-аналитическое объяснение явления E_f можно представить в виде следующей схемы:

- а) Система S – саморегулирующаяся: для каждой потребности существует механизм, который меняет состояние системы таким образом, что она удовлетворяется.
- б) Системе S ставятся в соответствие потребности $D_1 \dots D_f$.
- в) Потребность V может обслуживаться функциональными эквивалентами $T_1 \dots T_f \dots E_n$.
- г) Между функциональными эквивалентами существует отношение $R(E_1, E_2, \dots, E_n)$.
- д) Структуры системы определяет существование отношения $Q(S_1, \dots, S_m)$ между элементами $S_1 \dots S_m$ системы S .

E_f – элемент системы со значением R_f .

Это объяснение действительно тогда, когда отношения R и Q таковы, что все другие возможности обслуживания D_i исключены или менее благоприятны, чем E_f . Такой функциональный анализ должен быть завершен заданием функций $R_i(E_1, \dots, E_n)$, которые для каждой потребности V_i определяют значения функциональных эквивалентов E_f таким образом, что V_i обслуживаются оптимально. Такие функции нельзя получить эмпирическим путем; они требуют теоретического вывода.

Предложенная схема дает лишь статическую картину связей между потребностями и свойствами системы. Однако, саморегулирующиеся системы следует обязательно рассматривать в динамическом плане. Теоретически и практически важным вопросом является временная потребность механизмов приспособления. Изменение в окружении системы (например, в используемых средствах коммуникации) заставляет систему постепенно менять некоторые из своих свойств. Поскольку свойства системы связаны

¹¹ Одна из фундаментальных логических проблем заключается в том, что не все функциональные эквиваленты известны. Теоретически может существовать бесконечно больше число различных эквивалентов для определенной функции.

¹² Существование не предусмотренных моделью величин при этом не исключается, напротив, речь идет, безусловно, об эмбриональной фрагментарной модели.

между собой, это снова вызывает другие изменения и т.д., до тех пор, пока не будет достигнуто новое равновесное состояние системы. Этот процесс может быть дополнительно замедлен за счет сверхкомпенсации. Из-за сложности системы такое равновесие всегда является, по-видимому, компромиссом между большим числом отклоняющихся (дивергирующих) требований к основным элементам системы, который, кроме того, отстает по времени от развития потребностей.

Определяемое окружением системы требуемое состояние тоже не является постоянной величиной. Общественные отношения, в которые в качестве неотъемлемого элемента входит язык, изменяются очень быстро, так как еще до достижения оптимального компромисса потребности снова могут получить новое значение. Поэтому мы говорим здесь также о динамическом равновесии свойств системы.

По всем этим причинам язык в любой момент времени не находится в состоянии, которое по всем свойствам соответствует требуемому состоянию, постоянно находясь в движении¹³ (динамику отдельного свойства можно наглядно представить в виде графика на рис. 3).

Таким образом, лингвистические данные поддаются описанию в основном только с помощью стохастических средств. Явлений, представленных в виде детерминистских правил, меньшинство, и вследствие механизма приспособления они также подвержены изменениям. При рассмотрении языка как саморегулирующейся динамической системы следует смотреть глубже этих поверхностных, преходящих явлений, исследуя более существенные законы структуры и динамики системы.

Как же выглядят механизмы приспособления? Как они влияют на свойства системы и как взаимодействуют между собой реальные и требуемые состояния? Допущение о саморегуляции ни в коем случае не означает попытки телеологического объяснения: происходящие в языке изменения не направлены на достижение определенной заданной цели. Связанное с саморегуляцией представление лучше сравнивать с биологической моделью развития видов, эволюции и приспособления организмов¹⁴. Аналогично принципу мутации и селекции при использовании языка постоянно возникают отклонения и вариации, из которых лишь некоторые пробивают себе дорогу. В течение длительного времени сохраняются только такие варианты, которые тем или иным образом способствуют удовлетворению потребностей системы. Объяснительный подход основан в сущности на утверждении, что необходимость приспособления в приспособлении вызывает возникновение или исчезновение тех свойств системы, которые сохраняют или повышают выживаемость системы в ее окружении.

¹³ Об этом и о проблеме адекватности выражения ср. Альтман (Altmann 1981: 28).

¹⁴ Среди языковедов 19-го века (ср. особ. Steintal 1881) преобладало понимание языка как организма. Однако их точка зрения была очень идеалистической и односторонней: в первую очередь у них отсутствовали методологические предпосылки, позволяющие преодолеть первоначальный наивный подход. Сходная ситуация возникла при введении структуралистической парадигмы с ее столь же ограниченным и односторонним пониманием системы, которое стало преодолимым только после разработки теории систем. Ниже делается попытка показать, каким образом эти оба взгляда включаются в значительную более емкую, динамическую концепцию системы в современной теории систем.

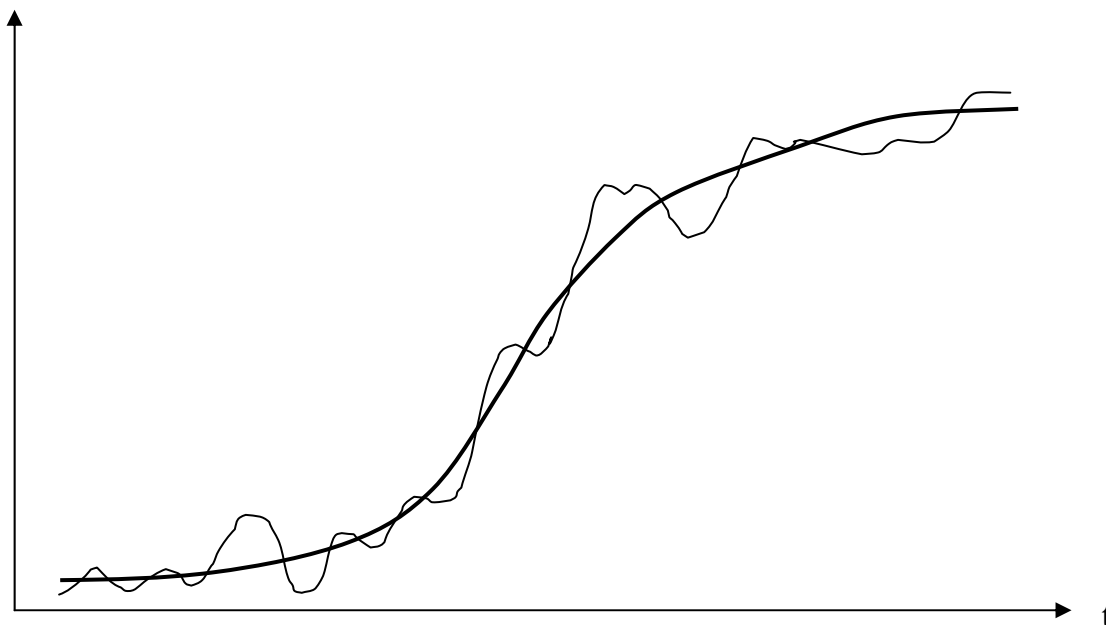


Рис. 3. Динамика языкового свойства. Жирная линия отражает теоретическую функцию $R(E_1, \dots, E_j)$, которая соответствует оптимальному удовлетворению определенной потребности. Тонкая линия отражает эмпирические значения функции.

Это высказывание вводится в модель в качестве аксиомы языковой саморегуляции. Для объяснения процессов мутации и селекции в языке мы снова возвращаемся к примеру потребностей $\min P$, $\min G$ и $\min W$. Каждое отдельное проявление определенного выражения отличается от всех других по большому числу свойств, каждая реализация определенного фона получается по-разному – даже у того же человека. Это постоянный источник фонетических вариаций, которые говорящими и слушающими в основном практически не замечаются.

В отношении отдельного говорящего потребность в минимизации затрат на порождение означает не что иное, как предпочтение тем фонетическим вариантам, которые обеспечивают экономию артикуляционных усилий. Потребность в минимизации затрат памяти проявляется как склонность чаще игнорировать различия, не использовать дистинктивные элементы, а в крайнем случае полностью исключать их из инвентаря. Обе тенденции отвечают за селекцию упрощающих мутаций.

Поскольку у каждого говорящего сходный процесс происходит независимо от всех других говорящих, при отсутствии обратной связи он привел бы теоретически к большому числу различных чрезвычайно простых фонетико-фонологических подсистем. Однако, эти локальные микропроцессы взаимодействуют, например, с противопоставленными им влияниями потребностей слушающего (самого говорящего и других людей). Потребность в минимизации затрат на понимание и декодирование в значительной степени исключает такой исход. Отсутствие дистинктивности при артикуляции затрудняет декодирование слушающим, так что потребуются дополнительные затраты (повышенная концентрация внимания, возможно, многократное повторение высказывания).

Когда возникающие в результате этого дополнительные затраты превышают экономию затрат за счет упрощения артикуляции, соответствующие варианты становятся нежелательными (мутации не закрепляются). Если абстрагироваться от локальных

микропроцессов, связанных с отдельными людьми, тогда их суммарный эффект предстает в виде механизма приспособления, обеспечивающего равновесие между потребностями, ориентированными на говорящего и слушающего.

Поэтому для каждого звука языка можно получить вероятность изменения как функцию артикуляционных усилий и затрат на декодирование. Йоб и Альтман (Job, Altmann 1985) выразили это с помощью дифференциального уравнения

$$\frac{dW}{W} = \left(\frac{k}{1-A} - \frac{b}{A} \right)$$

где W – стремление к изменению, A – артикуляционные усилия, а $1 - A$ – затраты на декодирование¹⁵. Решение $W = c(1 - A)^{-k} A^{-b}$ графически представлено на рисунке 4. Легко можно увидеть, что вероятность¹⁶ какого-то звукового изменения меньше всего тогда, когда между действиями обеих потребностей существует баланс:

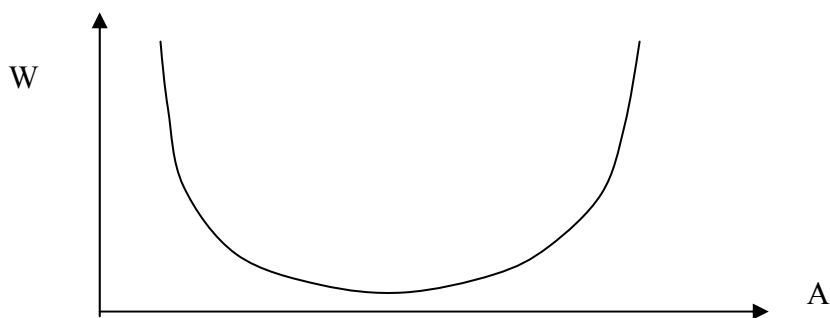


Рис. 4. Зависимость стремления (W) к изменению артикуляционных усилий (F).

4.3 ЛИНГВИСТИКА И СИНЕРГЕТИКА

Явления саморегуляции в других дисциплинах изучаются уже в течение определенного времени; в первую очередь в биологии (Eigen 1971), химии (Prigogine 1979: V) и физике (Naken 1978). Физик Hermann Naken построил на этом, а также на своих исследованиях в области лазеров, универсальную теорию кооперативных явлений: синергетику. Ее основополагающие принципы носят такой общий характер, что языковая саморегуляция без каких-либо изменений подходит под их описательно-объяснительные схемы. Для иллюстрации снова рассмотрим приспособление фонетико-фонологических свойств системы. Эффективный набор фонем определенного языка напрямую зависит от числа признаков, которые применяются носителями языка для дистинкции. Когда вследствие потребностей $\min P$ и $\min G$ возникают фонетические варианты,

¹⁵ Затраты на декодирование тем выше, чем меньше артикуляционные усилия, поэтому $1 - A$.

¹⁶ Вероятность можно вычислить путем нормирования площади поверхности под кривой.

стирающие определенные различия, тогда сливается по крайней мере две фонемы и соответствующим образом сокращается фонемный инвентарь. Это наблюдается, например, тогда, когда признак, до сих пор бывший различным, по причине упрощения артикуляции все больше и больше встречается в дополнительной дистрибуции, и две фонемы превращаются в обусловленные аллофоны. Обратный процесс происходит, когда вследствие потребности $\min W$ выбираются более дистинктивные мутации, что может вызвать появление новой фонемы (напри мер, в результате расщепления дополнительно распределенных аллофонов на отдельные фонемы).

Хотя фонемы являются абстрактными, теоретическими конструктами, они являются непосредственной мерой суммарных затрат на фонетическую дистинкцию, существующую в определенном языковом коллективе. То, что было сказано в связи с потребностями $\min P$, $\min W$ в соотношении фонетических признаков, касается в равной степени и всей фонемной системы: чем меньше фонем, тем больше затраты на декодирование (или приходится обрабатывать более длинные морфемы и слова, или имеются дополнительные затраты при снятии двусмысленности омофонов, чем больше фонем, тем больше затраты на порождение (повышение дистинктивности связано с затруднением артикуляции).

Как показано выше, для вероятности изменения каждой отдельной фонемы существует минимум, наблюдающийся при определенном соотношении между обоими видами затрат A и $1-A$. В переводе на язык синергетики это означает, что в точке этого минимума существует стабильное равновесие. Фонема, затраты на порождение которой находятся в какой-то другой точке, с повышенной вероятностью изменяется таким образом, что в конце концов ее координаты фиксируются в точке минимума – так же, как шар, скатившись со стенки блю да, после нескольких колебаний останавливается в самом глубоком месте (см. рис.5).

Однако это равновесие остается стабильным только при прочих равных условиях. Если меняется какой-то другой параметр, какая-то связанная с подсистемой величина, то сдвигается и минимум. При определенных обстоятельствах может возникнуть особая ситуация.

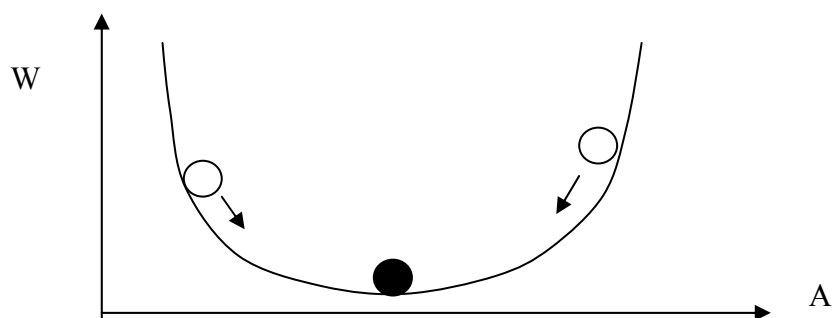


Рис. 5. Стабильное равновесие между видами затрат A и $1-A$ для одной фонемы.

Если из-за повышения частотности заимствованных из других языков выражений начинает возрастать роль до сих пор несуществовавшей или избыточной дистинкции, тогда импортируется дополнительная фонема. Для одной, а чаще всего нескольких

старых фонем это приводит (вследствие максимизации дистинктивности¹⁷) к нарушению симметрии (ср. рис. 6)

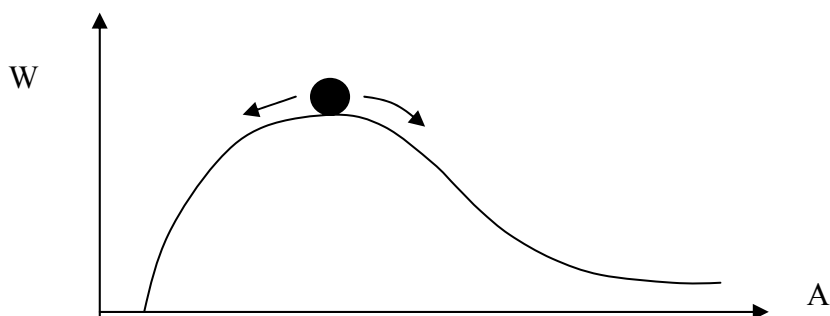


Рис. 6. Нестабильное равновесие

Поскольку коэффициенты функции стремления к изменению фонемы являются функциями других свойств системы, подобные изменения могут так исказить кривую¹⁸, что стабильное равновесие станет нестабильным. Теперь случай решает, в каком направлении пойдет изменение звуков; малейшей фонетической флуктуации достаточно для того, чтобы разрушить равновесие между затратами на артикуляцию и декодирование.

Если сравнить процессы при возникновении и выборе мутаций на фонетическом уровне с процессами изменений на фонологическом уровне, то можно заметить, что первые происходят на порядок быстрее последних. Изменение фонемных систем происходит в течение жизни нескольких поколений, в то время как процессы спонтанной фонетической вариации происходят в любом высказывании конкретного носителя языка. Поэтому можно сказать, что условия для того, чтобы определенная мутация вела или не вела в направлении равновесия, чтобы она могла закрепиться, определяются рамками данной фонологической ситуации в любой момент времени. Эту связь в синергетике называют "порабощением" одного процесса другими. Фонемная система является "упорядочивающим параметром" для фонетического процесса.

4.4 К ВОПРОСУ ОБ ОПРЕДЕЛЕНИИ ПОНЯТИЙ

Для лучшего понимания структуры модели, ее элементов и их соотношений между собой необходимо определить ряд понятий, которые в ней используются; в результате будет более четко очерчен моделируемый объект. Четыре величины – длина, семантическая сложность (полилексия), контекстуальная связанность (политекстия) и частота употребления, которые должны стать ядерными элементами базовой модели, являются свойствами таких языковых единиц как морфемы, слова или сигтагмы. Можно исходить из того, что постулируемые в базовой модели отношения и процессы относятся ко всем типам выражений, которые можно рассматривать как лексикализованные. Такие выражения в дальнейшем называются лексическими

¹⁷ Ср. Strauss 1980, Moulton 1961.

¹⁸ Чтобы вывести это математически, потребуется соответствующее обобщение функции.

единицами. Без потери общности мы также будем говорить в дальнейшем просто о словах.

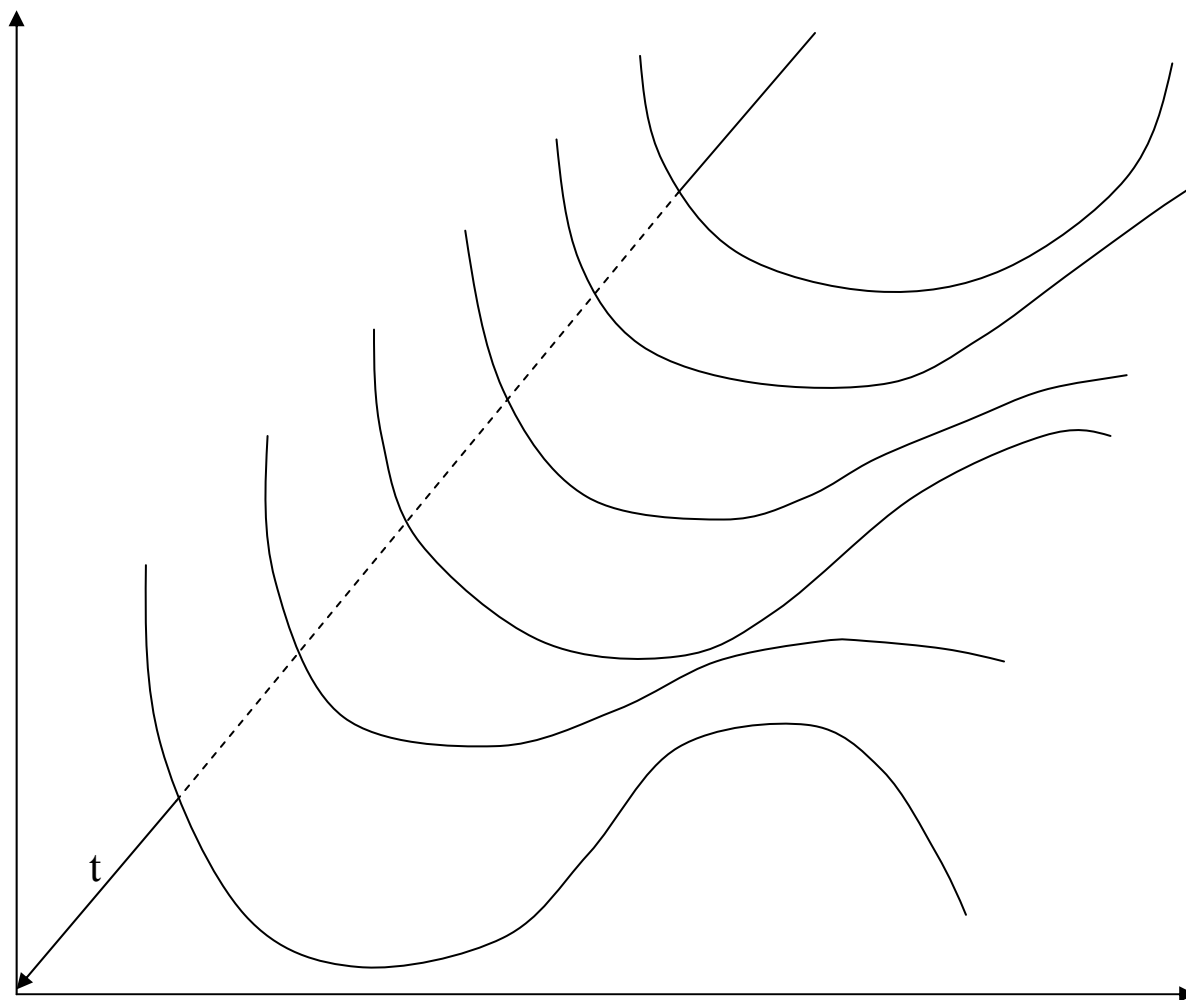


Рис. 6а. Переход от стабильного равновесия к нестабильному

Лексическая единица вводится как неопределяемое лексическое понятие: под ней мы понимаем наблюдаемый в речи, несущий значение, состоящий из одной или нескольких частей элемент лексикона естественного языка. Этим лексическим единицам, как и всем языковым единицам, может быть приписано множество лингвистических свойств. При этом имеется в виду любое свойство, которое данная единица имеет в рамках теоретического лингвистического рассмотрения соответствующей единицы, исходя из структуры модели (по этому вопросу и по вопросу классификации лингвистических свойств ср. Altmann, Lehfeldt 1980: 37-39).

При конструировании этой базовой модели не всякое лингвистическое свойство представляет интерес: те свойства, которые рассматриваются в качестве элементов в структуре модели, т.е. в случае данной модели четыре вышеуказанных следует назвать структурными свойствами. Использование на данном этапе только этих величин ни коем образом не исключает возможности включения в дальнейшем других свойств: такое включение будет даже необходимо, если теоретические рассуждения или эмпирическая проверка модели выявят соответствующую неполноту. Как только этим

вновь учтенным свойствам лексических единиц будет найдено место в модели, т.е. когда будут определены их отношения к другим величинам, они тоже получают статус структурных свойств.

Определение 1

Структурное свойство лексической единицы – это любое свойство, которое наблюдается при лингвистическом анализе применения этой единицы в речи в качестве явного или неявного признака и которое является элементом модели, лежащей в основе наблюдения.

На основе приписанных им свойств лексические единицы можно разбить на лексические классы: частотные классы, возрастные классы, части речи и флексивные классы.

Определение 2

Множество лексических единиц называется лексическим классом, если эти единицы имеют одно или несколько одинаковых свойств. Если общее свойство называется структурным, то и класс называется структурным. В качестве иллюстрации: класс единиц одной частоты или длины.

Определение 3

Структурным классом называется лексический класс, элементы которого имеют одно или несколько структурных свойств.

Мы устанавливаем, что любая лексическая единица должна иметь однозначное значение любого структурного свойства. В результате становится возможным однозначно представить эту единицу в виде точки в лексическом пространстве, координаты которой задаются n – вектором $\langle l_1, l_2, \dots, l_n \rangle$ значений ее структурных свойств.

Определение 4

Пространство, которое определяется измерениями структурных свойств лексических единиц, называется лексическим пространством.

Предшествующее рассмотрение ограничивается областью синхронии; однако конструируемая модель должна включать и динамический аспект, диахронию. Изменения лексических единиц можно отразить, добавив к системе осей структурных свойств дополнительную временную ось. Тогда синхронное представление соответствует моментальному снимку лексического пространства в определенный момент времени t , а диахронное представление – схеме изменения структурного свойства с течением времени. Соответственно определяются динамические свойства.

Определение 5

Динамическими свойствами лексической единицы являются длительность, скорость и ускорение движения этой единицы в лексическом пространстве.

Лексические свойства связаны между собой многообразными способами, например, парадигматико-семантическими отношениями, вероятностями синтагматического сходства, деривацией, образованием сложных слов композицией, и многими другими способами.

Таким образом, множество лексических свойств является объектом, имеющим внутреннюю структуру. Снаружи этот объект относительно замкнут и автономен. Структура, которую мы в нем видим, является отражением наших теоретических знаний об отношениях элементов между собой. Одним словом, мы рассматриваем это множество как систему.

Определение 6

Множество лексических свойств с их отношениями между собой образуют лексическую систему. Все элементы связаны между собой таким образом, что не возникает никаких изолированных подмножеств.

Само собой разумеется, что в рамках определенной теории не все мыслимые отношения в системе естественно являются релевантными, так же как не все свойства, которые во всевозможных отношениях можно приписать лексической единице, представляют интерес. В качестве структурных элементов модельной системы учитываются таким образом только те отношения, которые взаимозависимы со структурными свойствами лексических единиц. Вследствие этого вся система может быть описана в любой момент времени путем указания всех значений всех свойств всех элементов.

Определение 7

Вектор структурных свойств лексических единиц системы в момент времени t называется состоянием системы в t .

Моделируемая таким образом лексическая система рассматривается как относительно автономная, но не как герметически замкнутая. Это значит, что она, как все системы, через одну или несколько точек соприкасается со своим окружением. Элементы внешней среды, с которыми система находится во взаимодействии, определяются как системное окружение.

Определение 8

Множество элементов, которые не рассматриваются в качестве части лексической системы, но могут вызывать изменение состояния системы или изменяться под действием системы, называется системным окружением.

Элементы системного окружения можно сгруппировать по двум критериям: по тем изменениям, которые они вызывают в системе (или по тем системным элементам, на которые они действуют), и по их роли в системном окружении. Элементы группы второго вида будем называть системными потребностями.

Системная потребность – это элемент (который сам может быть системой) системного окружения, который рассматривается с точки зрения его отношений с другими элементами вне системы и по крайней мере с одним элементом внутри рассматриваемой системы. В качестве примера можно назвать минимизацию затрат на понимание; этот теоретический конструкт связан с элементами таких областей, как физическая акустика, физиология органов чувств, обработка символов, с одной стороны, и сходство знаков, с другой стороны.

Определение 9

Под системной потребностью мы понимаем элемент системного окружения, который находится в отношении с другими элементами вне системы и может вызывать изменения в системе.

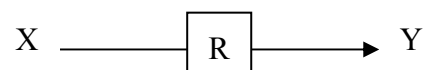
Распространенный (в первую очередь в социологии) термин "системная потребность" выбран очень неудачно и постоянно приводит к недоразумениям. Поэтому следует еще раз подчеркнуть, что такая потребность не принадлежит к системе, а является требованием, которое предъявляется к системе извне.

4.5 СИСТЕМНО-ТЕОРЕТИЧЕСКИЕ СРЕДСТВА: АЛГЕБРА ГРАФОВ И ЛИНЕЙНЫЕ ОПЕРАТОРЫ

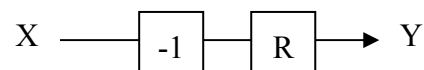
Форма представления, в которой подаются выше рассмотренные связи, еще не является достаточно точной, чтобы удовлетворить требования, предъявляемые к модели. Она создает возможность для неопределенности в отношении степени зависимости (до сих пор указывалось лишь направление зависимости) и с точки зрения взаимодействия зависимостей. (Что происходит, когда две или несколько стрелок указывают на одну и ту же величину?) Здесь в качестве теоретического средства может послужить общая теория систем: она базируется на множестве правил, с помощью которых из высказываний о структуре могут быть получены высказывания о функции. Эти правила образуют специальную алгебру: алгебру линейных графов и линейных операторов. Насколько это возможно, их представление будет соответствовать тем условностям, которые использованы в работе Cortes, Przeworski, Sprague (1974), дополненным выше понятием потребностей¹⁹ в нашей системе обозначений.

4.5.1 Операторы и графы

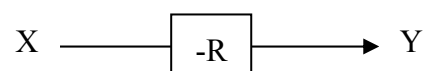
Утверждение, что величина X действует на другую величину Y, принимает в графической форме следующий вид:



Прямоугольником обозначается оператор пропорциональности, который указывает, в какой степени X действует на Y. Если нужно отразить обратно-пропорциональную связь, то это можно сделать или путем добавления оператора обратной пропорциональности

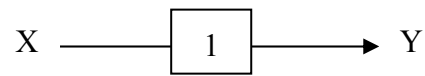


или просто с помощью отрицательного знака перед коэффициентом пропорциональности

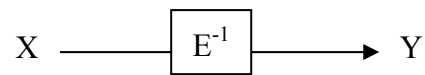


¹⁹ Ср. также раздел 4.2.

Для сокращения формы представления в дальнейшем мы по возможности не будем эксплицитно использовать оператор идентичности:



Еще один последний оператор – оператор запаздывания



который сообщает, что X действует на Y с запаздыванием.

4.5.2 Структура и функция

С помощью понятия операторов можно сформулировать следующие определения:

1. Структура системы – это множество ее связанных между собой операторов
2. Функция системы – это совокупность воздействий всех операторов системы.

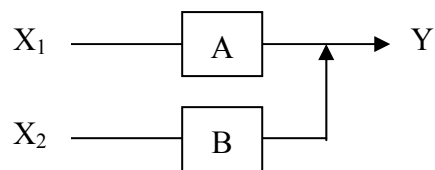
Для вывода функции системы или подсистемы из ее структуры алгебра графов устанавливает следующие правила:

1. При последовательном соединении операторы перемножаются. Из структуры

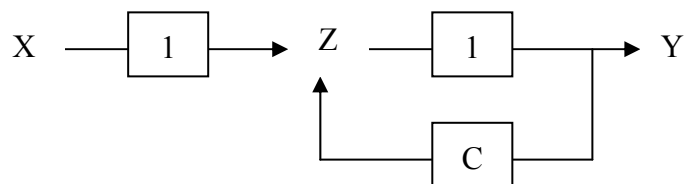


получается функция $Y = -ABX$

2. При присоединении стрелок происходит сложение. Структуре



соответствует функция $y = Ax_1 + Bx_2$. Этих двух правил уже достаточно для того, чтобы отсчитывать обратные связи. Структуре



соответствуют уравнения

$$y = 1 z$$

$$z = 1 x + cy$$

откуда

$$y = x + cy$$

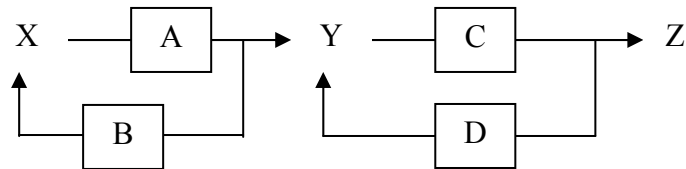
$$y - cy = x$$

$$y = \frac{x}{1-c}.$$

4.5.3 Упрощение графов

Отсчет функции сложных структур можно упростить, если сначала получить функции подструктур, а затем соединить полученные результаты между собой, как простые операторы.

Пусть дана структура



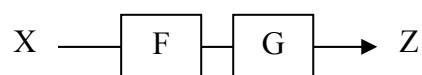
Функция подструктуры, соответствующая левой обратной связи, согласно уже известным правилам алгебры графов имеет следующий вид:

$$y = \frac{a}{1-ab} x$$

а правой подструктуре соответствует функция

$$z = \frac{c}{1-cd} y$$

Если эти две функции обозначить через F и G, то общая структура может быть представлена в следующем упрощенном виде:



Функция этой структуры имеет теперь простой вид:

$$z = fg x.$$

Если выше полученные функции подставить вместо F и G, то получим

$$z = \frac{ac}{(1-ab)(1-cd)}.$$

Аналогичным образом, путем постепенного упрощения можно легко отсчитывать и более сложные системы, в которых необходимо одновременно учитывать большое число входных и выходных величин.

Разумеется, такой вид представления очень сильно сокращен, но для наших целей он на данном этапе полностью достаточен. Подробное введение в алгебру графов и дополнительную литературу можно найти, например, в указанных источниках. В последующих рисунках системные величины и изображаются прямоугольниками, операторы – квадратами, а системные потребности – кружками. Системные потребности изображаются связанными через оператор с системной величиной или в виде интерпретации оператора. Во втором случае потребность (без стрелки) связывается с соответствующим оператором.

4.6 ЭЛЕМЕНТЫ, СТРУКТУРА И ФУНКЦИЯ ЛЕКСИЧЕСКОЙ СИСТЕМЫ

4.6.1 Инвентарь фонем

При разработке базовой модели рассмотрим сначала снова точку соприкосновения между фонемно-графемной системой и лексической системой. Оказывается, что эта точка совпадает с влиянием числа фонем²⁰ на длину слова. Действующие при этом потребности рассматриваются здесь с учетом уровня, когда это упрощает представление; на уровне фонем все потребности можно объединить в две противоположно направленные:

- минимизация затрат на кодирование (minK) и

- минимизация затрат на декодирование (minD)

Модифицируя и уточняя ранее предложенный подход, мы будем указывать не только направление, но и степень влияния этой потребности на число фонем. Чтобы избежать сметения инвентарей знаков, мы больше не будем использовать термин "объем инвентаря фонем"; отныне эта величина будет называться числом фонем. Кроме того, будет осуществлена следующая трансформация: чтобы модель можно было представить линейно, будет использоваться не само число фонем, а его логарифм, который в дальнейшем будет называться L-число фонем.

Потребности minK на этом уровне соответствует цифровский фактор унификации, который понижающим образом действует на L-число фонем. Этот фактор представляется в дальнейшем в виде оператора обратной пропорциональности – Y_2 (ср. рис. 7). В противоположность этому потребность minD соответствует цифровскому фактору диверсификации, который передается оператором Y_1 .

²⁰ Здесь и в дальнейшем всегда имеется в виду возможность рассматривать вместо фонем графемы, когда речь идет о числе фонем. Различие, проявляющееся между этими величинами в устной и письменной речи, можно выразить с помощью постоянного коэффициента.

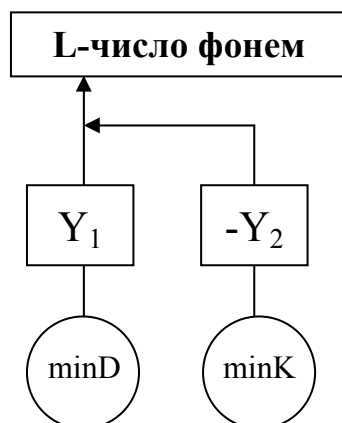


Рис. 7. L-число фонем вытекает из равновесия между факторами и диверсификации.

Непосредственное соединение стрелок означает, что суммарное влияние, действующие на величину L-число фонем, следует вычислять путем сложения двух противоположно направленных факторов²¹. Согласно правилам алгебры графов можно поэтому записать: L-число фонем = $Y_1 \min D - Y_2 \min K$, причем для определенного числа фонем существует бесконечное число значений для двух членов в правой части уравнения, удовлетворяющих этому условию. Поэтому без дополнительной информации числовое определение операторов на данном этапе невозможно.

4.6.2 Объем словаря

Еще одним параметром, определяющим другие элементы лексической системы, является величина словаря. Она непосредственно связана с элементарной системной потребностью:

- потребностью в кодировании (Kod),

т.е. необходимостью выражать несколько различных значений. Чем больше требуется кодировать значений, тем больше должно быть число лексических единиц в языке.

Для линеаризации эта величина тоже трансформируется; в модели используется L-величина словаря (см. рис. 8).

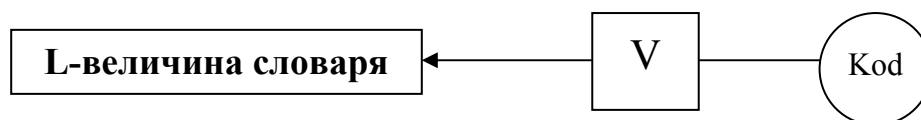


Рис. 8. Потребность кодирования действует через оператор пропорциональности V на L-величину словаря

²¹ Аналогичную линейную связь Альтман (1985) устанавливает для соответствующих факторов при моделировании семантической диверсификации стохастическими методами.

Для выражения влияния потребности кодирования на величину словаря используется не оператор идентичности, а оператор пропорциональности, поскольку отношение кодируемых значений к числу лексических единиц никогда (кроме искусственных языков) не составляет 1:1 (см. также 4.6.4)

4.6.3 Длина

В качестве первого системного элемента, соответствующего структурному свойству лексических единиц, мы вводим фундаментальную величину "длина". Длина языкового выражения может измеряться многими различными способами, в зависимости от того, какая выбирается единица отсчета, и какие аспекты такого выражения считаются релевантными в рамках модели, лежащей в основе исследования. Обычно в лингвистике под длиной понимают число составляющих элементов: в зависимости от модели это фонемы, графемы, слоги²², морфемы или слова. Однако, во многих случаях длина рассматривается также как временная длительность акустического сигнала²³. В нашей базовой модели за длину лексической единицы принимается число составляющих его фонем. Однако, чтобы можно было представить отношения переменной "длина" к другим элементам системы в линейной форме, осуществляется следующая трансформация:

Определение 10

L-длина лексической единицы – это логарифм числа составляющих ее фонем.

L-длина определена на интервале $(0, \infty)$, поскольку наименьшая длина языкового выражения составляет 1 фонему ($\log 1 = 0$), а теоретический верхний предел указать невозможно. В отличие от L-числа фонем и L-величины словаря L-длина не характеризует лексическую систему в целом, а представляет собой свойство, которое имеется у любой лексической единицы. Поэтому следует точнее определить, что имеется в виду под зависимостью L-длины от двух других величин.

Как уже указывалось, длина элементов лексики зависит от числа элементов на фонемном уровне. Это утверждение ни коим образом не касается определенного отдельного лексического элемента: конкретное слово не становится короче от того, что в инвентаре языка появляется новая фонема. Лучше исходить из того, что здесь наблюдается стохастическая связь между числом фонем и зависящей от нее по законам комбинаторики средней длины выражения. Отношения аналогичного типа существуют между величиной словаря и длиной выражения: чем больше требуется лексических единиц, тем длиннее они должны быть в среднем (при постоянном числе фонем)²⁴ x24].

Таким образом, системные величины L-число фонем и L-величина словаря оказывают глобальное стохастическое влияние на L-длину лексической единицы, об индивидуальном воплощении которой, отклонении ее от среднего значения, еще ничего не было сказано. Как мы знаем, ни один естественный язык не исчерпывает

²² Слог – это понятие наблюдения, связанное со звуковым представлением выражений. При изучении свойств слогов на письменных текстах, как это обычно и делается, необходимо эксплицитно показать соотношение между слогами и комбинациями графем.

²³ Ср. напр. Fónagy, Magdics (1960).

²⁴ Помимо этого отношения следует учитывать в соответствующих случаях тоны и ударение.

комбинаторно возможный набор форм для построения лексических единиц. Как правило, степень реального использования потенциалов очень мала. В качестве примера можно привести перечень потенциально возможных последовательностей из трех фонем /t/, /n/, и /u:/:

/tt/ /tu:t/ /tnt/
 /ttu:/ /tu:u:/ /tnu:/
 /ttn/ /tu:n/ /tnn/

 /u:tt/ /u:u:t/ /u:nt/
 /u:tu:/ /u:u:u:/ /u:nu:/
 /u:tn/ /u:u:n/ /u:nn/

 /ntt/ /nu:t/ /nnt/
 /ntu:/ /nu:u:/ /nnu:/
 /ntn/ /nu:n/ /nnn/

В современном (литературном) немецком языке для построения морфем используются только четыре подчеркнутые последовательности фонем. Можно даже сформулировать типичные для конкретного языка фонотактические правила, описывающие принципы выбора в этом языке. Основание для этого ограничения очевидно и давно известно: для повышения надежности передачи и для снижения опасности возникновения недоразумений любая естественная коммуникационная система имеет соответствующую требованиям избыточность. Поэтому здесь следует принять в качестве постулата системную потребность

- обеспечение надежности передачи (Red),

которая оказывает прямо пропорциональное воздействие на L-длину (рис. 9).

В качестве функции подсистемы, состоящей из введенных до сих пор элементов, L-длину можно представить по правилам алгебры графов следующим образом:

$$\log(l) = A \log(g) + Z \text{Red} - P \log(p)$$

где $\log(l)$ – L-длина, $\log(g)$ – L-величина словаря, а $\log(p)$ – L-число фонем.

При использовании нелогарифмических величин длина (l), величина словаря (g) и число фонем (p) мы получаем уравнение

$$l = \frac{g^A \text{red}^Z}{p^P} \quad \text{или} \quad l = g^A \text{red}^Z p^{-P}.$$

Следует еще раз напомнить, что при учете только глобальных влияний двух рассматриваемых системных параметров длина лексической единицы еще не определяется полностью.

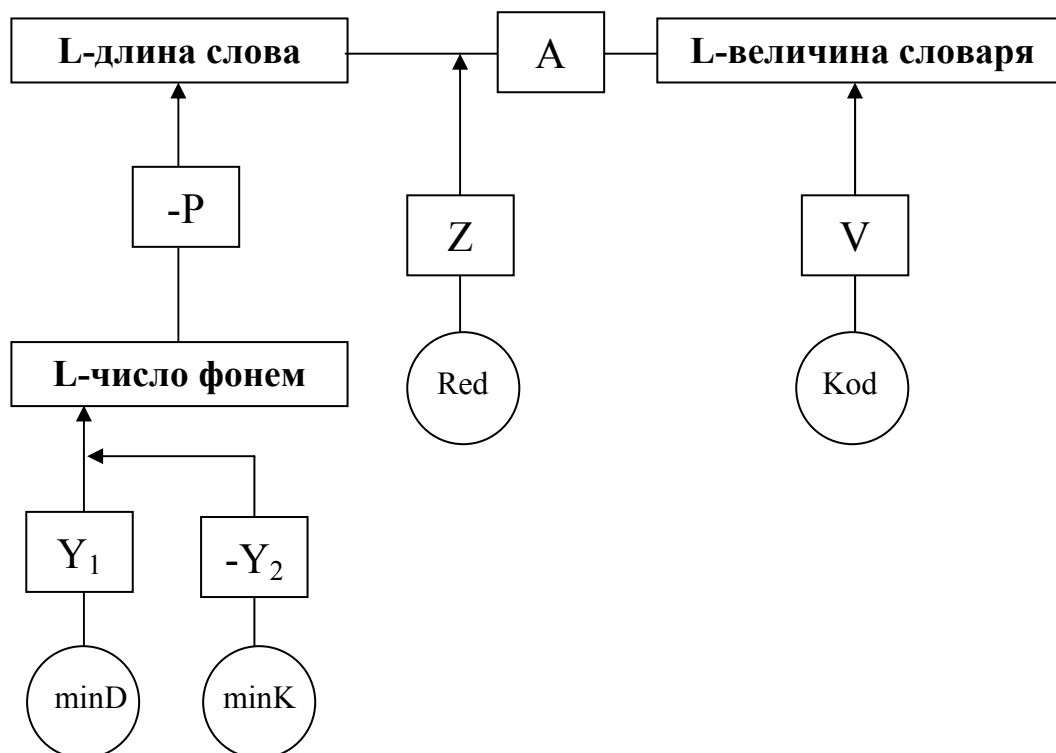


Рис. 9. Средняя L-длина определяется двумя противоположно действующими системными величинами L-величиной словаря и избыточностью, с одной стороны, и L-число фонем, с другой стороны.

4.6.4 Полилексия

Общим свойством всех естественных языков является их многозначность, которая наблюдается на всех уровнях анализа: существует фонологическая, морфологическая, синтаксическая, семантическая и другие виды многозначности. Поэтому и лексические единицы чаще всего не имеют единственного, четко очерченного значения.

Принять ли точку зрения, что значение языкового знака существует только при его использовании в соответствующем контексте или считать значения реальностями, которые существуют и отдельно от употребления, – в обоих случаях значение является абстракцией большого числа ситуативных и контекстуальных условий применения.

В результате мы получаем или самостоятельные конструкты, или более или менее обширные наборы ситуаций и контекстов, в которых может быть реализовано значение рассматриваемого выражения. Число различных значений, которые лексическая единица имеет в определенный момент времени (в работах Altmann, Веóthy, Best 1982 – сложность значения), следует назвать полилексией этой единицы. В этом понятии не должно проводиться различие между семантическими и грамматическими значениями, с той целью чтобы и функциональные слова имели полилексию больше нуля.

Таким образом, среднее значение полилексии лексической единицы лежит в интервале $<1, \infty$, а значение его логарифма – L-полилексии – в интервале $<0, \infty$). Очевидно, что

средняя полилексия влияет на величину словаря: чем больше значений имеет выражение, тем меньше выражений должно быть в лексике (при постоянном числе кодируемых значений). Здесь снова возникает потребность в минимизации затрат на кодирование; унификация действует в направлении увеличения полилексии.

В экстремальном случае можно представить себе лексику, состоящую из одного единственного слова, которое при необходимости может иметь все значения, выражаемые отправителем. Но и здесь существует действующая в противоположном направлении потребность в минимизации затрат на декодирование: текст тем легче декодировать, чем более однозначны лексические единицы. Диверсификация действует в направлении уменьшения полилексии. Этому соответствует гипотетический экстремальный случай, когда в лексике для каждого значения есть свое отдельное лексикализованное выражение.

Эмпирически наблюдаемые значения полилексии в различных языках являются результатом равновесия между действующими при этом влияниями.

Мы снова исходим из допущения, что факторы (логарифмически трансформированные) складываются; для отражения влияния L-полилексии на L-величину словаря вводится оператор пропорциональности -L (ср. рис. 10).

Если теперь определить влияние потребности кодирования как число (тоже логарифмическое) кодируемых значений, то становится возможным численно определить коэффициенты пропорциональности L и V: если в наборе правил не требуется учитывать еще какие-то влияния, то оба коэффициента имеют значение 1.

Пример: предположим, что язык различает в определенный момент времени 100000 значений. Если каждая лексическая единица имеет в среднем 4 значения, то требуется инвентарь в 25000 слов, поскольку

$$\text{Величина словаря} = \text{Число значений} / \text{Полилексия.}$$

При логарифмических переменных, с подстановкой 1 вместо L и V, получаем:

$$\text{L-величина словаря} = \text{Kod} - \text{L-полилексия}$$

Связь между числом значений и другими величинами была замечена еще раньше: Ципф (1949) предложил зависимость этого числа от частоты слова, а Guiter (Гитэ 1974) опубликовал эмпирические функции между частотой и числом значений, длиной и числом значений для нескольких романских языков и английского языка. Однако, эти авторы не дали интерпретацию своих результатов с лингвистической точки зрения и не предложили теоретического вывода. Убедительный вывод был впервые предложен Альтманом²⁵.

Допустим, что определенное выражение языка имеет какое-то значение полилексии больше 1 и существует коммуникативная ситуация, в которой возникающая в результате этой полилексии многозначность нежелательна: мы постулируем

- потребность в уточнении (Spz).

²⁵ Ср. Altmann 1980, Altmann, Веóthy, Best 1982

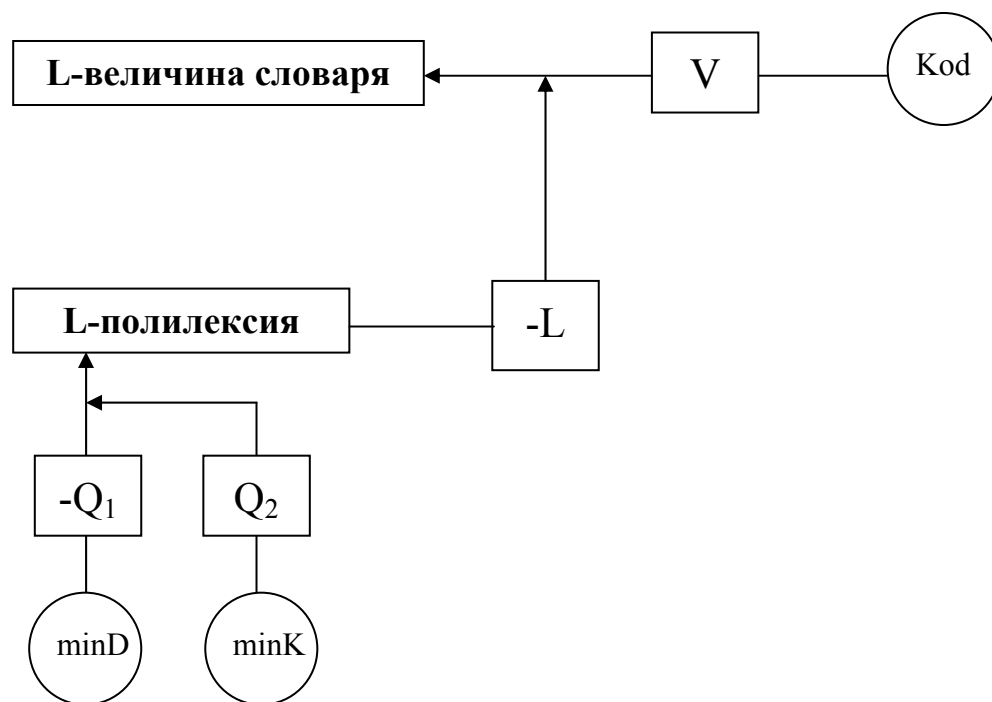


Рис. 10. Средняя полилексия и потребность кодирования совместно определяют величину словаря

Для удовлетворения этой системной потребности естественные языки имеют различные средства (функциональные эквиваленты): спецификация так же как и модификация может быть достигнута с помощью синтаксических (напр., "синтаксическое средство" в противоположность к "статистическому средству") или морфологических средств, например соединения (Komposition) и аффиксации (например, "Hilfsmittel" в отличие от "Mittelwert"). Во всех случаях задействована длина получаемого выражения²⁶; но из-за относительной редкости лексикализованных синтагм по сравнению со словами при практических измерениях на длину влияет почти только морфологическое решение. Это значит, что зависимость полилексии от длины тем сильнее, чем больше морфологическими средствами по сравнению с синтаксическими пользуется язык для спецификации значения. Это типологическое свойство языка следует назвать синтетичностью, обозначив ее буквой Т. О связи величин мы делаем следующие допущения:

- а) Скорость изменения полилексии обратно пропорциональна длине
- б) Коэффициент пропорциональности определяется только синтетичностью языка

Если обозначить полилексию через y , ее первую производную через y' , а длину через x , то при учете синтетичности Т можно составить дифференциальное уравнение, которое соответствует этим допущениям:

²⁶ Влияние на длину может быть модифицировано или даже устранено с помощью дополнительных связей, например, когда в языке используются тоны или ударения.

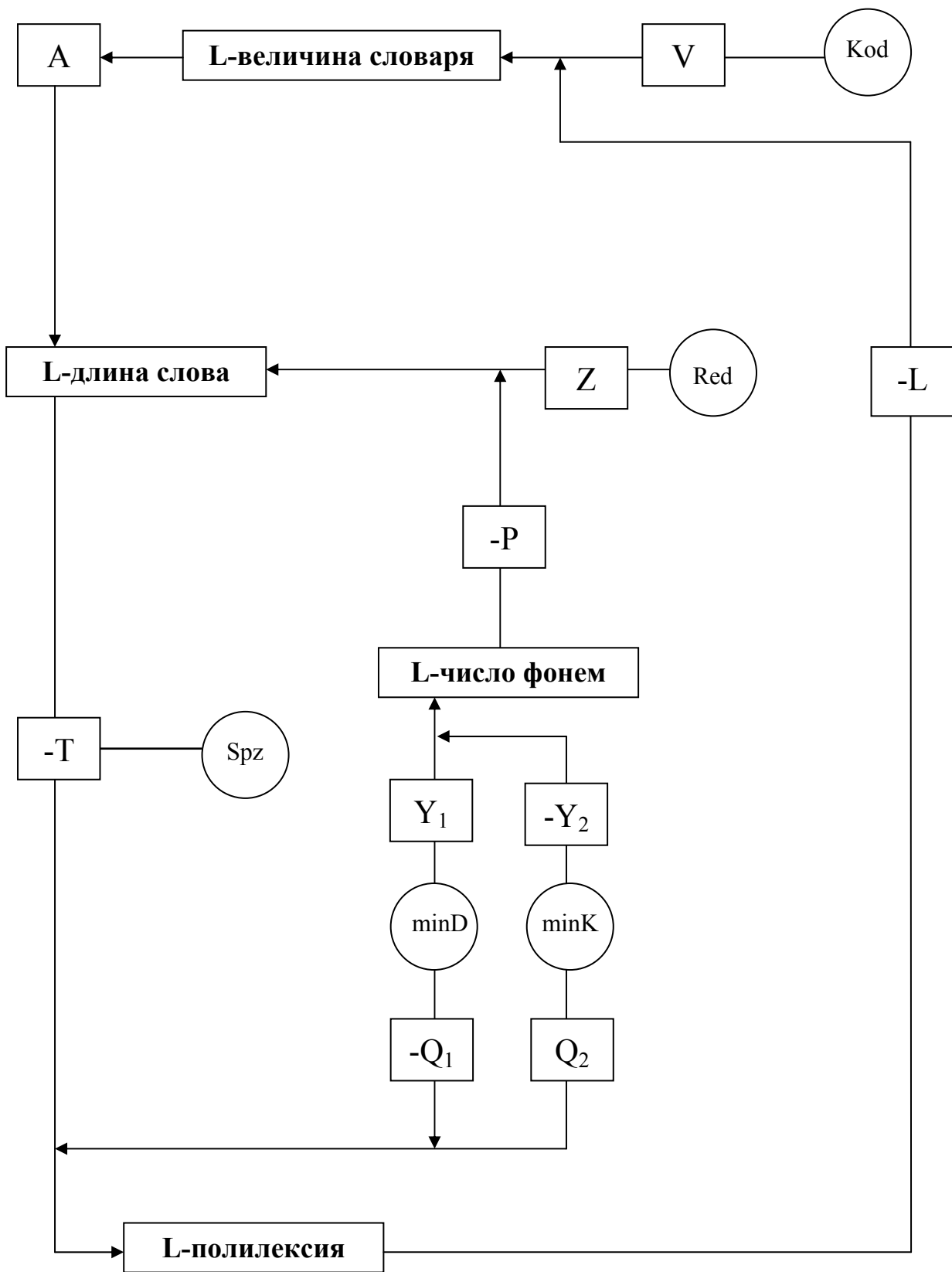


Рис. 11. В этом наборе правил *L*-полилексия представлена в зависимости от *L*-длины и глобального влияния потребностей *minD* и *minK*.

$$\frac{y'}{y} = \frac{-T}{x}$$

Это дифференциальное уравнение имеет решение

$$y = Cx^{-T}$$

где C – произвольная константа. Для переменных модели уравнение логарифмируется, откуда следует

$$L\text{-полилексия} = C - T L\text{-длина}$$

при $c = \ln C$. Из рассуждений в последнем разделе величина, вводимая здесь как C , уже известна; речь идет о глобальном влиянии равновесия между унификацией и диверсификацией. Поэтому полностью уравнение записывается в следующем виде:

$$L\text{-полилексия} = Q_2 \min K - Q_1 - T L\text{-длина}.$$

На рисунке 11 это отношение показано в связи с другими отношениями, рассмотренными ранее.

4.6.5 Политекстия

Значение лексической единицы абстрагировано от конкретных ситуаций употребления в речи — во всяком случае в такой степени, что она может использоваться в значительной степени независимо от контекста. Было бы в высшей степени не экономично, если бы в языке лексикализовались такие значения как $\text{дом}_{t,o,s}$ для обозначения определенного дома в момент времени t в месте o в ситуации s . Контекстуально-специфические компоненты значения в естественных языках обычно кодируются дополнительно или разгадываются реципиентом. Это позволяет сохранять конечность величины словаря. Поэтому следует постулировать соответствующую потребность:

- экономию контекста (КО)

Способность лексических единиц применяться независимо от контекста назовем политекстией. Она определяется как метрическое понятие:

Определение 11

Политекстия лексической единицы – это число контекстов, в которых она используется.

Экономия контекста действует в направлении повышения средней политекстии; соответствующий процесс мы называем глобализацией. Эта потребность имеет, естественно, своего контрагента. Во многих случаях для оптимальной коммуникации требуется настолько большая однозначность, что

а) лексическая единица получает дополнительное и недвусмысленное, контекстуально-специфическое значение (напр. "группа", "семейство", "кольцо", "тело" в математике),

б) создается новый термин, имеющий значение только в своем исходном контексте (напр. "лазер", "полилексия"),

в) импортируется иностранное выражение или снова вводится устаревшее или редко употребляемое слово из общеязыкового словаря (напр. "токамак", "клад").

Мы будем называть эту системную потребность

- контекстным уточнением (KS)

Эта потребность действует на политекстию в направлении, обратном влиянию потребности в экономии контекста, так что в результате равновесия между глобализацией и этим процессом в системе устанавливается средняя, оптимальная политекстия. Это представлено на рисунке 12 в виде линейного графа, в котором используется аддитивность (логарифмических) влияний. В качестве компонента системы мы снова используем логарифмически преобразованную величину: L-политекстию

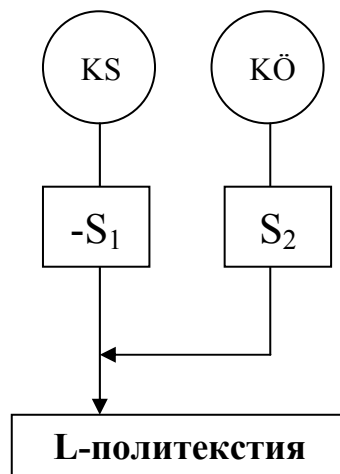


Рис. 12. Зависимость средней L-политекстии от потребностей KS и KÖ

Далее, политекстия распределена по лексическим единицам языка неравномерно. Так, выражение "иметь" имеет очень высокое, выражение "светлый" – менее высокое, а выражение "контрапункт" – очень низкое значение политекстии. Поэтому потребуются ввести отношение, которое отвечает за отклонения значений политекстии конкретных лексических единиц от средней политекстии. Представляется разумным допустить, что число контекстов, в которых выражение может быть использовано, в первую очередь зависит от гибкости его значения. Слово, имеющее только одно единственное узкое значение, не найдет применения в таком большом числе коммуникативных ситуаций, как слово, имеющее более высокую семантическую потенцию. Эту связь можно пояснить с помощью гипотетического примера, представленного в следующей таблице:

	Слово 1	Слово 2	Слово 3	Слово 4	Слово 5
контекст 1	3	5	2	2	2
контекст 2	1	3	1	1	
контекст 3	1	2		1	
контекст 4	1	1			
контекст 5		1			
Сумма значений	6	12	3	4	2
Число контекстов	4	5	2	3	1

Эту гипотезу мы сформулируем следующим образом:

Политекстия лексической единицы изменяется пропорционально ее полилексии.

Если обозначить политекстию через Y , полилексию через X , а коэффициент пропорциональности через G , то эту гипотезу можно представить в виде дифференциального уравнения

$$\frac{y'}{y} = \frac{G}{x}$$

имеющего решение

$$y = Cx^G.$$

В логарифмическом виде

$$L\text{-политекстия} = c + G L\text{-полилексия}$$

при $c = \ln C$. В этом уравнении неизвестную величину c нетрудно интерпретировать как влияние равновесия KS и $K\ddot{O}$:

$$L\text{-политекстия} = S_2 K\ddot{O} - S_1 KS + G L\text{-полилексия},$$

что точно соответствует функции, которая получается по правилам алгебры графов из структуры, представленной на рисунке 13.

4.6.6 Частота

Частотой лексической единицы называют частоту ее использования в речи. Она тем больше, чем чаще соответствующая лексическая единица в среднем используется говорящими, чтобы выразить определенное значение. Это справедливо независимо от того, используется ли говорящим или пишущим уже лексикализованное значение данного выражения, придается ли лексической единице в актуальной ситуации новое, может быть, только окказиональное значение или создается совершенно новое выражение. Частота зависит поэтому главным образом от того, как часто одно из значений, которое может иметь единица, становится коммуникативно релевантным.

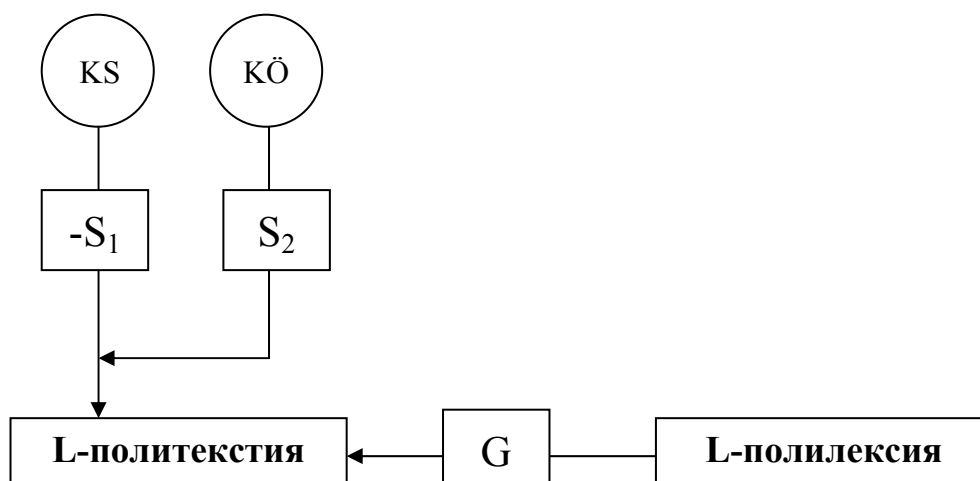


Рис. 13. L-полилексия влияет на L-политекстию в степени G, так что лексические единицы имеют индивидуальные значения L-политекстии.

Теперь введем системную потребность под названием

- потребность в использовании (Anw).

Эта потребность действует на (трансформированную) L-частоту лексической единицы в той степени, которую задает оператор пропорциональности K.

Вероятность того, что определенное выражение подходит для данной коммуникативной ситуации не является, однако, постоянной, а зависит от ситуации или чаще всего от контекста, и притом в тем большей степени, в скольких дополнительно различных контекстах выражение может быть использовано. Поэтому можно предположить, что скорость изменения L-частоты зависит от L-политекстии, причем пропорционально коэффициенту K. Этой гипотезе соответствует дифференциальное уравнение

$$\frac{y'}{y} = \frac{K}{x}$$

где y – частота, а x – политекстия. Его решение

$$y = Cx^K$$

позволяет получить логарифмическое уравнение

$$\text{L-частота} = c + K \text{ L-политекстия}$$

Если вместо $c = \ln C$ подставить влияние потребности использования, то мы получим функцию

$$\text{L-частота} = R \text{ Anw} + K \text{ L-политекстия},$$

которая получается из структуры на рис. 14.

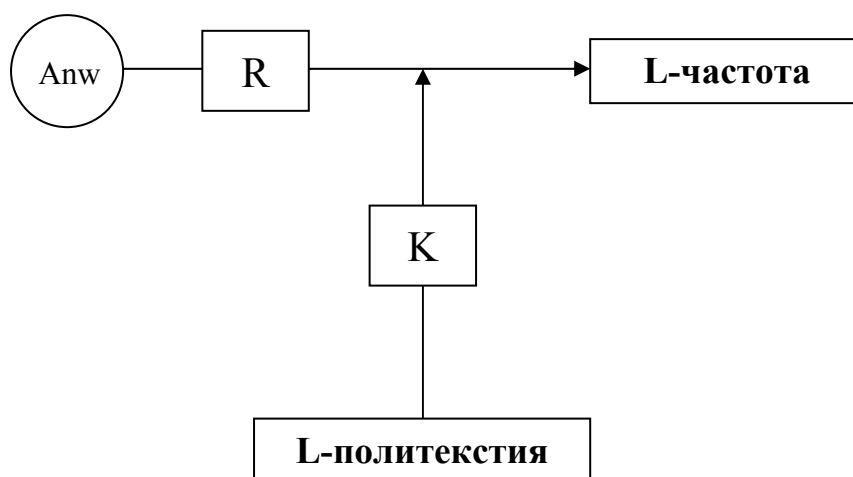


Рис. 14. Зависимость L-частоты от потребности в использовании и L-политекстии

4.6.7 Длина (продолжение)

Выше уже названы величины, глобально определяющие среднюю длину лексических единиц: величина словаря, число фонем и избыточность. В этом разделе введем в модель отношение, которое отвечает за изменчивость длины между индивидуальными лексическими единицами.

Связь между длиной и частотой слов, по-видимому, самое известное количественное отношение между двумя лингвистическими свойствами (см. Zipf 1949; Guiter 1974; Herdan 1966). Ципфовская гипотеза об этой связи заключается в предположении, что при частом употреблении слова укорачиваются. Mandelbrot (1953) использовал отношение частоты и длины для вывода частотно-рангового распределения, известного под названием закона Ципфа-Мандельброта, причем в этом выводе длина слова (проинтерпретированная как "затраты") минимизировалась в среднем. В противоположность этому в нашей модели связь между этими двумя величинами служит для объяснения индивидуальных значений длины.

Очевидно, что длина лексических единиц зависит от их частоты приблизительно так, как это представлял Ципф. Когда слово "автомобиль" вошло в лексику ряда языков, оно было еще очень редким (из-за нераспространенности автомобилей десигнат этого нового слова часто еще не был коммуникативно релевантным). С увеличением частоты (из-за роста потребности использования слова "автомобиль" вследствие все большего распространения автомобиля как транспортного средства) это выражение во всех языках значительно сократилось: нем. "Auto", сканд. "bil", индон. "mobil" и т.д.

Однако существует не только процесс сокращения; при уменьшении частоты имеется тенденция снова удлинять короткие выражения или заменять их более длинными (Lüdtke 1980) наряду со многими другими примерами приводит современное французское выражение "сегодня": "aujourd'hui", которое заняло место короткого "hui" (из лат. "hodie"). Поэтому правило лучше сформулировать следующим образом: каждая лексическая единица имеет длину, которая соответствует ее частоте. Эту зависимость мы рассматриваем как функцию еще одной системной потребности и постулируем поэтому минимизацию затрат на порождение (minP) в качестве причины этого. Гипотезу о том, что скорость изменения длины обратно пропорциональна частоте лексической единицы мы представляем в виде дифференциального уравнения

$$\frac{y'}{y} = \frac{-N}{x}$$

где y – длина, x – частота, а $-N$ – коэффициент пропорциональности потребности $\min P$.

Решение данного уравнения

$$y = Cx^N$$

после логарифмирования приобретает вид

$$\text{L-длина} = c - N \text{ L-частота.}$$

Неизвестную величину $c = \ln C$ легко проинтерпретировать как глобальное влияние на среднюю длину, от которой в зависимости от своих частот отклоняются индивидуальные длины лексических единиц. Таким образом, полное функциональное уравнение имеет вид

$$\text{L-длина} = A \text{ L-величина словаря} + Z \text{ Red} - P \text{ L-число фонем} - N \text{ L-частота}$$

или

$$\text{Длина} = (\text{Величина словаря})^A \text{ Red}^Z (\text{Число фонем})^{-P} \text{ Частота}^{-N}$$

Такую же функцию можно получить из структуры на рис. 15.

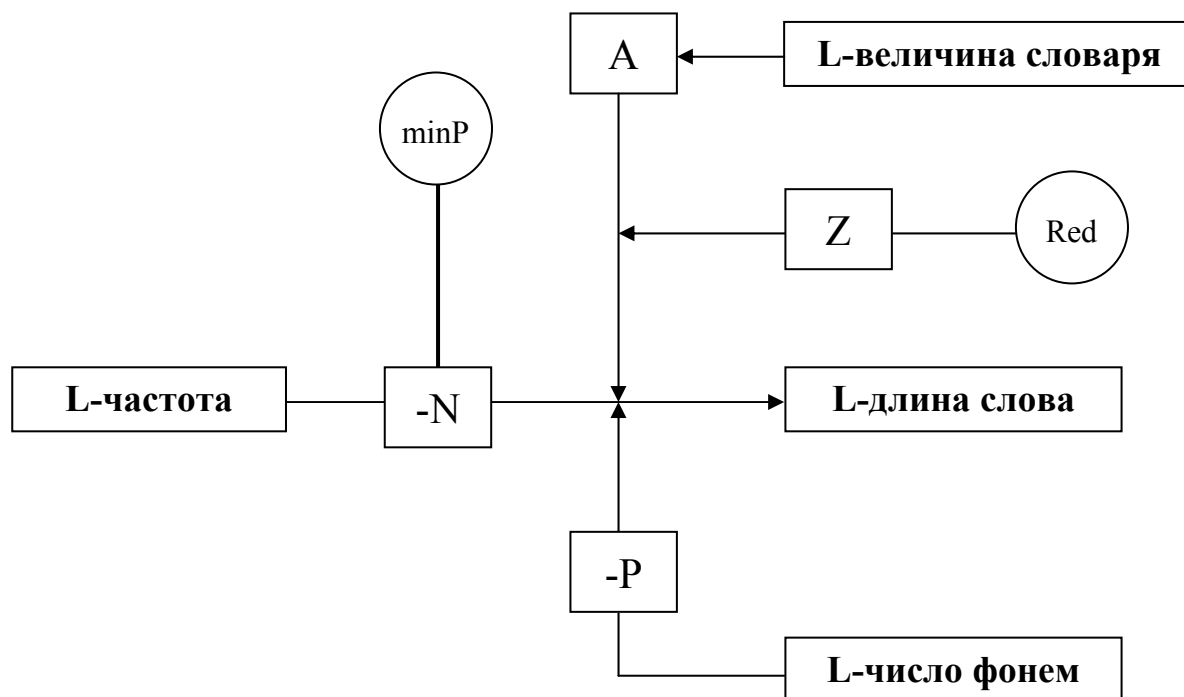


Рис. 15. Длина лексических единиц определяется балансом четырех системных величин и двух потребностей.

Между длиной и частотой можно установить и другие имеющие под собой основу отношения, отражающие более сложную связь. В частности, можно допустить, что сокращение выражения зависит также от его текущей длины. Такое слово как "Auto" при дальнейшем увеличении частоты не сократится в такой степени, как "Automobil". Это предположение соответствует структуре, представленной на рис. 16. Оператор запаздывания E^{-1} означает, что на сокращение влияет та длина выражения, которую выражение имело "за мгновение" до сокращения.

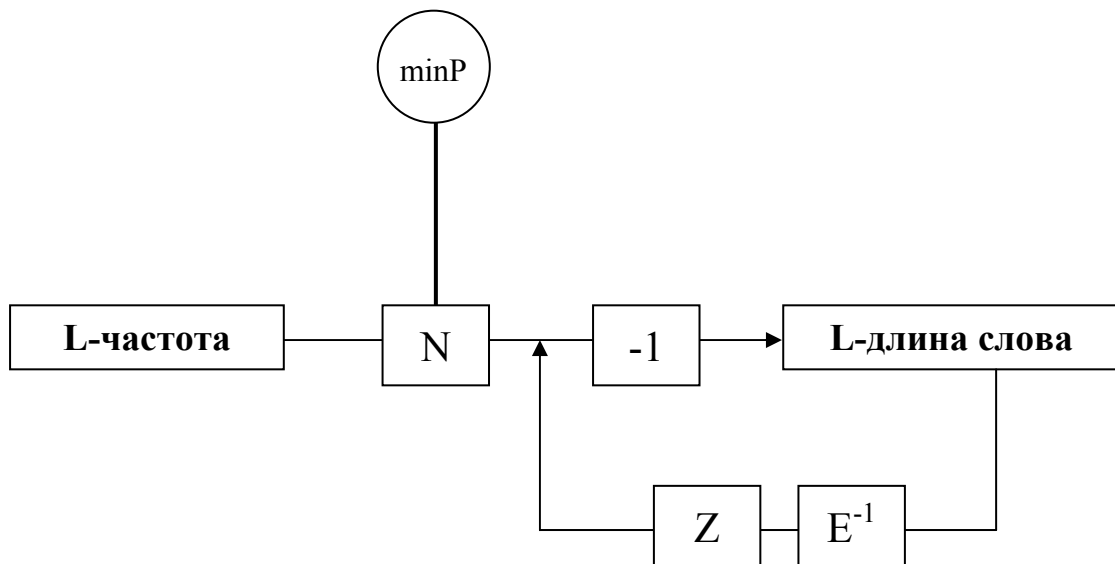


Рис. 16. Обратное влияние длины на сокращение.

Еще одно дополнительное допущение могло бы заключаться в учете влияния длины на частоту: из двух выражений, подходящих с точки зрения значения и контекста чаще используется более короткое. Соответствующая структура представлена на рисунке 17.

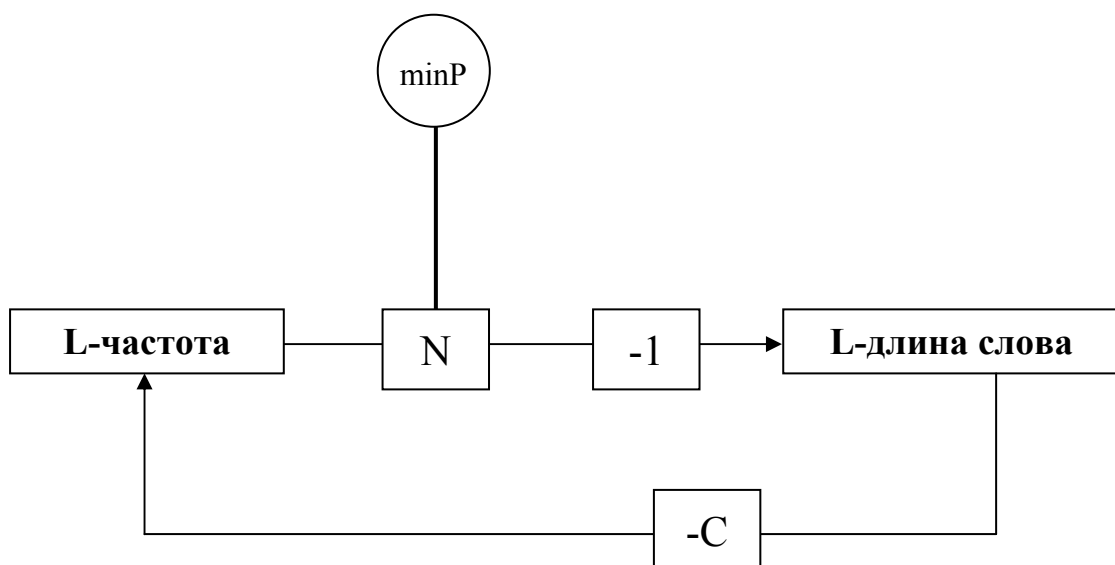


Рис. 17. Обратное влияние длины на частоты.

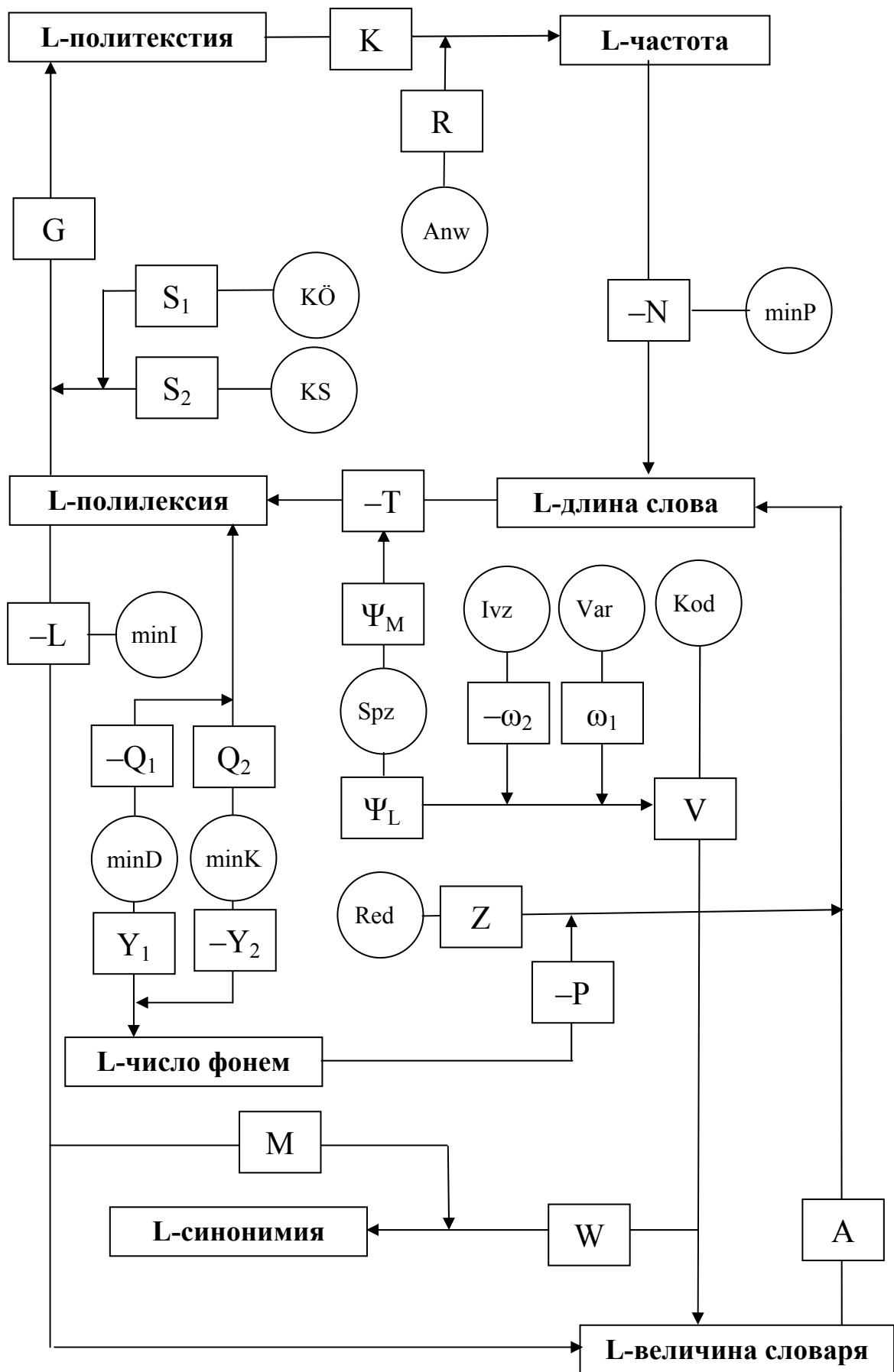


Рис. 18. Целостная структура базовой модели

Вопрос о том, позволяют ли эти или последующие уточнения получить более реалистическую модель языковой действительности, на данном этапе исследований вряд ли можно решить. Поэтому мы отказываемся от всех возможных уточнений модели до тех пор, пока ее существенные структурные черты не пройдут эмпирической проверки (см. ниже).

На рис. 18 представлена целостная структура базовой модели в ее нынешнем виде.

4.7 ФУНКЦИИ ПОДСТРУКТУР

Из подструктур модели для всех четырех структурных свойств лексических единиц выводятся функции вида

$$y = Ax^B$$

которым соответствует решение дифференциального уравнения

$$\frac{y'}{y} = \frac{B}{x}.$$

Для всех зависимостей между системными величинами, выражающих структурные свойства, здесь также использована эта форма дифференциального уравнения²⁷. Эти четыре уравнения составляют ядро модели. Две переменные – величина словаря и число фонем – получены из комбинаторных соображений или из величин, находящихся за пределами модели, которые поэтому сейчас можно рассматривать как эмпирические параметры.

4.8 ПОТРЕБНОСТИ, ВЕЛИЧИНЫ И ПРОЦЕССЫ

Ниже представлены в абличной форме системные потребности и их влияние на систему (см. таблицу 1).

В таблице 2 представлены системные величины с их функциями, в таблице 3 – процессы в модельной системе, непосредственно связанные с потребностями.

²⁷ В зависимости от знака перед константой В уравнения выражают прямую или обратную пропорциональность.

Таблица 1. Общий обзор системных потребностей

Потребность	Действует на величину	В направлении
Минимизация затрат на кодирование (min K)	число фонем полилексия	- +
Минимизация затрат на декодирование (min D)	число фонем полилексия	+ -
Надежность передачи (Red)	длина	+
Кодирование (Kod)	величина словаря	+
Уточнение (Spz)	полилексия	-
Экономия контекста (KO)	политекстия	+
Контекстное уточнение (KS)	политекстия	-
Использование (Anw)	частота	+
Минимизация затрат на порождение (min P)	длина	-
Минимизация величины инвентаря (min I)	величина словаря	-

Таблица 2. Системные величины и функции

Системная величина	Функция
Величина словаря	$LG = KOD^V PL^{-L}$
Число фонем	$PH = \min D_1^y \min K_2^{-y}$
Длина	$L = LG^A Red^Z PH^{-P} F^{-N}$
Полилексия	$PL = \min K_2^{Q_2} \min D_1^{-Q_1} L^{-T}$
Политекстия	$PT = KO^{S_2} KS^{-S_1} PL^G$
Частота	$F = Anw^R PT^K$

Таблица 3: Процессы

Процесс	Обслуживаемая потребность	Результат
Фонологическая унификация	minK	Уменьшение числа фонем

Фонологическая диверсификация	minD	Увеличение числа фонем
Фонологическое ограничение	Red	Увеличение длины
Лексикализация	Kod	Увеличение размера словаря
Лексическая унификация	minK	Увеличение полилексии
Лексическая диверсификация	minD	Уменьшение полилексии
Лексическая редукция	minI	Уменьшение величины словаря
Уточнение	Spz	Уменьшение полилексии
Глобализация	KÖ	Увеличение политекстии
Централизация	KS	Уменьшение политекстии
Использование	Anw	Увеличение частоты
Сокращение	minP	Уменьшение длины

5. ПОСЛЕДСТВИЯ, ВЫТЕКАЕМЫЕ ИЗ МОДЕЛИ

Из модели можно вывести ряд следствий, которые используются для ее эмпирической проверки²⁸. Некоторые из них имеют самостоятельное теоретическое значение.

5.1 ПРЯМЫЕ СВЯЗИ

Выше из дифференциальных уравнений получен общий вид зависимости между переменными. Из этого следует, что базовая модель не только описывает, но и объясняет и предсказывает связь между наблюдаемыми величинами. Объяснение заключается в дедуктивном выводе функций²⁹ из системной структуры, т.е. из гипотез, законов, предложенных выше; функции в свою очередь объясняют связи, наблюдаемые между данными. Поскольку законы (гипотезы, претендующие на статус законов) являются всеобщими высказываниями, из модели следует прогноз, что эти связи переменных соответствуют выведенным функциям во всех языках и во все времена, пока выполняются граничные условия.

Это открывает уже прекрасные возможности для теоретической и эмпирической проверки модели. Использованный подход, однако, позволяет извлечь еще более

²⁸ Ср. главу 6

²⁹ См. таблицу 2

далеко идущие следствия для выводимых функций: параметры функций тоже выводятся из структуры системы; в результате они задают определенную интерпретацию эмпирически найденным значениям параметров; в некоторых случаях уже сейчас возможен прогноз о значениях параметров. Функция

$$PL = QL^{-T}$$

соответствует³⁰ прогнозу, что в любом языке с переменной длиной выражений

- а) полилексия выражения зависит от его длины,
- б) форма этой зависимости гиперболическая,
- в) изменение длины в тем большей степени действует на поли лексию, чем синтетичнее язык.

Первое высказывание имплицитно содержит утверждение о направлении зависимости. Поэтому речь идет не о простой корреляции, а о квазикаузальной функциональной связи. Последнее высказывание – это уже количественный³¹ прогноз о параметре N, что позволяет подвергнуть модель очень существенному эмпирическому испытанию на фальсификацию.

Если модель найдет подтверждение, то с помощью утверждения в) станет также возможным уже наоборот использовать параметр N в качестве количественной меры степени синтетичности. Аналогичные соображения применимы к другим функциональным параметрам.

5.2 ОПОСРЕДОВАННЫЕ СВЯЗИ

Системно-теоретические средства позволяют также вывести из модели опосредованные связи переменных. Можно, например, задаться вопросом, как зависит полилексия выражения от его частоты. Для этого берем два функциональных уравнения: то, в котором полилексия выступает как зависимая переменная, и другое, в котором частота выступает как независимая переменная.

В любой момент времени влияние потребностей постоянны, то же можно сказать о величине словаря и числе фонем, а также операторах. Поэтому мы кратко записываем

$$PL = QL^{-T} \text{ и}$$

$$L = RF^{-N}.$$

³⁰ Символ Q выражает влияние равновесия между действием потребностей minK и minD.

³¹ В первую очередь на ординальной шкале. После углубленных типологических исследований станут возможны и численные прогнозы: путем сравнения большого числа выборок из различных языков и типологических признаков этих языков можно найти факторы, отвечающие за значение параметра N. Если в результате этого удастся определить точную связь между ними и N, то станет возможно закономерно вывести из этого вид зависимости PL(L).

Здесь константа Q обозначает $\min K^Q_2 \min D^{-Q}_1$, а константа R – выражение $LG^A Red^Z PH^P$. Чтобы ответить на вопрос об опосредованной зависимости поли лексии от частоты, мы исключаем промежуточную переменную "длина", подставляя правую часть второго уравнения вместо независимой переменной L:

$$PL = Q(RF^{-N})^{-T} \\ = QR^{-T} F^{NT}$$

При $S = QR^{-T}$ и $Z = NT$ получаем

$$PL = SF^{TZ}.$$

Если исходные уравнения отражают монотонно убывающие гиперболические зависимости, то для опосредованной зависимости полилексии от частоты мы получаем монотонно возрастающую степенную функцию. Это утверждение тоже легко проверяется эмпирически.

Как и в пункте 5.1, еще более плодотворным является рассмотрение параметров: степенная функция, которую мы получили, полностью определяется исходными уравнениями: ее параметры определяются, как было показано, исключительно из них. Этот формальный результат соответствует утверждению, что для описания связи между полилексией и частотой не требуются никаких других величин, кроме уже полученных в двух уравнениях.

Таким образом, эмпирическое изучение этой опосредованной связи на материале любого языка должно дать следующие результаты:

а) степенную функцию $PL = SF^{TZ}$

б) параметры, которые можно предсказать численно по формулам $S = QR^{-T}$ и $Z = NT$.

В принципе, со всеми опосредованными связями между величинами, используемыми в модели, тоже можно поступить аналогичным образом. При этом такие непрямые зависимости могут быть получены через несколько промежуточных переменных по очереди в соответствующие уравнения. Этим способом из системы модели можно получить все отношения между структурными свойствами лексических единиц.

5.3 РАСПРЕДЕЛЕНИЯ СТРУКТУРНЫХ ХАРАКТЕРИСТИК

Помимо результатов Ципфа и Мандельброта существует уже целый ряд исследований по распределениям свойств лингвистических единиц (ср. напр. Carrol 1967, 1968, 1970/72, Fucks 1955, Grotjahn 1982, Herdan 1966: 201-206, Крылов 1982). Во всех этих работах изучаемые свойства рассматриваются как изолированные явления, эмпирически наблюдаемые распределения которых необходимо возможно лучше аппроксимировать статистическими средствами. Исключение составляет работа Крылова, в которой он предлагает для распределения числа значений слов в словаре модель, полученную на основе теоретических соображений, т.е. дедуктивным путем³². Другие авторы (Арапов,

³² Однако эмпирическая проверка модели Крылова, проведенная им самим, не может рассматриваться как подтверждение, хотя сам он считает, что это аргумент в пользу модели. Позднее (1982) он предложил более совершенную модель.

Херц 1983; Вебб, Altmann 1985) убедительно выводят распределение лингвистических величин из их взаимозависимости с другими величинами.

В предлагаемой здесь модели значение структурного свойства в любом случае зависит от двух главных компонентов – это влияние потребности, равновесия между противодействующими потребностями или величины, опосредованным образом определяемой потребностями³³. Общим для вариантов этих компонентов является то, что в любой момент времени они являются константами, которые обеспечивают в данный момент переменным "частота", "длина", "полилексия" и "политекстия" фиксированную базисную долю³⁴.

В случае полилексии мы обозначили эту постоянную долю, обеспечиваемую равновесием между разнонаправленными потребностями $\min D$ и $\min K$, символом Q . Второй компонент, напротив, включает влияние независимых переменных. Это влияние не постоянно, а варьирует в любой момент времени в области значений, которые может принять структурная величина, рассматриваемая как независимая, и образует тем самым ту долю значения переменных, которая отвечает за индивидуальные отклонения свойств лексических единиц от фиксированной базисной доли³⁵.

Для полилексии независимой величиной является длина лексических единиц. Ее влияние, как указывалось выше, модифицируется потребностью спецификации через параметр N . Каждая лексическая единица имеет, таким образом, одно индивидуальное значение полилексии, отклонение которого от фиксированного значения, определяемого $\min D$ и $\min K$, зависит как раз от длины этой единицы.

Этот механизм позволяет получить еще одно следствие. Поскольку каждое значение независимых переменных дает определенное значение зависимых переменных, число лексических единиц, имеющих определенное значение независимых переменных, определяет число встречаемости значения функции (т.е. значение зависимых переменных). Так, модель позволяет сделать вывод, что вероятное распределение структурного свойства лексических единиц в словаре является функцией распределения определяющего его структурного свойства. Если, например, $n(x)$ – число лексических единиц с длиной x , то из-за того, что $y = Qx^{-T}$, число лексических единиц с полилексией – в простейшем случае равно теоретически $n(y) = n(x)$.

Эта связь эмпирически не очевидна, т.к. действующие при этом механизмы не носят детерминистского характера. Так, длина выражений не уменьшается при повышении частоты ни сразу, ни одновременно для всех выражений: приспособление системных величин происходит с определенным запаздыванием, причем степень инертности неоднобразна и зависит от многих, по большей части еще не известных факторов. Следует ожидать, что большинство значений переменных постоянно отстает по времени от своего оптимума. Таким образом, поскольку зависимости системных величин между собой имеют стохастический характер, зависимость вероятностных распределений друг от друга становится необходимым выводом с помощью стохастического процесса, который учитывает флуктуации. Решение этой задачи потребует, однако, отдельного исследования.

³³ Здесь – число фонем и величина словаря

³⁴ При линеаризованном представлении соединение аддитивно.

³⁵ Переменная доля соответствует мультипликативной связи.

5.4 ДИНАМИКА

Дальнейшие следствия, которые можно вывести из модели, касаются динамики системы. Наблюдение за значением одной или нескольких используемых в модели величин с течением времени соответствует диахроническому рассмотрению моделируемых лингвистических свойств. Теперь развитие одной величины можно сформулировать в зависимости от развития другой, подобно тому, как было бегло показано для распределения используемых в модели величин, так что все структурные свойства в их временном развитии можно предсказать, зная динамическое поведение одной из них.

В результате впервые представляется возможность для лингвистического имитационного моделирования. Для изменения частоты лингвистических единиц Altmann et al (Altmann 1983, Altmann, v. Buttler, Rott, Strauss 1983) вывели - на основе работы (А.А. Пиотровская, Р.Г. Пиотровский 1974) - дифференциальное уравнение из интеракционистской модели распространения новых элементов в языке; временные функции, получаемые в качестве решений, хорошо подтверждаются эмпирически (ср. Müller-Hasemann 1983, Imsiepen 1983, Best 1983, Kohlhase 1983, Best, Kohlhase 1983). Г. Альтман дал этому закону название "закон Пиотровского". Согласно этому закону доля каждого языкового явления в общем числе использований в зависимости от времени t при неполном изменении определяется как

$$p = \frac{c}{1 + ae^{-bt}},$$

при полном изменении (когда прокладывает себе дорогу новое явление, а все другие использования подавляются) как

$$p = \frac{1}{1 + ae^{-bt}},$$

а при обратимом развитии (когда явление после достижения максимума снова становится более редким) как

$$p = \frac{1}{1 + ae^{-bt + ct^2}}.$$

Этот закон создает предпосылки для проверки разработанной здесь модели на дихронических данных. Однако такая задача связана с усилиями, которые потребуют много времени: для большинства языков получение представительных данных представляется почти бесперспективным. Поэтому здесь будет достаточно указать на принципиальную возможность. Практические исследования следует предположительно ограничить от дельными аспектами.

Попытку более широкого имитационного моделирования, проверка которого обещает дать ценные данные, можно мыслить в данном случае тоже только в перспективе.

6. ЭМПИРИЧЕСКАЯ ПРОВЕРКА МОДЕЛИ

Теоретическая обоснованность и последовательность гипотез – вот первые и важнейшие требования к модели с претензией на объяснительную силу; кроме того, наша научно-теоретическая точка зрения требует эмпирической проверки высказываний³⁶. Для проверки развиваемой здесь базовой модели используются некоторые из вышерассмотренных следствий, вытекающих из модели.

6.1 ОПРЕДЕЛЕНИЕ РАБОЧИХ ПОНЯТИЙ

Для сопоставления полученных выше теоретических результатов с эмпирическими данными требуются еще сведения о том, какие лингвистические понятия наблюдения соотносятся с теоретическими конструктами модели; необходимо определить, как следует идентифицировать соответствующие единицы и каким способом измерять их свойства.

Правильность того или иного рабочего определения (операционализации) практически никогда нельзя доказать – во всяком случае логически обоснованным способом, хотя здесь мы не будем рассматривать общую проблематику построения правил соответствия для перевода теоретических конструктов в понятия наблюдения (общие сведения о возможных подходах можно найти, например, в работе Scheuch, Zehnpfennig 1974: 165-170); стоит сделать одно замечание о следствиях осуществленной операционализации.

Против дефиниций и способов измерения, которые предлагаются ниже для теоретических понятий модели, можно, без сомнения, выдвинуть много возражений, вытекающих из различных точек зрения на рассматриваемый предмет. Однако, даже хорошо обоснованное возражение против одной только операционализации не обесценивает результатов исследования. В зависимости от операционализации следует ожидать различных результатов анализа; конкретные результаты претендуют на обоснованность только в отношении данного множества правил соответствия и поддаются оценке только при данной предпосылке (ср. тж. Galtung 1967). На основе этой предпосылки в последующих разделах осуществляется поиск операционализаций и способов измерения для конструктов модели.

Основополагающим понятием в модели является понятие лексической единицы, хотя модель построена так, что в качестве лексических единиц можно рассматривать все виды лексикализованных выражений, здесь из практических соображений мы для начала ограничимся только словами. Для операционализации понятия "слово" существует несколько возможностей, из которых ни одна не может быть признана полностью удовлетворительной. Если отвлечься от трудностей, постоянно встречающихся при анализе прерывных величин, то существует принципиальная дилемма: В качестве структурных величин центральную роль играют длина и частота лексических единиц. Эти характеристики проявляются раньше всего на уровне словоформ, которые поддаются наблюдению в речи. Следует заметить, что значения характеристик слов подчиняются отраженным в модели процессам только при их

³⁶ Поэтому для построения лингвистической теории представляют интерес только такие модели, которые (по крайней мере в принципе) доступны для таких испытаний. Неэмпирические модели и другие не поддающиеся проверке гипотезы не могут быть полезными в этом отношении.

реальном использовании; при этом слово используется всегда в морфосинтаксическом окружении, от которого зависит актуализация его определенной словоформы³⁷.

В пользу выбора словоформы в качестве наблюдаемой величины говорит также то обстоятельство, что различные словоформы одной парадигмы различаются с точки зрения свойств "частота" и "длина". В отличие от этого, структурные величины полилексия и политекстия, которые тоже играют в модели центральную роль, ведут себя безразлично по отношению к флексивно-морфологическим различиям. Значения и контексты – это реальности, которые возникают при абстрагировании от актуализации слов в определенной грамматической структуре. Это аргумент в пользу применения лемм вместо словоформ.

Идея использования различных дефиниций в зависимости от изучаемой связи должна быть отвергнута по двум причинам, хотя для проверки зависимости длины от частоты применение словоформ, а для других функций применение лемм само по себе без сомнения не представляет трудностей; получаемые данные измерений из-за несоответствия дефиниций будут несопоставимы³⁸. Второе возражение полностью исключает такой подход: модель трактует лексические единицы как точки в лексическом пространстве, измерения которого составляют структурные характеристики. Тем самым устанавливается, что эмпирические измерения этих характеристик должны производиться как векторные измерения, т.е. для каждой исследуемой лексической единицы следует одновременно задать все структурные характеристики³⁹, что является предпосылкой для единообразной операционализации.

В противном случае проверке поддавались бы только отдельные выводимые из модели гипотезы законов, но не модель в целом и не непосредственно. Модель не отдает теоретического предпочтения ни одному из двух решений, поэтому здесь придется параллельно использовать обе операционализации⁴⁰. Теперь необходимо операционализировать четыре структурные характеристики лексических единиц. Наименьшую трудность представляет длина. Поскольку мы ограничиваемся словами, в качестве лингвистически осмысленных единиц длины рассматриваются только морфемы, слоги, фонемы и графемы. Поскольку в модели морфемы и слоги не представлены эксплицитно⁴¹, а уровень фонем/графем, напротив, представлен, последний лучше подходит для измерения длины. Выбор между этими двумя вариантами не играет никакой роли; конечно, в большинстве языков не существует однозначного соответствия между фонемами и графемами, но обе эти системы знаков являются отражением языковой реальности, которые равным образом возникли в результате абстрагирования и, по-видимому, одинаково отдалены от языковой реальности.

Поскольку материал для исследования, например, массивы текстов и словари, существует преимущественно в орфографической форме, в качестве меры длины лексической единицы следует принять число ее графем. В случае словоформ

³⁷ Для любого синтаксического, лексического и морфологического толкования конструкции.

³⁸ Тогда измеренные частоты и длины будут относиться к различным объектам.

³⁹ Тем самым исключаются более простые способы, например, установление связи между частотой слов определенной длины из одной выборки и полилексией слов той же длины из другой выборки.

⁴⁰ Разумеется, независимо друг от друга и полностью изолированно.

⁴¹ При моделировании отношения полилексии и длины и влияния потребности спецификации имплицитно делается ссылка на морфемы

учитывается последовательность графем между двумя пробелами⁴², в случае лемм – последовательность графем соответствующей основной форм⁴³.

Для операционализации длины лемм, естественно, возможны и другие решения. Так, можно возразить, что длина произвольно выбранной "основной формы" не является той, которая наблюдается при использовании лексической единицы, а таковой является длина фактически используемой словоформы. Если исходить из этого, то в качестве меры длины следовало бы использовать "действительную длину", среднюю среди длин фактически используемых словоформ отдельных лемм (взвешенную по их отдельным частотам). Эта возможность не была использована в первую очередь из-за больших затрат на измерения, однако она привлекалась для уточнения измерений в тех случаях, когда вышеописанный, более грубый способ измерения представлялся нам недостаточным. Еще одно возражение может заключаться в том, что различная длина флексивных морфем лемм различных частей речи приводит к систематической ошибке (ср. не имеющую окончания предикативную форму основной формы прилагательного с инфинитивным окончанием глагола). Здесь тоже осуществлялась соответствующая корректировка способа измерения (измерение длины без флексивных морфем) в той степени, насколько это было необходимо с точки зрения состояния данных. Значительно большую трудность представляет измерение полилексии.

К сожалению, в настоящее время нет алгоритма для определения семантического потенциала выражений, и поэтому единственным практическим способом представляется тот, который используют Altmann et al⁴⁴.

Для оценки числа различных значений лексической единицы используется число дефиниций в словаре под этим словом. Хотя, как можно ожидать, между различными словарями существуют значительные расхождения, можно исходить из того, что предсказанные моделью закономерности проявятся при этом по крайней мере в общих чертах. Выражаясь другими словами, если даже при таком грубом способе измерений будут получены соответствующие прогнозам результаты, то это будет хорошим подтверждением модели. Если же испытание даст отрицательные результаты, то необходимо будет сделать дополнительные усилия, чтобы установить, требуется ли более точная мера или следует считать модель в этом отношении ошибочной.

Политекстия вводится как мера числа контекстов, в которых возможно использование определенного выражения. Совершенно ясно, что для такой величины возможны только очень грубые измерения. Для эмпирической проверки модели мы установили, что политекстия лексической единицы определяется числом текстов, в которых эта единица встречается хотя бы один раз. Поскольку число всех текстов бесконечно, здесь возможна лишь оценка выборок из массива текстов⁴⁵.

⁴² При игнорировании прерывных величин

⁴³ Как это принято в традиционной грамматике: инфинитив; именительный падеж, единственное число, мужской род и т.д. Разумеется, любая флексивная форма парадигмы была бы в равной степени пригодной.

⁴⁴ См. напр., Altmann, Веbthy, Best (1982).

⁴⁵ Никакой массив текстов не может претендовать на представительность для всех текстов языка в определенный интервал времени. Однако проверяемая зависимость политекстии от полилексии должна проявиться, когда в качестве текстовой основы используется очень односторонний массив, включающий большое число сходных текстов.

Ряд нетривиальных проблем возникает при операционализации понятия частоты. В работе (1983 г) Ю. К. Орлов убедительно показал, что вероятность появления слов применительно к одному языку в целом неустановима⁴⁶, поскольку любая выборка и, следовательно, любой массив текстов слишком малы для надежной вероятностной оценки, даже если они включают несколько миллионов словоупотреблений. Правда, предсказанная гиперболическая функция может принять экстремальную форму, но должна соответствовать данным. Это, в частности, является следствием в высшей степени неблагоприятного для статистических методов резко падающего распределения покрытия частот слов, которое известно из работы Г. К. Ципфа, но влияние которого на параметры выборки до сих пор не учитывалось. Большинство слов текста (приблизительно половина словоупотреблений) являются одноразовыми, в то время как для надежной оценки нужны частоты по крайней мере в 10 встречаемостей. Однако вместе с тем именно повторение слов в текстах является причиной того, что число содержащихся в выборке слов всегда очень мало по сравнению с лексическим инвентарем словаря, в результате чего недостаточная представленность большинства слов ведет к значительно завышенной оценке частот использованных слов.

Аргументация Ю. К. Орлова не оставляет никакого реального выхода; получение текстовой выборки, объем которой значительно превышает объем словаря, т.е. порядка 10 миллионов словоупотреблений, в настоящее время нереально. Однако, последствия этого для возможности проверки модели не являются столь катастрофическими, как это может показаться на первый взгляд. С одной стороны, безусловно, нет никакой возможности получить неискаженную оценку частотного распределения, но, с другой стороны, это распределение само по себе не является предметом исследования. Речь идет о связи между значениями этих переменных и значениями длины и политекстии. Разумеется, все соответствующие расчеты из-за орловского искажения тоже будут искажены, но все совершенно одинаковым образом. Эмпирически определяемые кривые будут слишком наклонными (пропорционально завышенной оценке частот) по отношению к недоступной языковой действительности и именно поэтому будут адекватны с точки зрения отражения зависимостей. Во всяком случае, нет никаких теоретических оснований для предположения об изменении типа кривой в результате этого искажения.

Поэтому следует просто принять частоту встречаемости лексической единицы в изучаемом массиве текстов как соответствие теоретической величине "частота". В зависимости от того, каким образом операционализуется понятие лексической единицы, нужно следовать определенному способу измерений, чтобы обеспечить соотносимость измеренных значений структурных характеристик. В качестве частоты словоформ непосредственно используется их абсолютная частота в массиве текстов; частота леммы является суммой абсолютных частот словоформ, принадлежащих лемме.

6.2 Получение и обработка данных

Для конкретной эмпирической проверки на основе вышеизложенных соображений по операционализации были получены данные, соответствующие теоретическим понятиям: величины длина, частота и политекстия слов немецкого языка – на основе анализа массива текстов, полилексия – из словаря. Из-за больших затрат, необходимых

⁴⁶ Это касается, несомненно, непременно своевременного состояния эмпирических исследований в языкознании.

для такого исследования, проведение испытаний для начала было ограничено одним языком; другие языки будут привлечены позднее. По соображениям наибольшей доступности в качестве первого языка для испытаний был выбран немецкий. В качестве материала для испытаний использовался массив LIMAS⁴⁷, включающий 500 текстов и фрагментов, при выборе которых авторы делали упор на представленность различных видов текстов⁴⁸. Массив построен таким образом, что каждый отдельный текст имеет длину (хотя не всегда точно) в 2000 слов. Поэтому, отвлекаясь от других преимуществ и недостатков такого построения, можно заключить, что массив LIMAS очень хорошо подходит для получения требуемых данных и обладает также представительным объемом (1 миллион словоупотреблений). Из этого материала были непосредственно получены величины частота, политекстия и длина словоформ⁴⁹. С помощью программы лемматизации LEMMA 2, составленной Виллэ⁵⁰, было осуществлено полуавтоматическое сведение словоформ в леммы и части речи. Поскольку доля ручной работы велика (все предлагаемые программой лемматизации приходилось проверять и корректировать, то же касается сведения в части речи), потребовалось ограничение выборки приблизительно на одну пятую словаря: из 125000 словоформ, встречающихся в корпусе LIMAS, в дальнейшем рассматривались только словоформы с начальными буквами D, G, H, L и R⁵¹.

В результате был получен список из 13000 лемм, являющийся выборкой из той же генеральной совокупности, что и выше, при получении которой, однако, в качестве коррелята лексической единицы использовалась другая единица.

Измерение структурных характеристик лексемы осуществлялось путем повторного подсчета встречаемости соответствующих словоформ в массиве, так что частота леммы получалась из суммы частот соответствующих словоформ. Политекстия леммы определялась как число текстов, в которых встречалась по крайней мере одна словоформа леммы. В качестве длины бралась фактическая длина леммы в буквах⁵². Все связи, полученные из модели, параллельно проверялись на обеих выборках. Кроме того, были сделаны контрвыборки с целью обнаружения систематических ошибок, которые могли быть вызваны явным или скрытым различием в трактовке частей речи. Для этого из обеих выборок были получены три дополнительные подвыборки, включающие открытые словарные классы существительных, глаголов и прилагательных. Таким способом каждое испытание было проведено 8 раз; в следующей таблице показано отношение выборок друг с другом.

Для проверки отношений, в которых участвует полилексия, использовались также соответствующие данные из словаря. Для этой цели был использован словарь "Deutsches Wörterbuch" Варига⁵³. Поскольку значения полилексии приходилось получать чисто ручным способом, было осуществлено дополнительное сокращение

⁴⁷ Ср. Schaefer (1976).

⁴⁸ Эти соображения не будем здесь рассматривать. В принципе при невыполнении этого требования возникают почти такие же проблемы как при анализе частоты (см. выше). Аналогичная оценка следствий тоже представляется обоснованной, таким образом, с этим материалом можно было работать.

⁴⁹ Получение исходных данных и их дальнейшая числовая обработка осуществлялись на ЭВМ вычислительного центра Эссенского университета.

⁵⁰ Ср. Willee (1979).

⁵¹ По техническим причинам пришлось отказаться от случайной выборки, хотя с методологической точки зрения она была бы предпочтительнее.

⁵² Однако в массиве LIMAS умляуты и буква "ß" представлены в виде двух графем, при отсчете, как это обычно принято, они рассматривались как один знак.

⁵³ Wahrig (1981)

объема выборки. Из выборки лемм была получена выборка меньшего объема – 1325 единицы. Для них по словарю Варига определялось значение полилексии как число указанных в нем различных значений.

Таблица 4: Выборки из списка слов по массиву LIMAS

Словоформы			Леммы		
В совокупности			В совокупности		
Существ.	Глаг.	Прилаг.	Существ.	Глаг.	Прилаг.

Таким образом, при решении вопросов, связанных с полилексией, помимо 8 вышеупомянутых использовалась также эта 9-я выборка.

6.3 МЕТОДИКА ЭКСПЕРИМЕНТА

Важнейшим элементом эмпирической проверки модели является соотношение теоретически утверждаемых связей между парами системных величин с зависимостями переменных, вытекающих из анализа выборок. Каждая структурная связь двух используемых в модели величин соответствует гипотезе по следующей схеме:

H_y : значение переменной y зависит от значения переменной x согласно функциональному уравнению

$$y = Ax^b$$

Из этой схемы путем подстановки вместо символов x и y переменных, соответствующих системным величинам, порождаются все необходимые отдельные гипотезы (все структурные характеристики модели связаны друг с другом подобным образом). Конкретно это гипотезы⁵⁴:

$$H_L : L = A F^B \quad (\text{длина в зависимости от частоты})$$

$$H_{PL} : PL = A L^B \quad (\text{полилексия в зависимости от длины})$$

$$H_{PT} : PT = A PL^B \quad (\text{политекстия в зависимости от полилексии})$$

$$H_F : F = A PT^B \quad (\text{частота в зависимости от политекстии})$$

Константы A в функциональных уравнениях объединяют в себе те факторы, которые не являются функцией независимой величины. Например, в H_L эта константа обозначает произведение

$$LG^A \text{Red}^Z \text{PH}^{-P},$$

⁵⁴ Символы констант A и B обозначают в разных гипотезах разные константы! Такие обозначения были выбраны для упрощения записи.

т.е. общее влияние величины словаря, потребности в надежности передачи и величины инвентаря фонемной системы.

Построенные таким способом гипотезы проверяются теперь на соответствие данных выборок. Для этого из всех выборок были построены множества пар значений измеренных структурных характеристик⁵⁵. Затем функциональные уравнения линеаризуются и по алгоритму Хаусхольдера⁵⁶ (см., напр. Stoer 1979: 164-171) определяются те значения параметров А и В, которые обеспечивают наименьшие отклонения данных от теоретических значений (аппроксимация кривой). В заключение с помощью F-теста определяется, являются ли отклонения эмпирически измеренных значений переменных y_i от теоретически ожидаемых достаточно малыми, чтобы их можно было рассматривать как случайные. Этот тест устанавливает соотношение суммы объясненных отклонений экспериментальных точек от регрессионной прямой линии с необъясненными отклонениями:

$$F_{f1, f2} = \frac{SSR / (k - 1)}{SSE / (n - k)},$$

$$SSR = \sum (\bar{y} - y_i)^2 \quad (\text{объясненная дисперсия}),$$

$$SSE = \sum (y_i - \hat{y}_i)^2 \quad (\text{необъясненная дисперсия}),$$

где k – число параметров, а n – число точек данных. Число степеней свободы определяется как $f1 = k - 1$; $f2 = n - k$ (ср. Walker, Lev 1953: 230 и далее). Получаемое значение F является мерой качества аппроксимации. Если еще вычислить вероятность случайного достижения столь же большого или большего значения A по сравнению с найденным, то возникает возможность сравнивать между собой различные аппроксимации.

Все отдельные гипотезы подверглись этому тесту, в результате чего было установлено, что гипотеза должна быть отвергнута в том случае, когда получаемое значение F может быть достигнуто с вероятностью, большей $P = 0,01$, даже случайно. Таким образом, гипотеза предварительно⁵⁷ принимается только тогда, когда установлено, что согласованность теоретических и эмпирических значений свидетельствует о правильности гипотезы по крайней мере с 99%-ной надежностью.

6.3.1 Полилексия

Из структуры модели следует гипотеза, что число значений выражения зависит от длины выражения. Константа A в функциональном уравнении $PL = A L^B$ отражает глобальное влияние равновесия между действием потребностей $\min D$ и $\min K$, а

⁵⁵ Поскольку в исследуемых явлениях речь идет о стохастических зависимостях, при измерениях для одного значения x часто бывает несколько различных y , из которых берется среднее значение. Поэтому при аппроксимации используется, как правило, значительно меньше точек, чем в выборке.

⁵⁶ В первом приближении может быть также применен метод наименьших квадратов погрешностей

⁵⁷ "Предварительно" потому, что индуктивные испытания не позволяют осуществить верификацию гипотез. Неудачная фальсификация позволяет рассматривать гипотезу как правильную до тех пор, пока она, может быть, все-таки будет опровергнута.

константа B – мера синтетичности языка и она объясняется потребностью спецификации.

Аппроксимация гиперболы данным выборки дает оценку, представленную в таблице 5⁵⁸. Из рисунка 19 можно составить визуальное впечатление о ходе кривой и о качестве аппроксимации. Здесь и во всех последующих графиках эмпирические точки данных обозначаются символом "x"; непрерывная линия соответствует теоретической функции с оцененными по выборке значениями параметров.

Таблица 5. Аппроксимация функции $PL = A L^B$

A	B	FG	F	$F_{0,01}$	P(F)
12,5	-0,828	1, 22	130	7,95	10^{-10}

A, B – эмпирические константы

FG – число степеней свободы ,

F – переменная F . Гипотеза принимается на уровне значимости 0,01, когда это значение превышает табличное F .

$F_{0,01}$ – теоретическое значение F на уровне 0,01

$P(F)$ – вероятность случайного получения равновеликого или большего значения F .

Полученные результаты полностью соответствуют ожиданиям: параметр B (синтетичность) имеет отрицательный знак⁵⁹, и теоретическая кривая превосходно отражает измеренные значения. На основании значения F , которое на несколько порядков превышает требуемое, гипотеза принимается. Для проверки связи этих двух величин может быть использована только одна выборка⁶⁰.

6.3.2 Политекстия

Выводимая из модели гипотеза предусматривает гиперболическую зависимость политекстии выражения от его полилексии. Параметры функции имеют следующее значение: A – равновесие воздействий экономии контекста и контекстной спецификации, B – влияние полилексии.

Результаты аппроксимации этой функции данным из той же выборки представлены в таблице 6 и на рисунке 20⁶¹.

Таблица 6: аппроксимация с помощью функции $PT = A PL^B$

A	B	FG	F	$F_{0,01}$	P(F)
3,92	1,56	1, 14	17,4	8,86	0,0009

⁵⁸ Эмпирические и теоретические значения выборки представлены в табличной форме в приложении А1.

⁵⁹ Ср. раздел 5

⁶⁰ Ср. раздел 6.1

⁶¹ Точки данных и соответствующие теоретические значения представлены в приложении А2.

Здесь гипотеза также подтверждается в полном объеме. Параметр V положителен, как того требуют допущения при конструировании модели: увеличение семантического объема вызывает увеличение множества контекстов (и, следовательно, текстов), в которых выражение может быть использовано. Значение F почти в два раза превышает требуемое для предварительного принятия гипотезы.

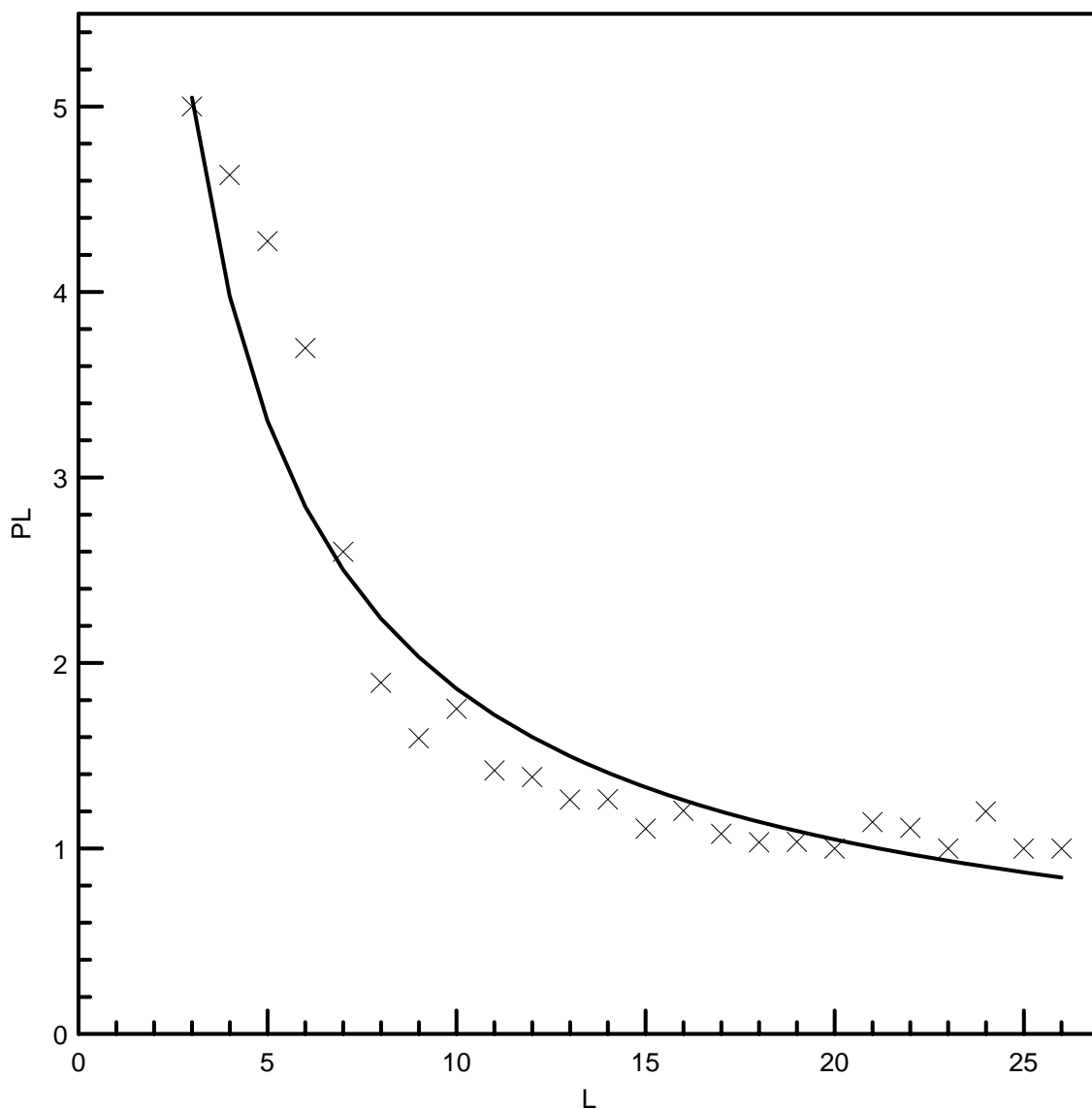


Рис. 19. Полилексия (PL) в зависимости от длины (L)

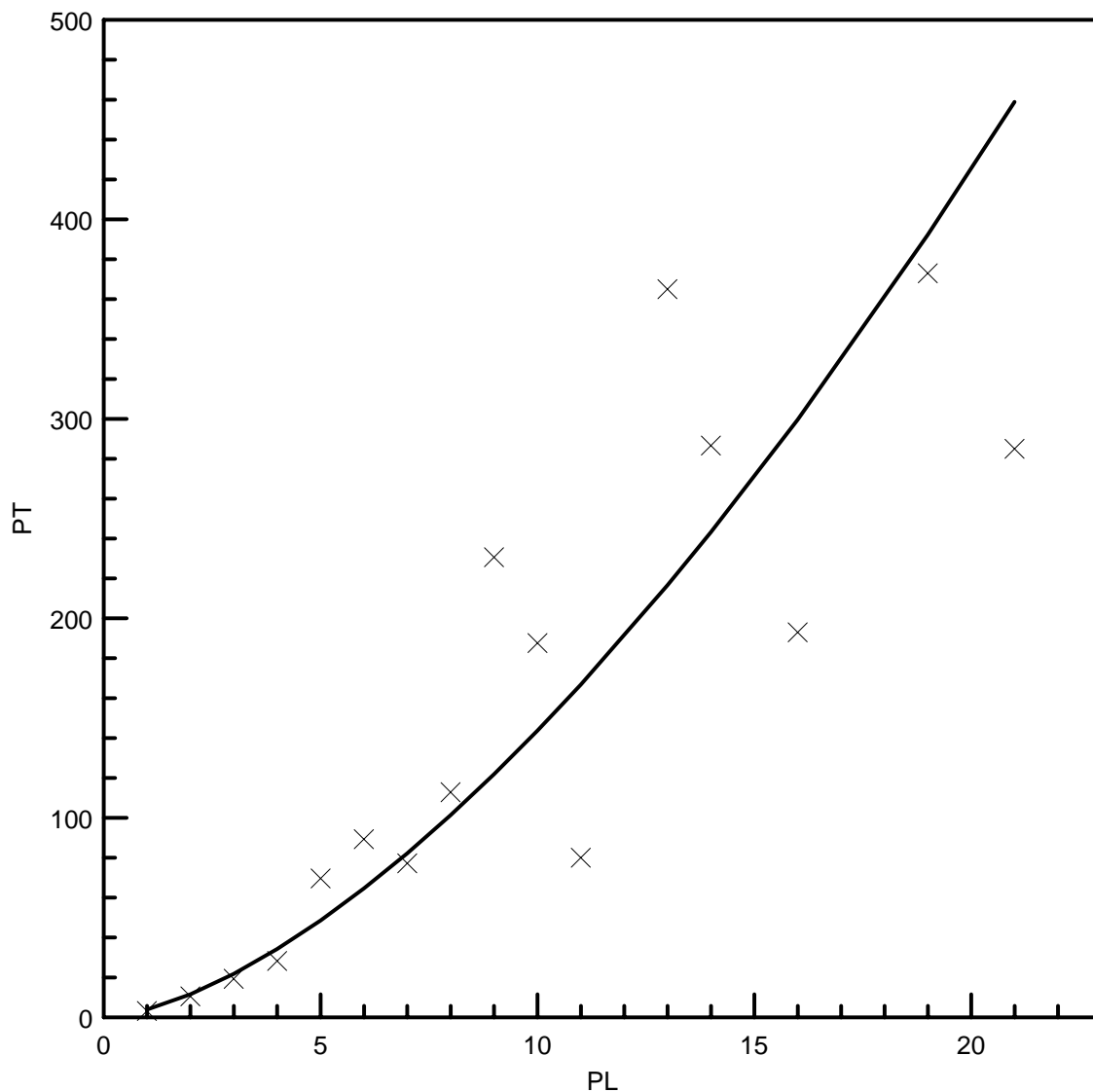


Рис. 20. Политекстия (PT) в зависимости от полилексии (PL)

6.3.3 Частота

Частота лексической единицы определяется согласно модели действием потребности Anw и политекстией единицы. В функциональном уравнении константа A обозначает влияние Anw , а константа B – действие политекстии.

Сначала следует представить результаты аппроксимации данным из той же выборки, которые (данные) использовались для первых двух связей. Таблица 7 и рисунок 21 свидетельствуют об очень хорошей аппроксимации⁶²: значение F почти в 50 раз превышает требуемое для принятия гипотезы.

⁶² Таблица данных см. приложение А3

Таблица 7: Аппроксимация функции $F=A PT^B$

A	B	FG	F	$F_{0,01}$	$P(F)$
1,12	1,10	1,122	332	6,86	$1,2 \cdot 10^{-11}$

Параметр B положителен и описывает ожидаемую тенденцию: чем в большем числе контекстов (ко-текстов) можно использовать выражение, тем чаще оно используется. В отличие от математически превосходного результата визуальное впечатление от аппроксимации несколько неудовлетворительное. График явно указывает на то, что более наклонная кривая легла бы на данную значительно лучше. Однако применение дифференциальных уравнений при конструировании модели не позволяет получить такую кривую.

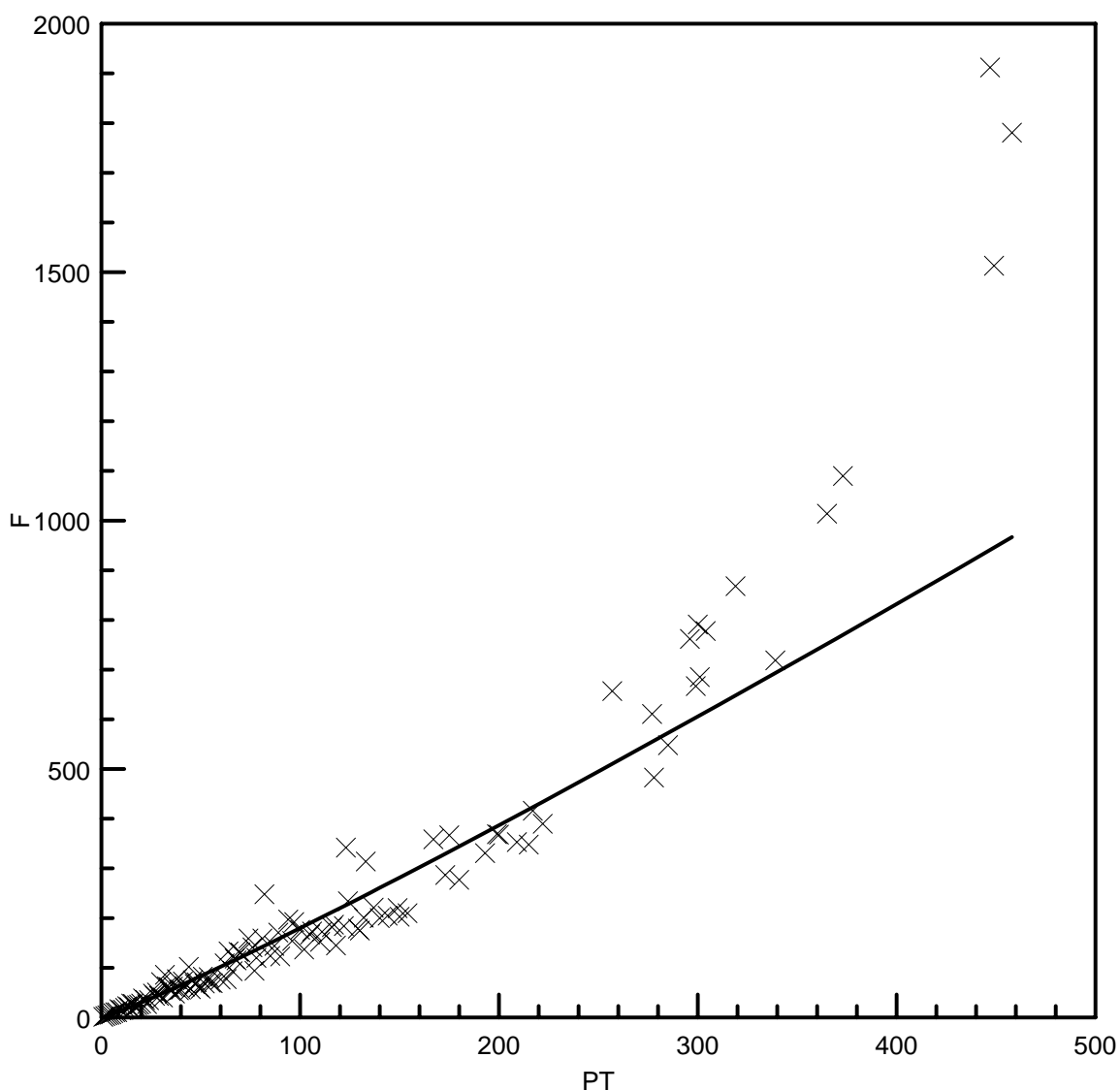


Рис. 21. Частота (F) в зависимости от политекстии (PT)

Такая ситуация может иметь две интерпретации. Либо этот подход правильный; тогда в данных массива LIMAS содержится искажение, которое следует рассматривать как нетипичное. Либо модель требует уточнения, поскольку не учтено какое-то влияние на системную величину "частота" (например, возраст лексических единиц). Какая из этих интерпретаций соответствует действительности, можно будет установить только после изучения достаточно большого числа массивов текстов на различных языках. Если первое предположение верно, то в отношении модели больше ничего не нужно делать, поскольку она соответствует результатам наблюдения. Если же верно второе предположение, то уже сейчас следует рассмотреть возможное уточнение модели.

До сих пор структура модели была основана в этом отношении на допущении, что скорость изменения частоты зависит только от политекстности. Выше это допущение было выражено с помощью дифференциального уравнения

$$\frac{y'}{y} = \frac{B}{x}$$

Наиболее простым дополнительным допущением в данном случае является существование постоянного члена, т.е. допущение о том, что изменение частоты лексической единицы помимо политекстности зависит еще от какой-то неизвестной постоянной величины. Тогда дифференциальное уравнение имеет вид

$$\frac{y'}{y} = \frac{B}{x} + c$$

и его решение дает функцию

$$y = Ax^B e^{cx}$$

Результаты аппроксимации тех же данных, что и выше, с помощью этой функции представлены в таблице 8 и на рисунке 22⁶³.

Таблица 8: Аппроксимация функции $F = AP^T e^{cPT}$

A	B	c	FG	F	F _{0,01}	P(F)
1,62	0,964	0,00193	2, 121	824	4,79	33 · 10 ⁻⁷¹

По сравнению с чисто гиперболической функцией полученный результат лучше с математической точки зрения и визуально ближе к данным. Нужно ли это уточнение модели, покажут в дальнейшем только эмпирические исследования. Однако, решающим в конечном счете будет возможность найти обоснованную интерпретацию такого дополнительного допущения. Эта интерпретация может заключаться в принятии предположения об обратной связи длины от частоты: по возможности предпочтение отдается более коротким выражениям (ср. раздел 4.6.2)

⁶³ Таблицу данных см. приложение А 4

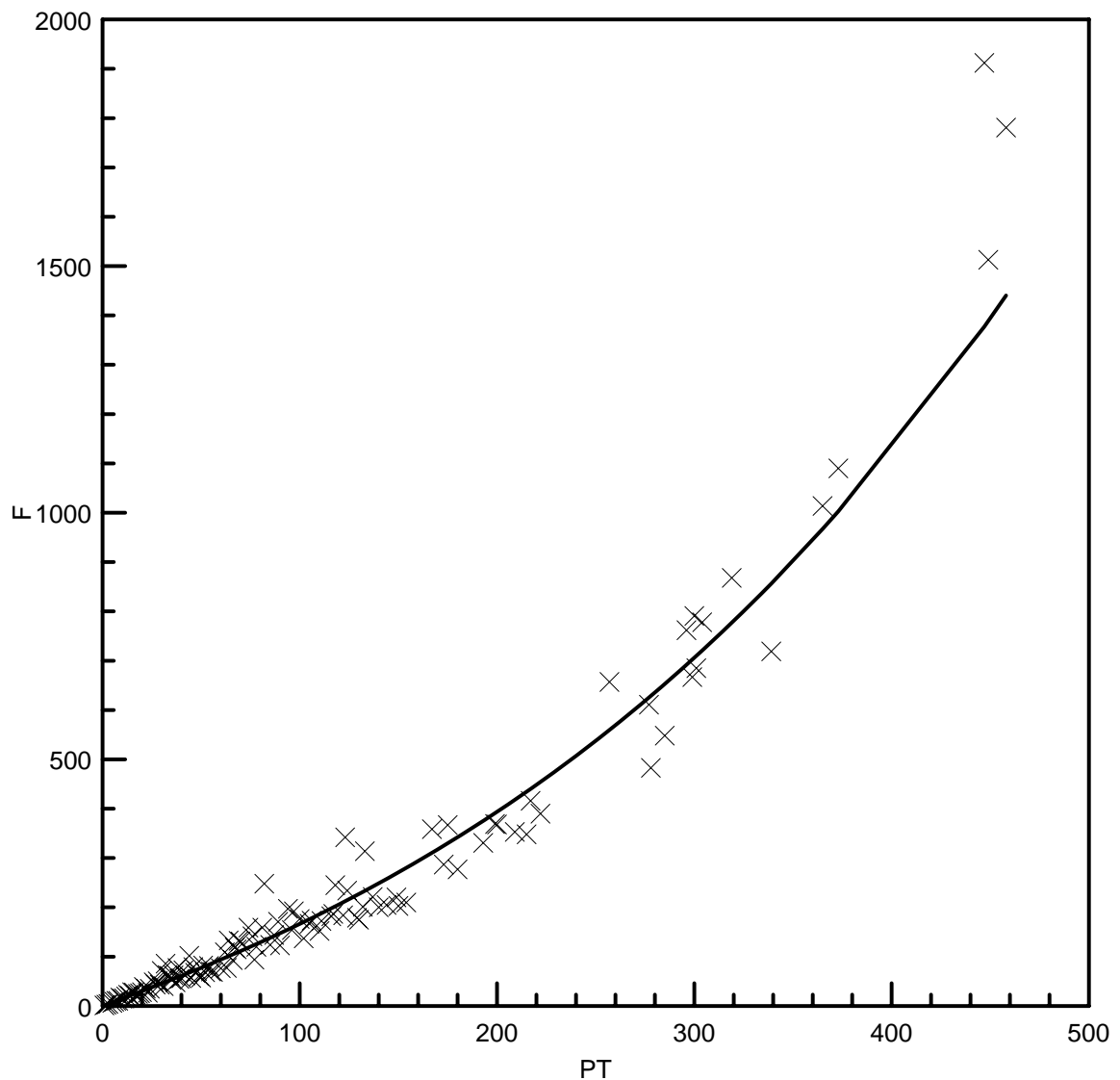


Рис. 22. Частота (F) в зависимости от политекстности (PT); уточненная модель

6.3.4 Длина

Длина лексической единицы определяется двумя факторами – глобальным компонентом, состоящим из элементов: величина словаря, число фонем и потребности в надежности передачи Red, и индивидуальным отклонением соответствующей единицы по причине ее частоты. В функциональном уравнении глобальный компонент представлен константой A . Параметру B , отражающему угол наклона графика уменьшения длины с течением времени, соответствует влияние частоты на длину.

Таблица 9 : Аппроксимация функции $L = A F^B$

A	B	FG	F	$F_{0,01}$	$P(F)$
10,9	-0,108	1, 158	105	6,80	$2,2 \cdot 10^{-11}$

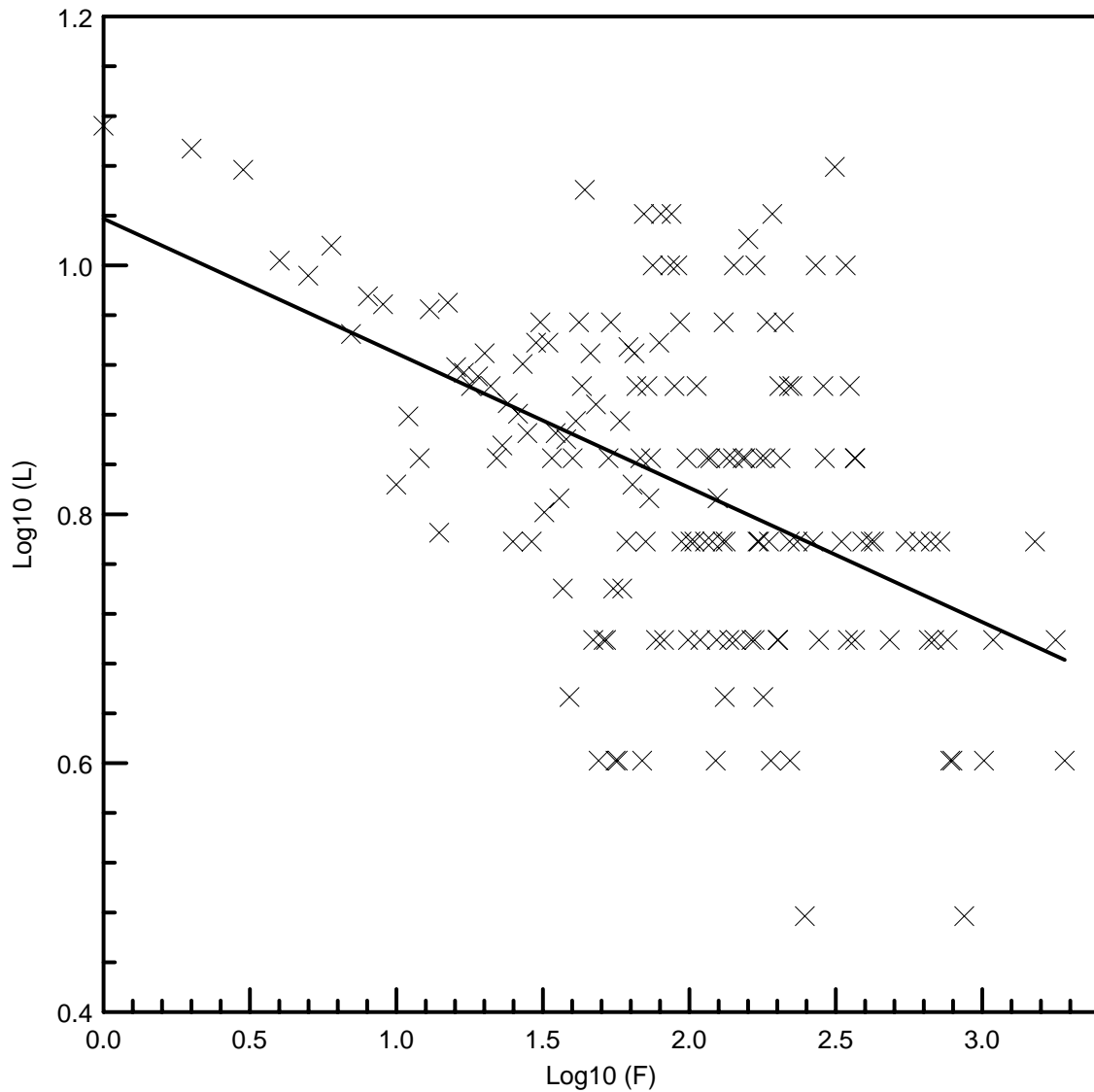


Рис. 23. Длина (L) в зависимости от частоты (F); по обеим осям логарифмический масштаб

Аппроксимация к данным (из той же выборки) дает снова превосходный результат: переменная A в 50 раз превышает требуемое табличное значение (см. таблицу 9). Визуальное впечатление от рисунка 23 также подтверждает, что теоретическая функция

хорошо отражает ход эмпирической кривой⁶⁴. Однако, следует признать, что по сравнению с другими аппроксимациями бросается в глаза очень большой разброс эмпирических точек вокруг графика функции. Это впечатление объясняется отчасти оптическим искажением в результате использования различных масштабов для F и L. С другой стороны, график действительно отражает большой разброс. К этому вопросу мы еще вернемся⁶⁵; пока достаточно сослаться на чрезвычайно хорошие результаты F-теста.

6.3.5 Проверка опосредованных зависимостей

Еще более жестким испытанием модели является проверка опосредованных зависимостей при одновременной увязке нескольких переменных, поскольку при проверке только прямой связи двух системных величин в форме гипотезы можно построить только одно изолированное высказывание о соответствующем фрагменте структуры.

Поэтому были построены дополнительные гипотезы о подструктурах с тремя связанными системными величинами и, наконец, гипотеза, заставляющая задействовать при испытаниях все ядро модели, т.е. все четыре структурные характеристики. С целью их проверки для соответствующих зависимостей из выборок были получены дополнительные эмпирические данные. Дополнительные гипотезы получены, как это описано в разделе 5, путем сопряжения двух или трех вышерассмотренных первичных гипотез за счет исключения промежуточной величины подстановкой соответствующей функции из функционального уровня. При этом константы A и B были вычислены из соответствующих констант исходных функций (эти исходные значения представлены в таблице 10).

Таблица 10. Эмпирические константы четырех прямых зависимостей

	L(F)	PL(L)	PT(PL)	F(PT)
A	10,9	12,5	3,92	1,12
B	-0,108	-0,828	1,56	1,1

Даже если придется исходить из того, что эмпирические точки, полученные от измерения таких опосредованных зависимостей, характеризуются значительно большим разбросом, чем у прямых зависимостей, в любом случае должна быть все-таки возможной достаточно хорошая аппроксимация функции $y = A x^B$. Однако, не следует ожидать, что оцененные по данным параметры будут очень близко лежать к расчетным значениям, за исключением, может быть, знака перед показателем и порядка величины. Отрицательный результат будет в этой ситуации свидетельствовать о том, что в соответствующем месте модельной структуры на одну из величин действует еще один, пока не учтенный параметр.

Ниже изложены в общем виде результаты этой попытки.

⁶⁴ Таблицу данных см. приложение А 5

⁶⁵ Ср. раздел 7

1. Для получения теоретического функционального уравнения зависимости полилексии от частоты первые две прямые зависимости связываются между собой. Из

$$L = 10,9F^{-0,108}$$

и

$$PL = 12,5L^{-0,828}$$

получаем путем подстановки

$$PL = 12,5(10,9F^{-0,108})^{-0,829}$$

и поэтому для этой связи мы имеем прогноз

$$A_V = 1,37 \text{ и } B_V = 0,09.$$

Аппроксимация функции по данным (ср. таблицу 11 и рисунок 24)⁶⁶ дала следующие оцененные параметры: $A_D = 1,03$ и $B_D = 0,283$ при снова превосходном качестве аппроксимации. Как видно из рисунка, в данных как будто бы наблюдается расщепление тенденции на две отдельные подфункции, что может объясняться различным поведением частей речи. Отдельное рассмотрение подфункций, согласно нашим предложениям, может дать еще лучший результат (ср. раздел 6.3.6).

Поразительно соответствие значений расчетных и эмпирических констант. Поэтому логично будет проверить, различаются ли параметры вообще значимым образом. Если это не так, то прогноз можно считать правильным. Интерес здесь представляет не столько параметр F , который характеризуется лишь параллельным сдвигом регрессионных прямых линий⁶⁷, сколько в первую очередь разность между B_V и B_D . Нулевая гипотеза

$$H_0 : B_V - B_D = 0$$

Таблица 11: Аппроксимация функции $PL = A F^B$

A	B	FG	F	$F_{0,01}$	$P(F)$
1,03	0,283	1, 158	121	6,80	$2,2 \cdot 10^{-11}$

означает принятие правильного прогноза. Ее следует отвергнуть, когда значение

$$t_{2n-4} = \frac{|B_V - B_D|}{s \sqrt{\frac{1}{\sum_i (x_{Vi} - \bar{x}_V)^2} + \frac{1}{\sum_i (x_{Di} - \bar{x}_D)^2}}},$$

⁶⁶ Таблицу данных см. в приложении А6

⁶⁷ Аппроксимация производилась после логаритмизации с помощью линейной регрессии.

$$\text{где } s^2 = \frac{(n-2)s_B^2 + (n-2)s_V^2}{2n-4}$$

больше, чем значение t-распределения с $2n-4$ степенями свободы. S_{BV}^2 и S_{BD}^2 – дисперсии X_{BV} и X_{BD} – средние значения X_{iBV} и X_{iBD} , теоретически вычисленных соответственно при V_V и V_D . В этом случае мы получаем значение $t = 0,079$ при 316 степенях свободы по сравнению с критическим значением 2,35 на уровне значимости 0,01. Таким образом, нулевую гипотезу отклонять не нужно, и теоретический прогноз считается подтвержденным.

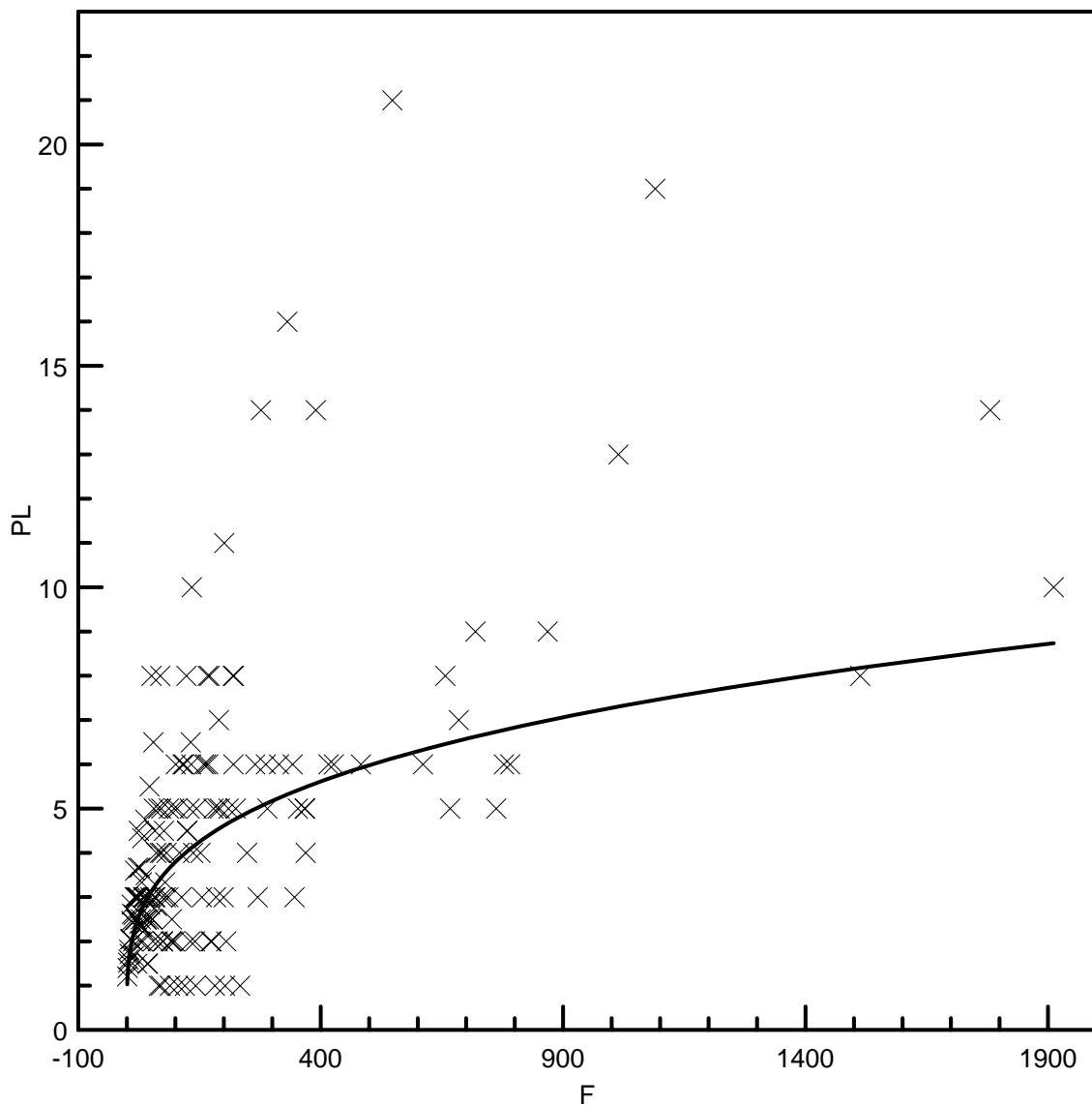


Рис. 24. Опосредованная зависимость полилексии (PL) от частоты (F)

2. Опосредованная зависимость полилексии от длины вычисляется аналогичным образом. Из

$$PL = 12,5 L^{-0,828}$$

и

$$PT = 3,92 PL^{1,56}$$

путем подстановки получаем функцию

$$PT = 3,92(12,5 L^{-0,828})^{1,56}$$

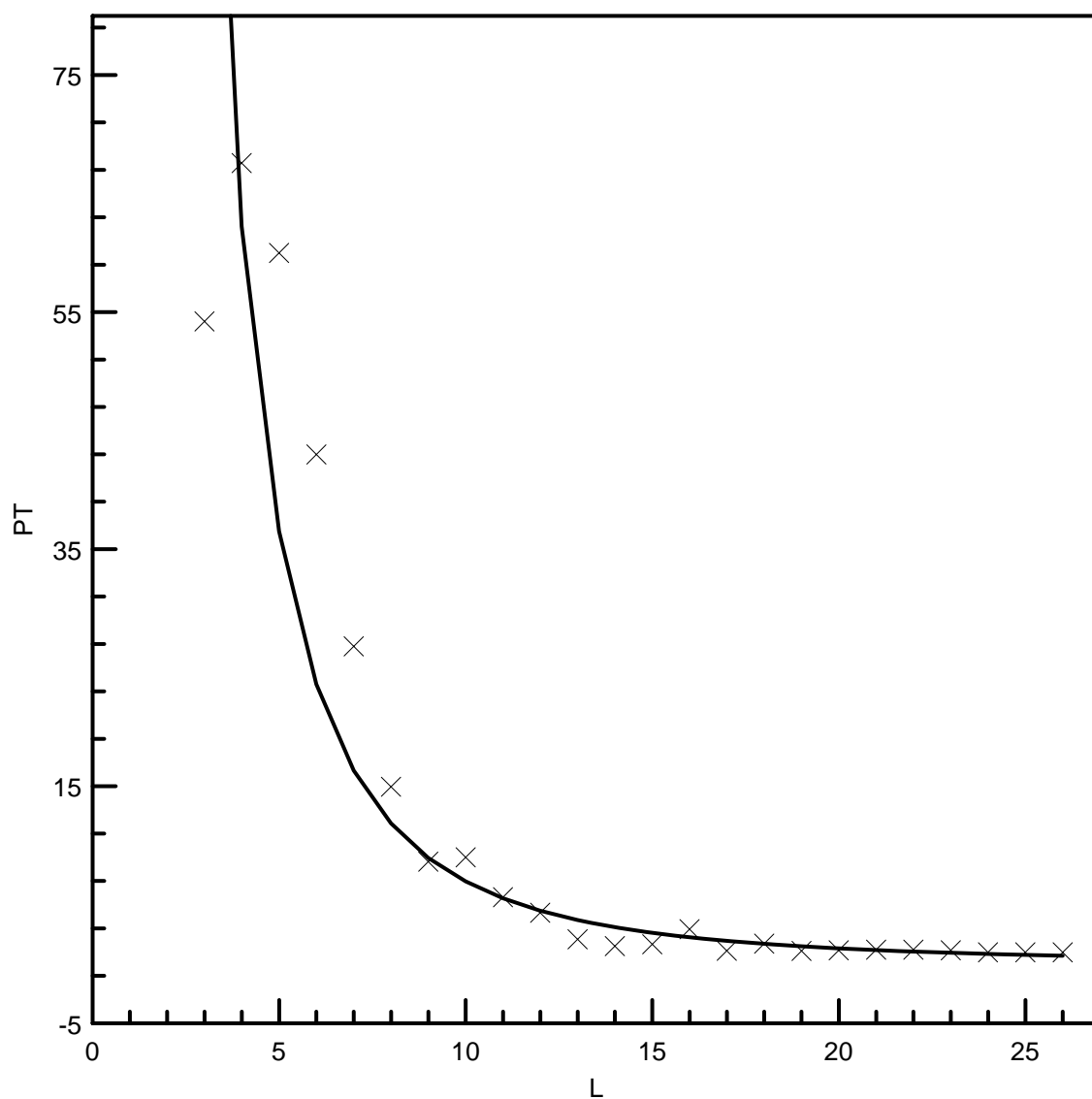


Рис. 25. Опосредованная зависимость политекстии (PT) от длины (L)

и, следовательно, теоретические значения $A_V = 201,6$ и $B_V = -1,29$. Аппроксимация гиперболы данным выборком⁶⁸ дала при хорошем значении F (ср. таблицу 12 и рисунок 25) эмпирическую оценку $A_D = 1710$ и $B_D = -2,39$.

Таблица 12. Аппроксимация функции $PT = A L^B$

A	B	FG	F	$F_{0,01}$	$P(F)$
1710	-2,39	1, 22	17,8	7,95	0,00035

Если параметр B определяется очень точно, то значение A_D достаточно велико, чтобы его можно было идентифицировать с расчетным A_V . Однако, это отклонение не должно нас особенно тревожить, если мы учтем значение этого параметра: он отражает политекстную слов длины 1, т.е. определяет ход кривой в первую очередь в зоне малых длин. Однако, не существует ни одной леммы с длиной 1, так что здесь придется считаться с особенно большими погрешностями измерений. Проверка $B_V - B_D$ проводилась, как и раньше, с помощью t -теста и дала значение $t=0,6502$ при 44 степенях свободы по сравнению с табличным значением 2,41 на уровне 0,01. В данном случае прогноз значения параметра тоже следует признать правильным.

3. Третья опосредованная связь существует теоретически между частотой и полилексией. Как и выше путем подстановки из

$$PT = 3,92 PL^{1,56}$$

и

$$F = 1,12 PT^{1,1}$$

получаем уравнение

$$F = 1,12(3,92 PL^{1,56})^{1,1}$$

что дает теоретические параметры $A_V = 5,033$ и $B_V = 1,716$. В этом случае аппроксимация (ср. таблицу 13 и рисунок 26)⁶⁹ дала оцененные значения параметров, которые почти численно совпадают с предсказанными. Проверка разности параметров также дала соответственно минимальное значение $t=0,0311$ с 28 степенями свободы при теоретическом $t_{0,01} = 2,47$.

⁶⁸ Таблица данных см. приложение А7

⁶⁹ Таблица данных см. приложение А8

Таблица 13. Аппроксимация функции $F = A PL^B$

A	B	FG	F	$F_{0,01}$	P(F)
5,50	1,76	1, 14	11,2	8,86	0,0048

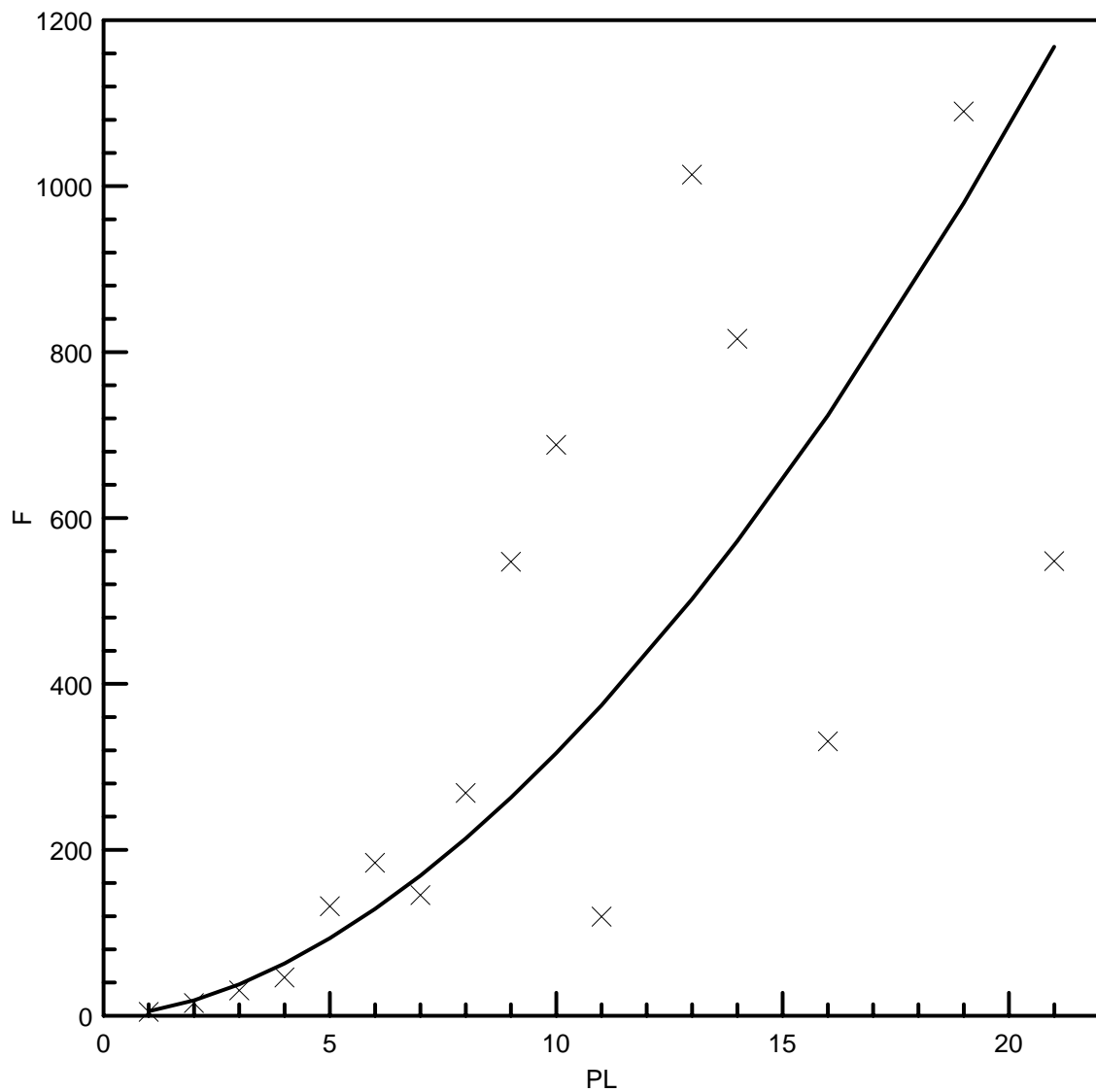


Рис. 26. Опосредованная зависимость частоты (F) от полилексии (PL)

4. Наконец, остается связь между длиной и политекстией. Исходные уравнения

$$F = 1,12 PL^{1,1} \text{ и}$$

$$L = 10,9 F^{-0,108}$$

дают

$$L = 10,9(1,12 PT^{1,1})^{-0,108}$$

откуда определяются $A_v = 10,77$ и $B_v = -0,12$. Эта аппроксимация (ср. таблицу 14 и рисунок 27)⁷⁰ снова обеспечивает числовое соответствие с оцененными значениями. Значение t еще меньше чем прежде, а именно 0,0056 (996 степеней свободы), в то время как критическое значение на уровне 0,01 составляет 2,33.

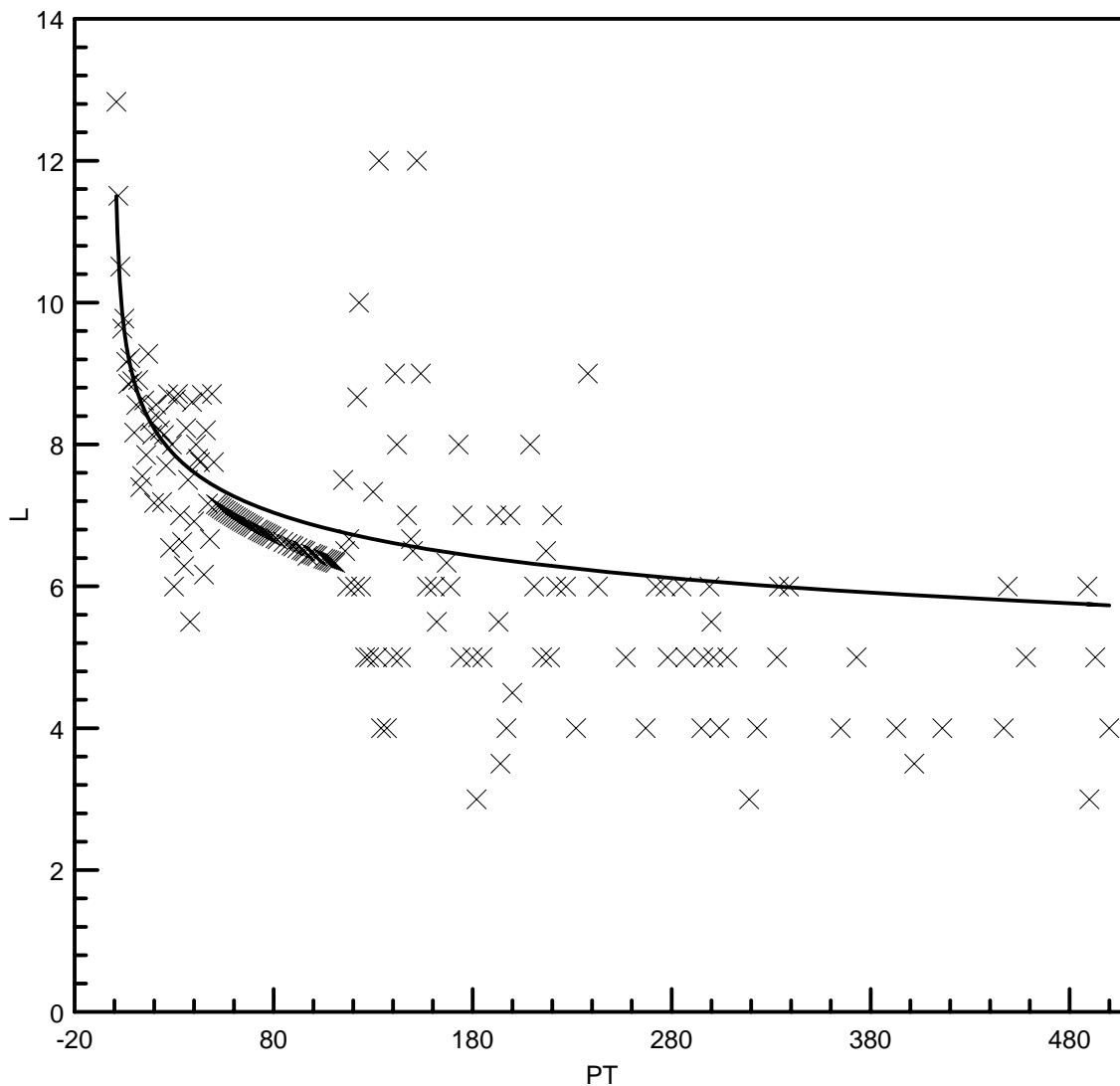


Рис. 27. Опосредованная зависимость длины (L) от политекстии (PT)

⁷⁰ Таблицу данных см. приложение А9

Таблица 14. Аппроксимация функции $L = A PT^B$

A	B	FG	F	$F_{0,01}$	P(F)
11,5	-0,112	1,180	110	6,77	$2 \cdot 10^{-11}$

5. На основе этих положительных результатов в заключении была сделана попытка также проверить структуру всего ядра модели с помощью опосредованной зависимости, включающей все четыре используемые в модели переменные. Как и раньше, путем подстановки

$$F = 1,12 PT^{1,1}$$

$$L = 10,9 (1,12 PT^{1,1})^{-0,108}$$

$$PL = 12,5 (10,767 L^{-0,1188})^{-0,828}$$

были вычислены теоретически прогнозируемые параметры функций, что дало $A_v = 1,747$ и $B_v = 0,10$. Несмотря на большое „расстояние“ между независимыми и зависимыми переменными соответствие этих данных значениям, получаемым на основе аппроксимации (ср. таблицу 1 и рисунок 28)⁷¹, только на соответствие упорядочения, а в данном случае получилось почти численное равенство. На основе критерия разности параметров ($t = 0,0868$ с 244 степенями свободы по сравнению со значением t – распределения 2,35 на уровне значимости 0,01) фактически тоже приходится исходить из равенства теоретического и эмпирического значения.

Таблица 15. Аппроксимация функции $PL = A PT^B$

A	B	FG	F	$F_{0,01}$	P(F)
1,11	0,31	1,122	88,8	6,86	$2,1 \cdot 10^{-11}$

Эта дополнительная проверка опосредованных зависимостей в целом тоже свидетельствует о правильности конструкции модели; предварительно можно даже сделать вывод, что смоделированная подсистема обладает относительной автономностью.

6.3.6 Сравнение возможных рабочих понятий

Во всех приведенных описаниях использовались данные выборки, полученных путем измерений, в которых в качестве наблюдаемого коррелята теоретического понятия ”лексическая единица” служила лемма найденной в тексте словоформы.

⁷¹ Таблицу данных см. приложение F10

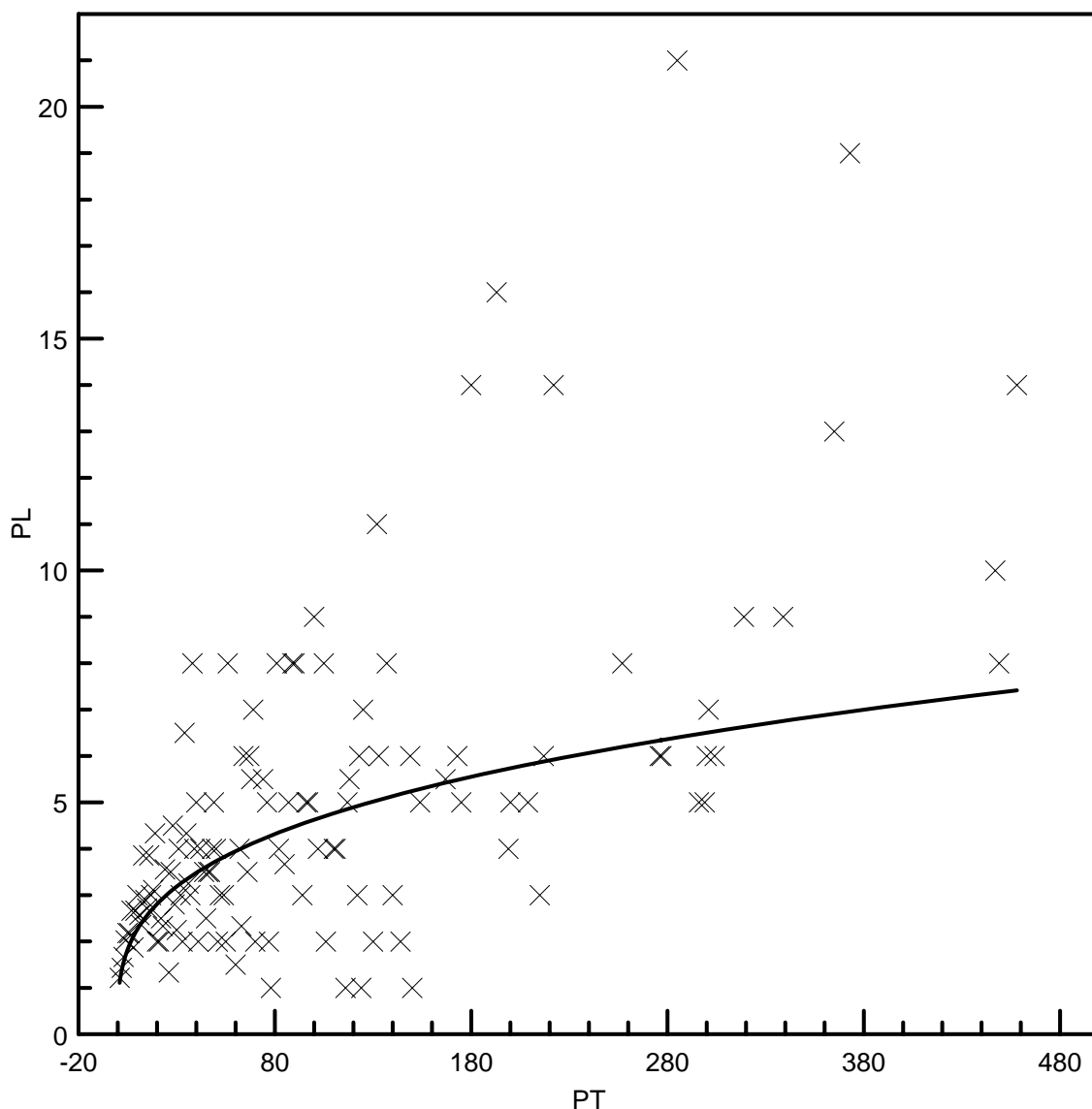


Рис. 28. Дважды опосредованная зависимость полилексии (L) от политекстии (PT)

Для контроля были проведены дополнительные независимые тестирования на базе словоформ. Это было возможно в отношении только тех связей переменных, в которых не играет никакой роли полилексия, т.е. для связи длины с частотой и частоты с политекстией. Поскольку искажения, которые могли возникнуть вследствие неправильной операционализации, имели по нашим предположениям частичечную специфику⁷², были учтены также различия в принадлежности слов к одному из этих открытых классов слов. В таблице 16 и на рисунке 29 представлены результаты аппроксимаций восьми выборок функции $y = A^x e^{cx}$ для зависимости частоты от политекстии⁷³. Сомнений в правильности гипотезы не возникало, почти все результаты

⁷² Количественное различие при анализе частей речи и лемм зависит в первую очередь от богатства форм соответствующей лексической единицы.

⁷³ Таблицы данных см. в приложениях с А 11 по А 18.

испытаний оказались исключительно высокими. Кроме того, можно констатировать, что даже оценки параметров лежали очень близко друг от друга. Несомненно, лексические единицы (как леммы, так и словоформы) различаются в зависимости от принадлежности к классам слов по своему зависимому поведению в отношении системных величин. Однако до включения этого различия в модель разности параметров также следует проверить на значимость.

Таблица 16. Аппроксимация функции $F = A P T^B e^{cPT}$ восьмью различными выборкам

	Глаголы	Существительные	Прилагательные	Всего
Леммы	A = 1,25 B = 0,96 C = 0,00234 FG : 2, 120 F = 508 F _{0,01} = 4,8 P(F) = 2,68·10 ⁻⁵⁹	A = 1,53 B = 1,03 C = 0,00705 FG : 2, 98 F = 237 F _{0,01} = 4,82 P(F) = 2,87·10 ⁻³⁸	A = 1,47 B = 0,96 C = 0,00234 FG : 2, 120 F = 508 F _{0,01} = 4,79 P(F) = 3,48·10 ⁻⁸¹	A = 2,16 B = 0,848 C = 0,00391 FG : 2, 180 F = 66,1 F _{0,01} = 4,72 P(F) = 2,99·10 ⁻²²
Словоформы	A = 1,46 B = 0,982 C = 0,00394 FG : 2, 94 F = 48,3 F _{0,01} = 4,85 P(F) = 3,38·10 ⁻¹⁵	A = 1,35 B = 1,12 C = 0,00215 FG : 2, 82 F = 215 F _{0,01} = 4,88 P(F) = 2,43·10 ⁻³³	A = 1,52 B = 1,02 C = 0,000838 FG : 2, 109 F = 242 F _{0,01} = 4,80 P(F) = 8,09·10 ⁻⁴¹	A = 1,77 B = 1,02 C = 0,0015 FG : 2, 157 F = 56,1 F _{0,01} = 4,74 P(F) = 4,14·10 ⁻¹¹

6.3.7 Обобщение результатов проверки

Все испытания, включая проверку дополнительных гипотез и применение контрольных выборок, однозначно свидетельствуют о соответствии модели рассматриваемым эмпирическим данным. Решения о приемлемости гипотез принимались на относительно высоком уровне значимости $P = 0,01$. Найденные значения испытанных переменных превзошли табличные значения почти во всех случаях на несколько порядков величины; вероятность ошибки поэтому очень мала. Во всех случаях теоретически прогнозируемые значения параметров соответствовали вычисленным значениям на основе данных даже численно.

Как большую удачу следует отметить тот факт, что выведенные из модели связи переменных столь же хорошо подтверждались даже в случае дважды опосредованной зависимости с участием четырех системных величин. В случае зависимости частоты от политекстии есть указание на то, что здесь может наблюдаться дополнительное, до сих пор еще неизвестное влияние. Однако этот вопрос может быть решен только на основе значительно более крупной базы данных, если не возникнет теоретической необходимости допустить существование такой новой величины. Проверка модели на полученных данных позволяет сделать однозначный вывод о ее правильности. Для большей очевидности и решения вопроса о сфере действия модели требуется еще большое число различных исследований, особенно на материале других языков.

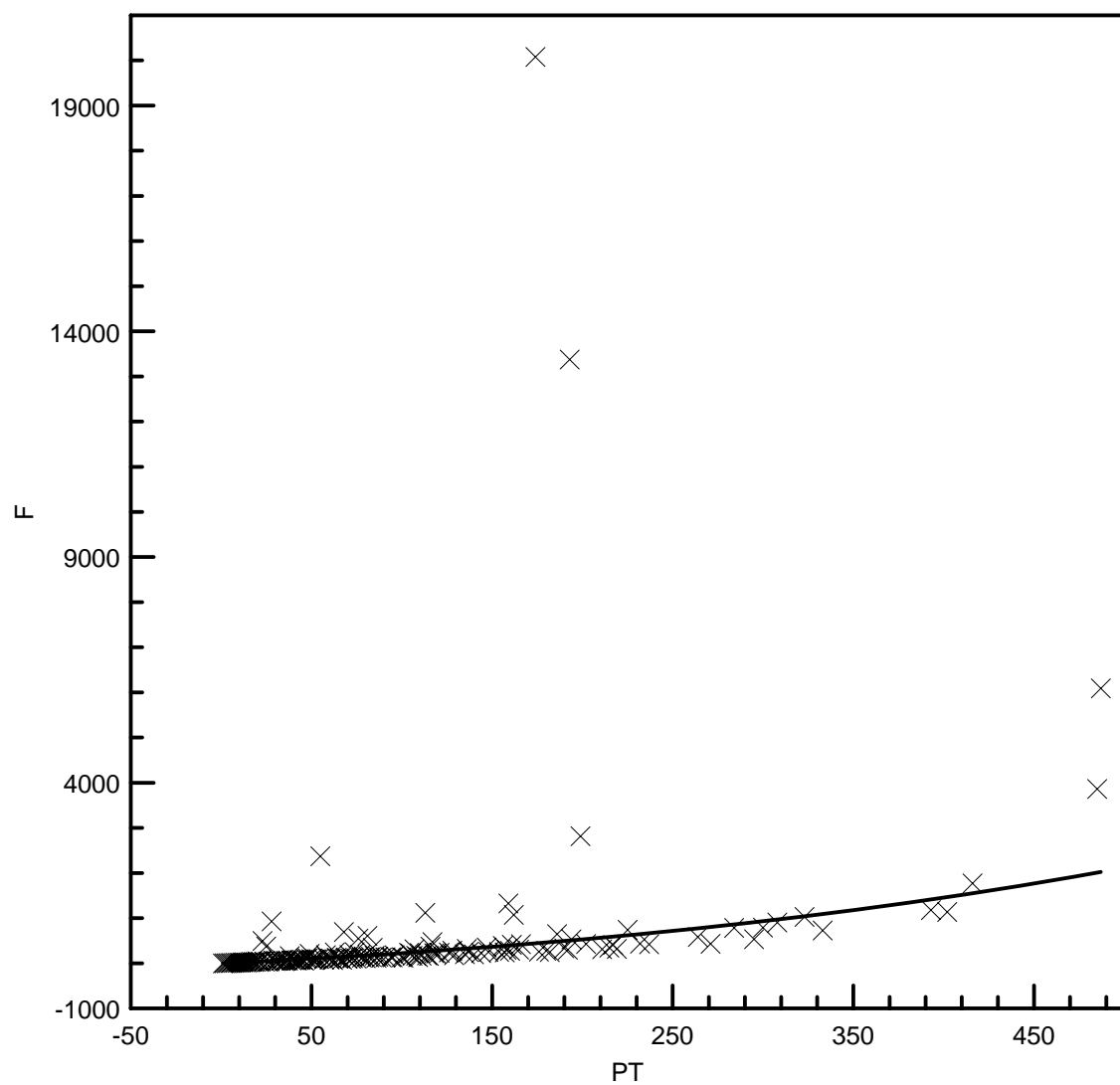


Рис. 29 а. Зависимость частоты (F) от политекстности (PT): словоформы в совокупности

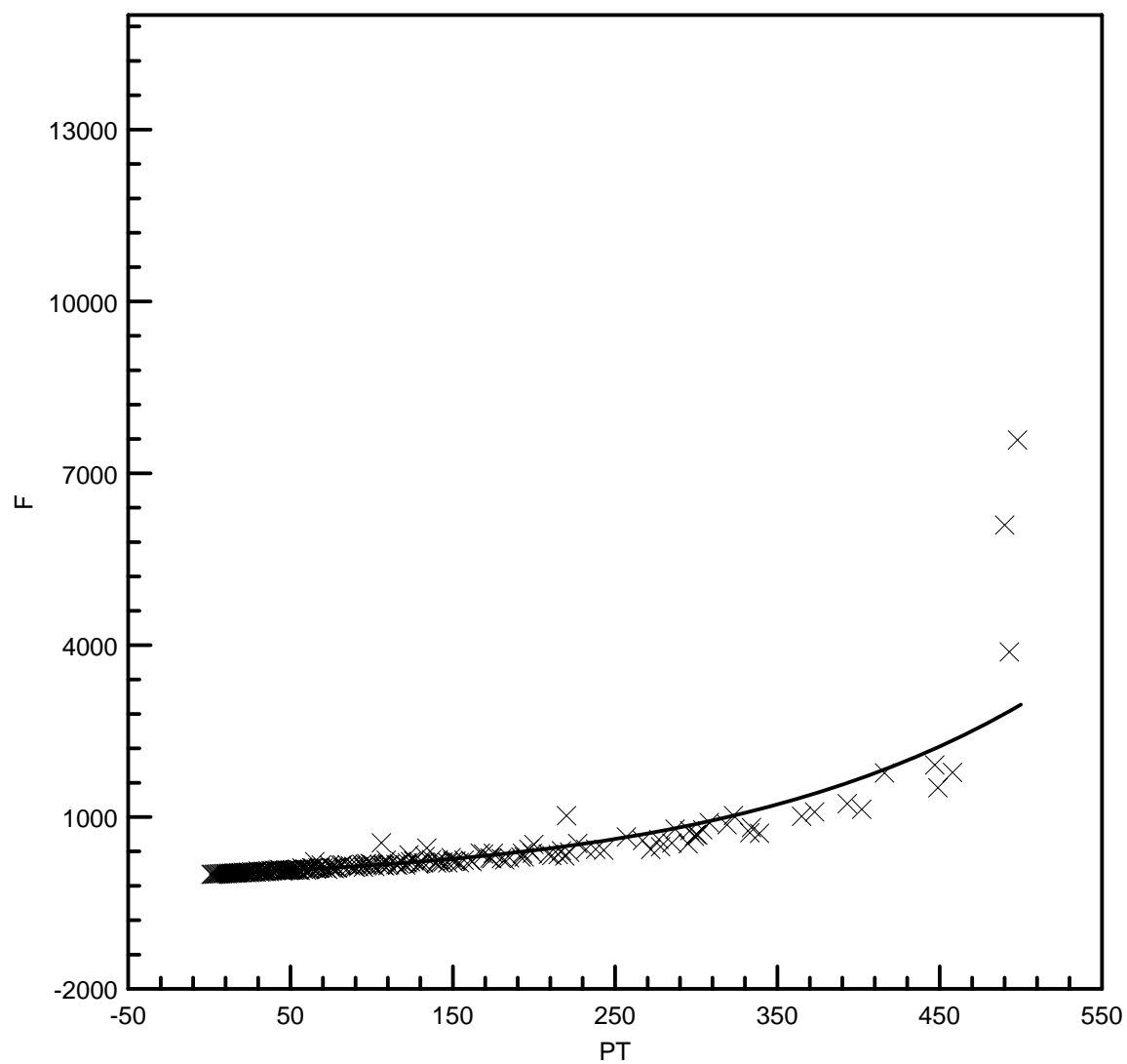


Рис. 29 б. Зависимость частоты (F) от политекстности (PT): леммы в совокупности

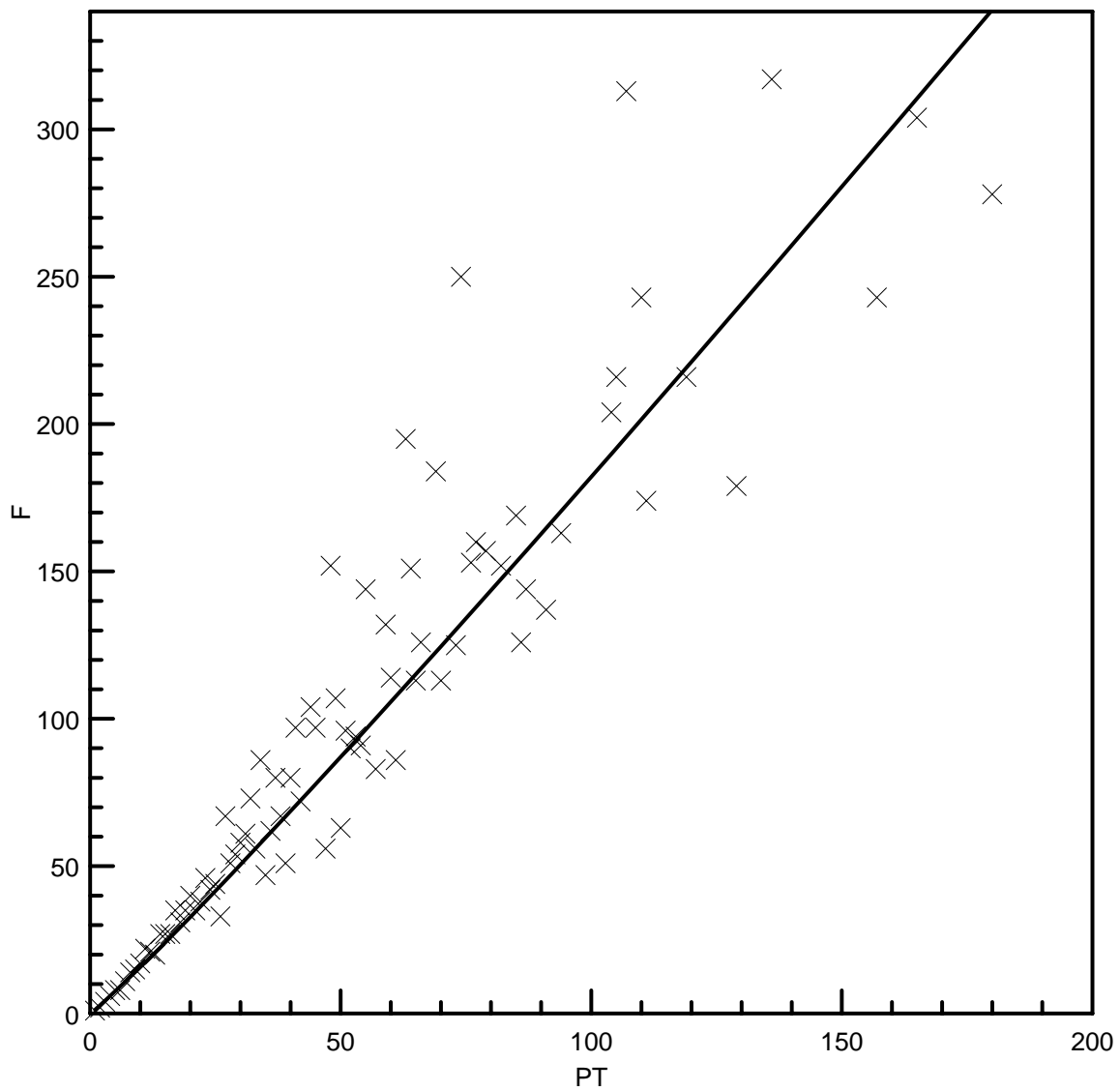


Рис. 29 в. Зависимость частоты (F) от политекстности (PT): словоформы существительных

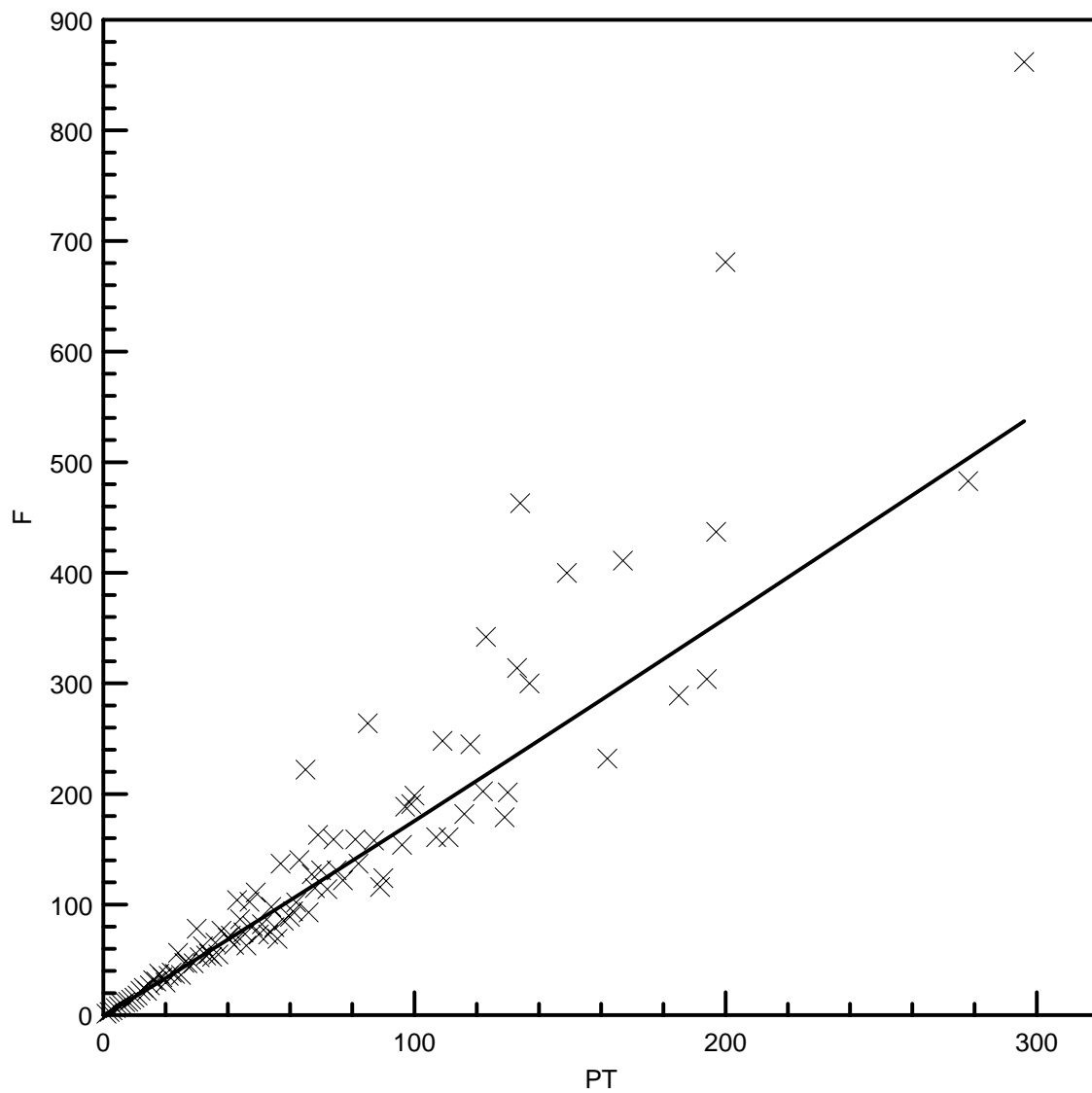


Рис. 29 г. Зависимость частоты (F) от политекстии (PT): леммы существительных

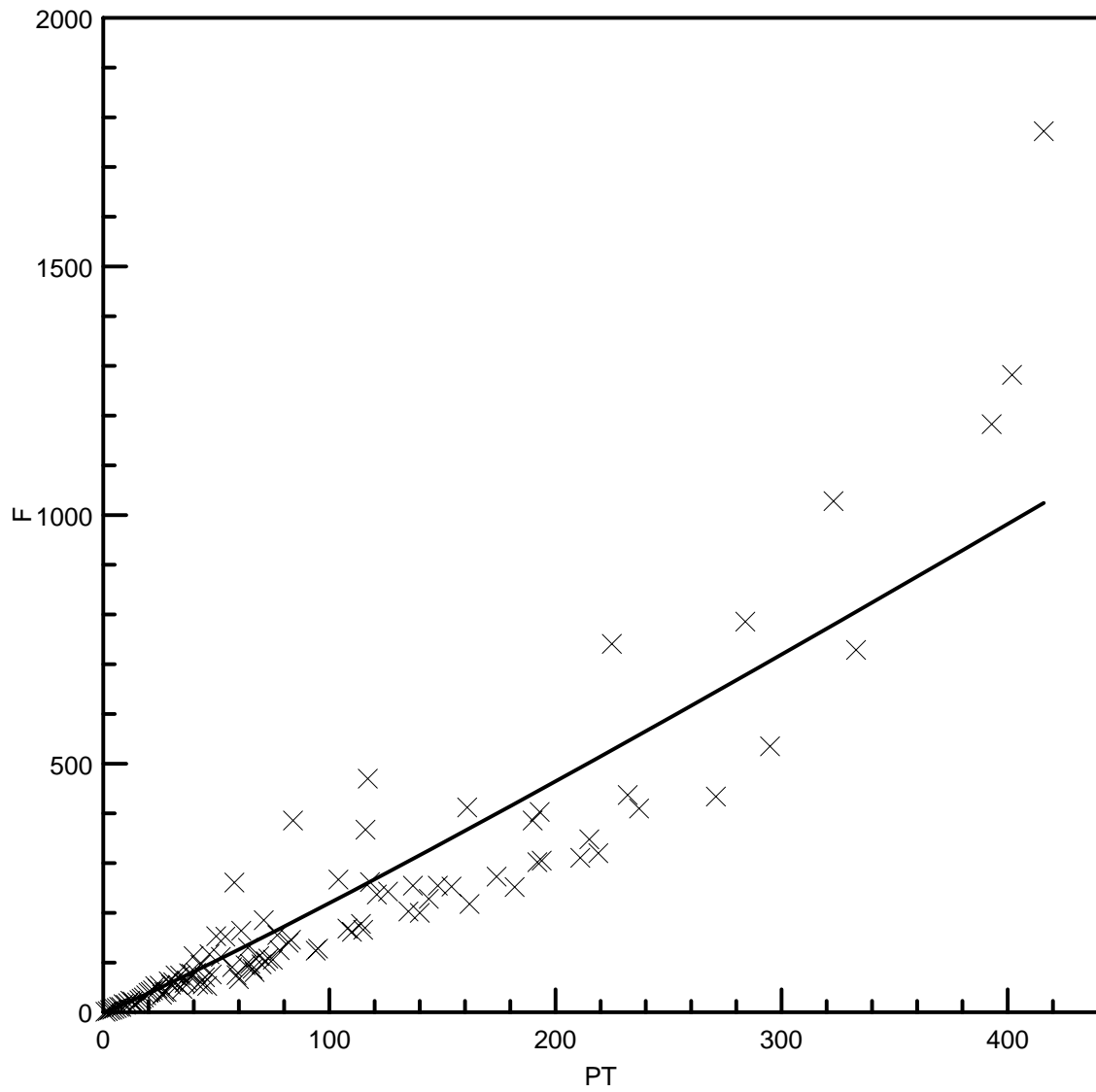


Рис. 29 д. Зависимость частоты (F) от политекстии (PT): словоформы прилагательных

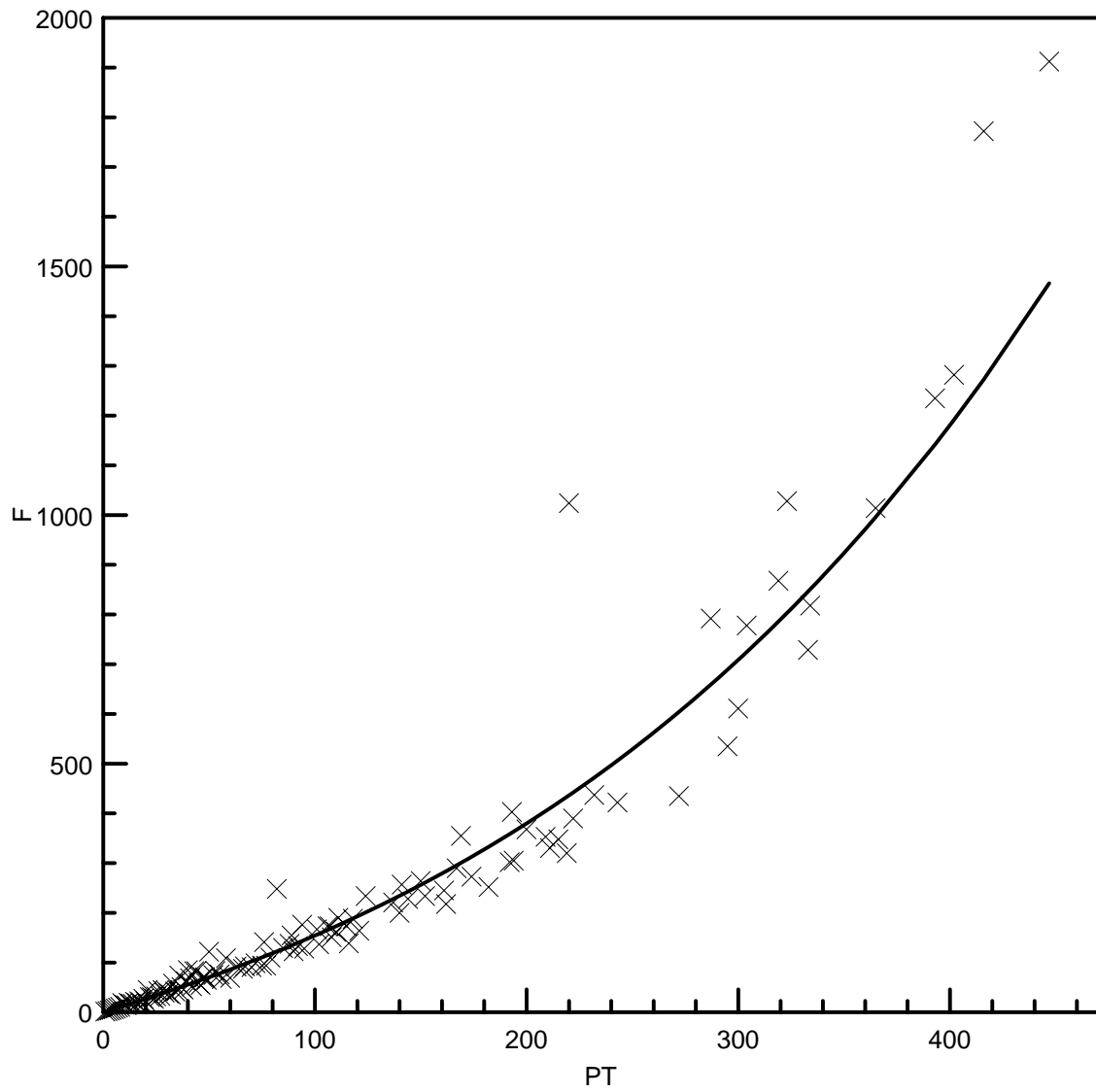


Рис. 29 е. Зависимость частоты (F) от политекстии (PT): леммы прилагательных

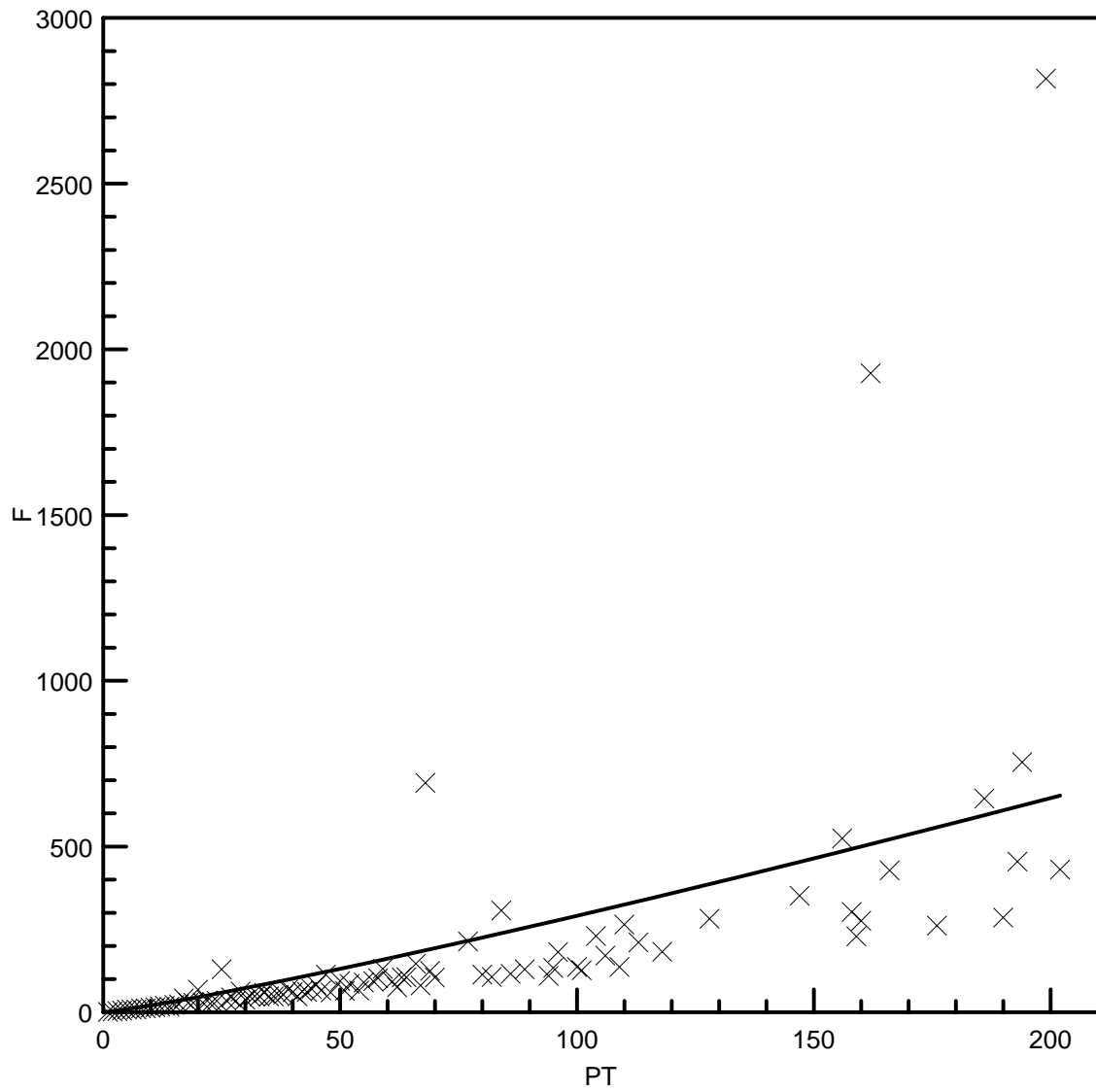


Рис. 29 ж. Зависимость частоты (F) от политекстности (PT): словоформы глаголов

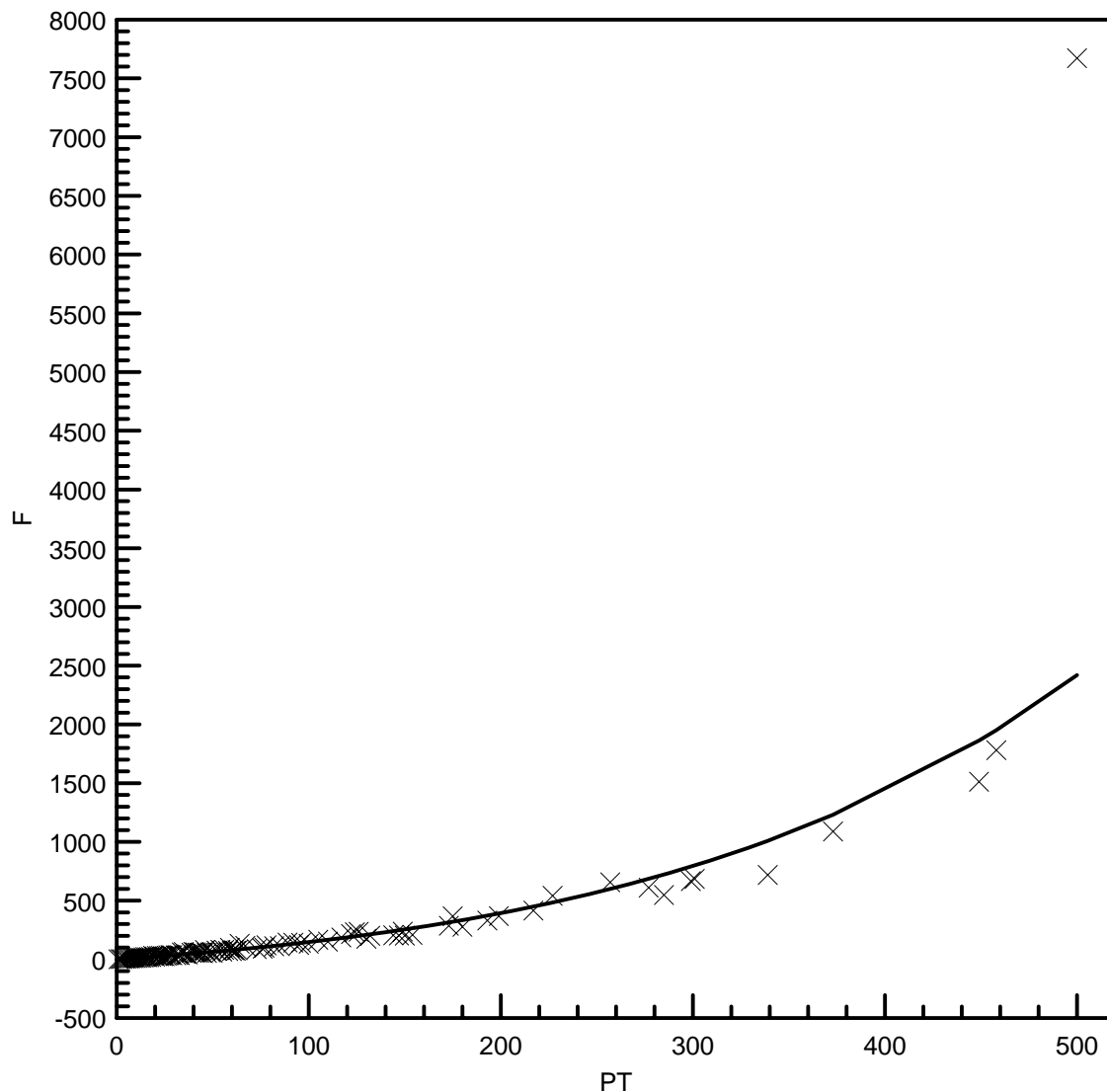


Рис. 29 з. Зависимость частоты (F) от политекстии (PT): леммы глаголов

7. ФЕНОМЕН ОСЦИЛЛЯЦИИ ЛЕКСИЧЕСКИХ ХАРАКТЕРИСТИК

При аппроксимации функции $y = Ax^B$ эмпирическим данным относительно зависимости длины от частоты возникла несколько неожиданная картина: качество аппроксимации согласно F-тесту не подлежит сомнению, и все-таки по сравнению с другими аппроксимациями создается впечатление, что точки данных сильно разбросаны вокруг теоретической функции. На рисунке 30 данные представлены в нелогарифмической форме. Эта тенденция несомненна, и результат статистического испытания свидетельствует о том, что объясненная дисперсия настолько велика по сравнению с необъясненной, что гипотезу нельзя отвергнуть. Однако можно

предположить, что за мнимо случайным разбросом скрывается систематическая методическая ошибка или еще не учтенное системное влияние. В первом случае все исследование следует повторить с целью исправления ошибки. Второй случай представляет интерес, поскольку такая "аномалия"⁷⁴ или должна привести к улучшению модели (в том случае, если соответствующее расширение или модификация позволит объяснить эту аномалию), или покажет несовместимость модели с наблюдениями. Если отклонения точек данных от теоретической кривой случайны, то их ход должен поддаваться сглаживанию по методу скользящих средних. Поэтому из той же выборки были получены данные, значения которых для длины получены из средних значений скользящего интервала длины в 20 единиц по шкале частоты.

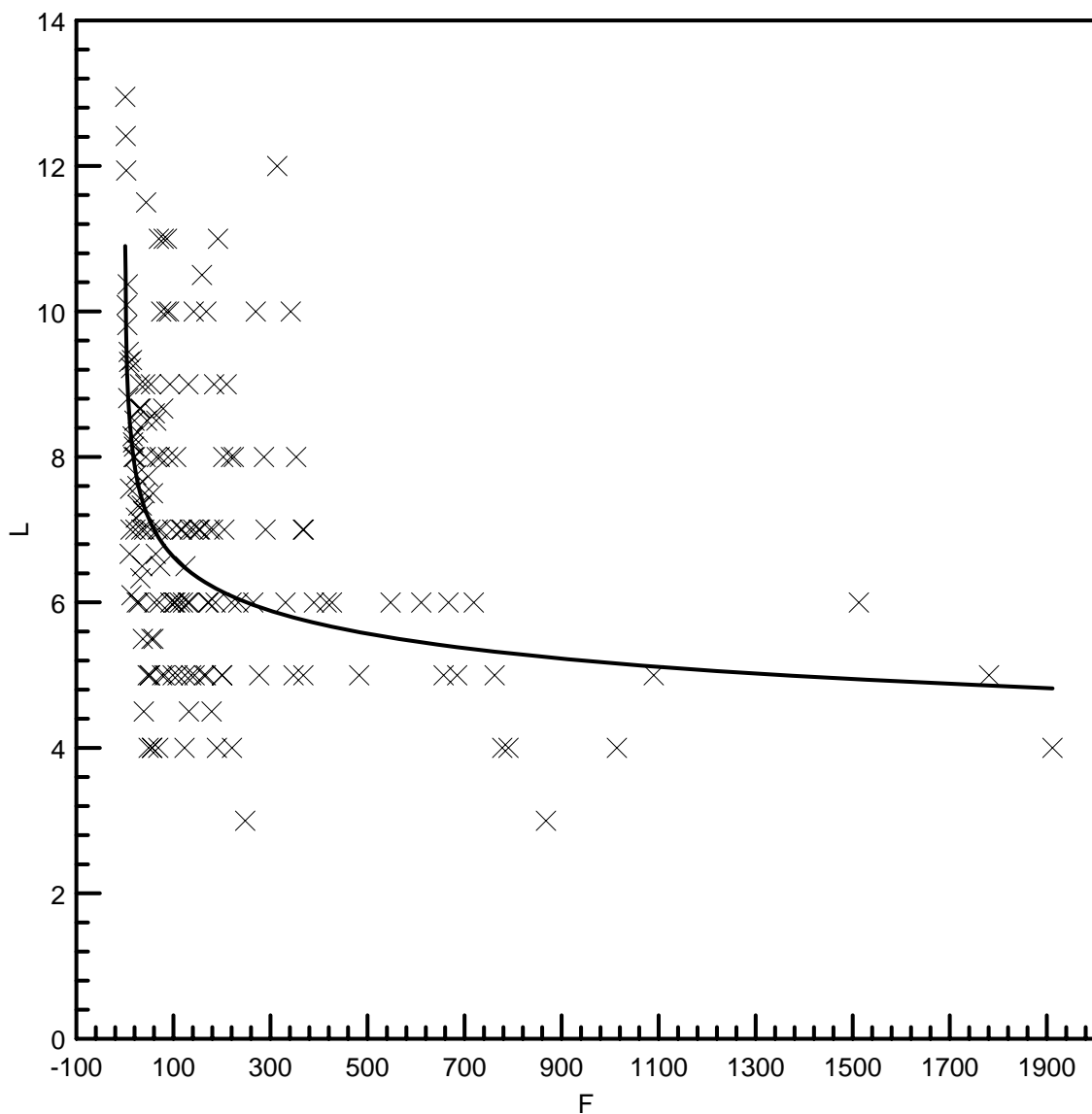


Рис. 30. Длина (L) в зависимости от частоты (F)

⁷⁴ Согласно терминологии Куна (Kuhn, 1967).

Вместо просто ожидаемого приближения точек к кривой возникла поразительная картина (ср. рис. 31): сглаженный ход точек данных образовал кривую, колеблющуюся вокруг теоретической гиперболической функции; качество аппроксимации (см. таблицу 17) функции, выведенной из модели, снова было превосходным.

Таблица 17. Аппроксимация функции $L = AF^B$ при скользящих средних значениях (интервал 20)

A	B	FG	F	$F_{0,01}$	$P(F)$
9,36	-0,066	1,138	138	6,83	$1,7 \cdot 10^{-11}$

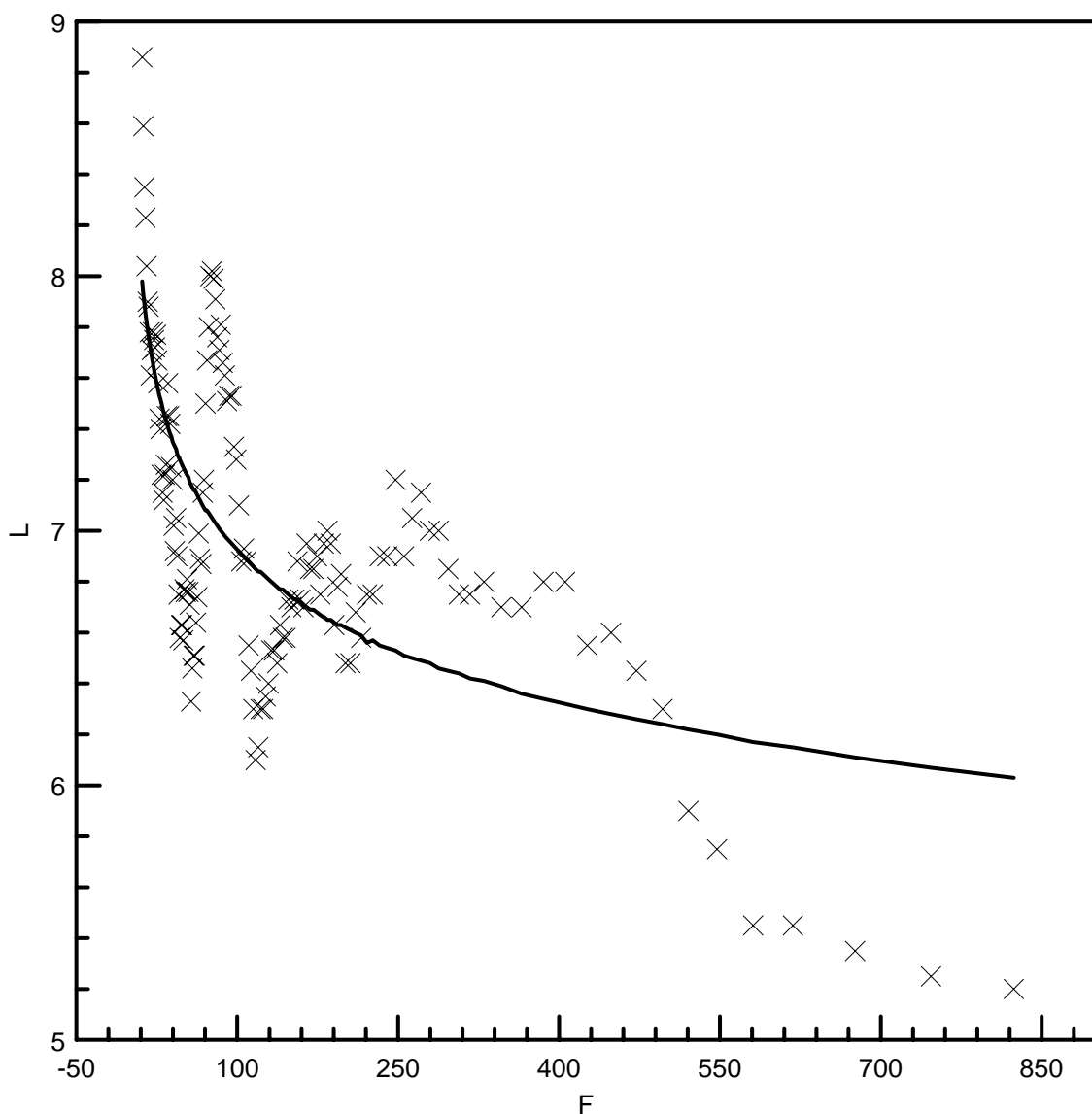


Рис. 31. Длина (L) в зависимости от частоты (F). Рассчитано с помощью скользящего среднего при интервале 20.

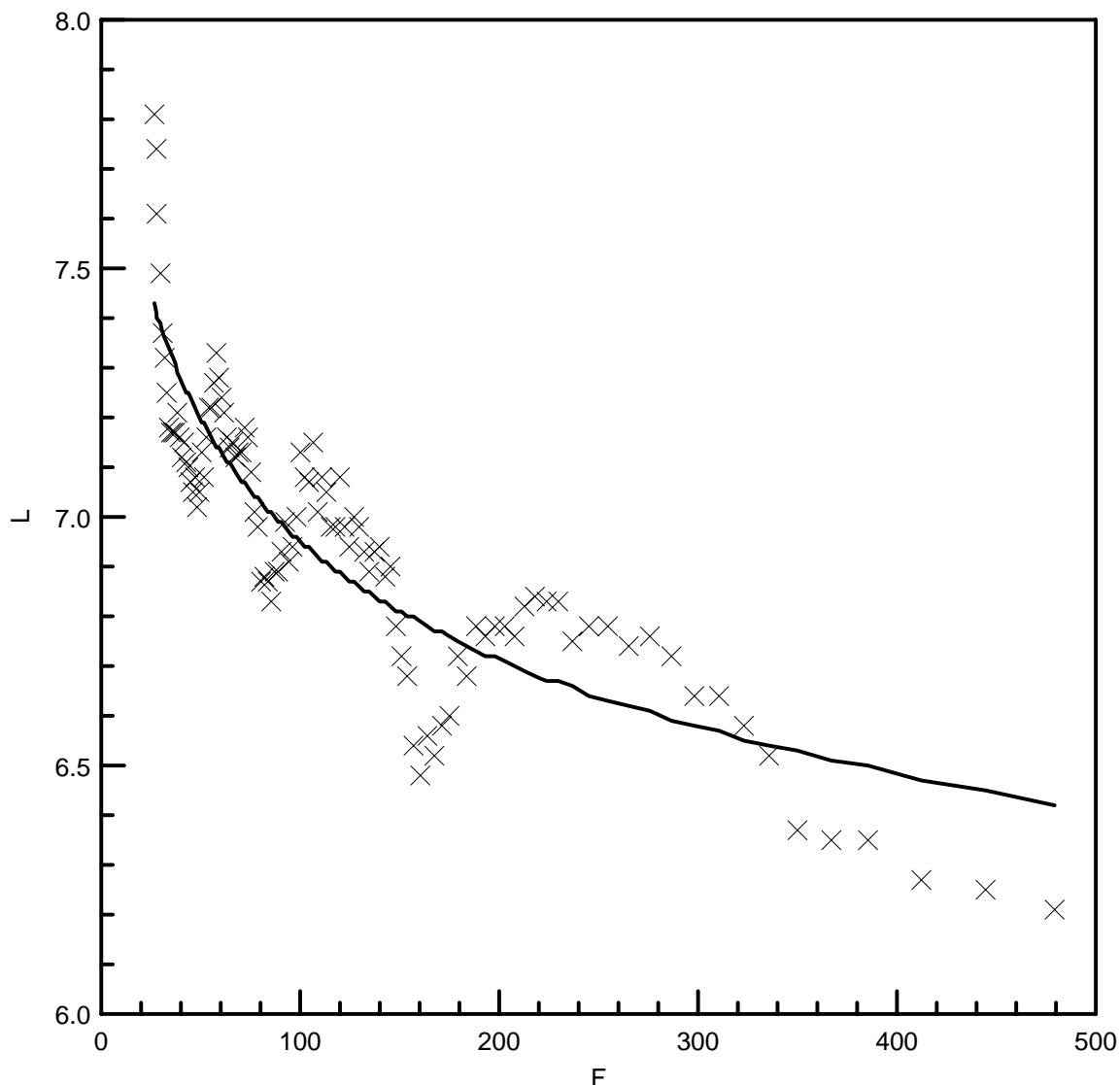


Рис. 32. Длина (L) в зависимости от частоты (F). Рассчитано с помощью скользящего среднего при интервале 50.

Для проверки стабильности этого явления аппроксимация была несколько раз повторена с различными интервалами (ср. рисунки 32 и 33 и таблицы 18 и 19). Наблюдаемая осцилляция вокруг предсказанной кривой проявлялась всякий раз с различным разложением, но с несомненной отчетливостью. Случайное происхождение этого явления можно почти с полной уверенностью исключить. В связи с этим перед нами встает задача – найти лингвистическое объяснение данного явления. На первый взгляд кажется, что не существует какой-либо известной связи, способной вызвать колебательный процесс в лексике языка, тем более, что осцилляция зарегистрирована не во временном измерении, а по оси частоты.

Таблица 18. Аппроксимация функции $L = AF^B$ при скользящих средних значениях (интервал 50)

A	B	FG	F	$F_{0,01}$	P(F)
8,76	-0,05	1,108	238	6,75	$1,5 \cdot 10^{-11}$

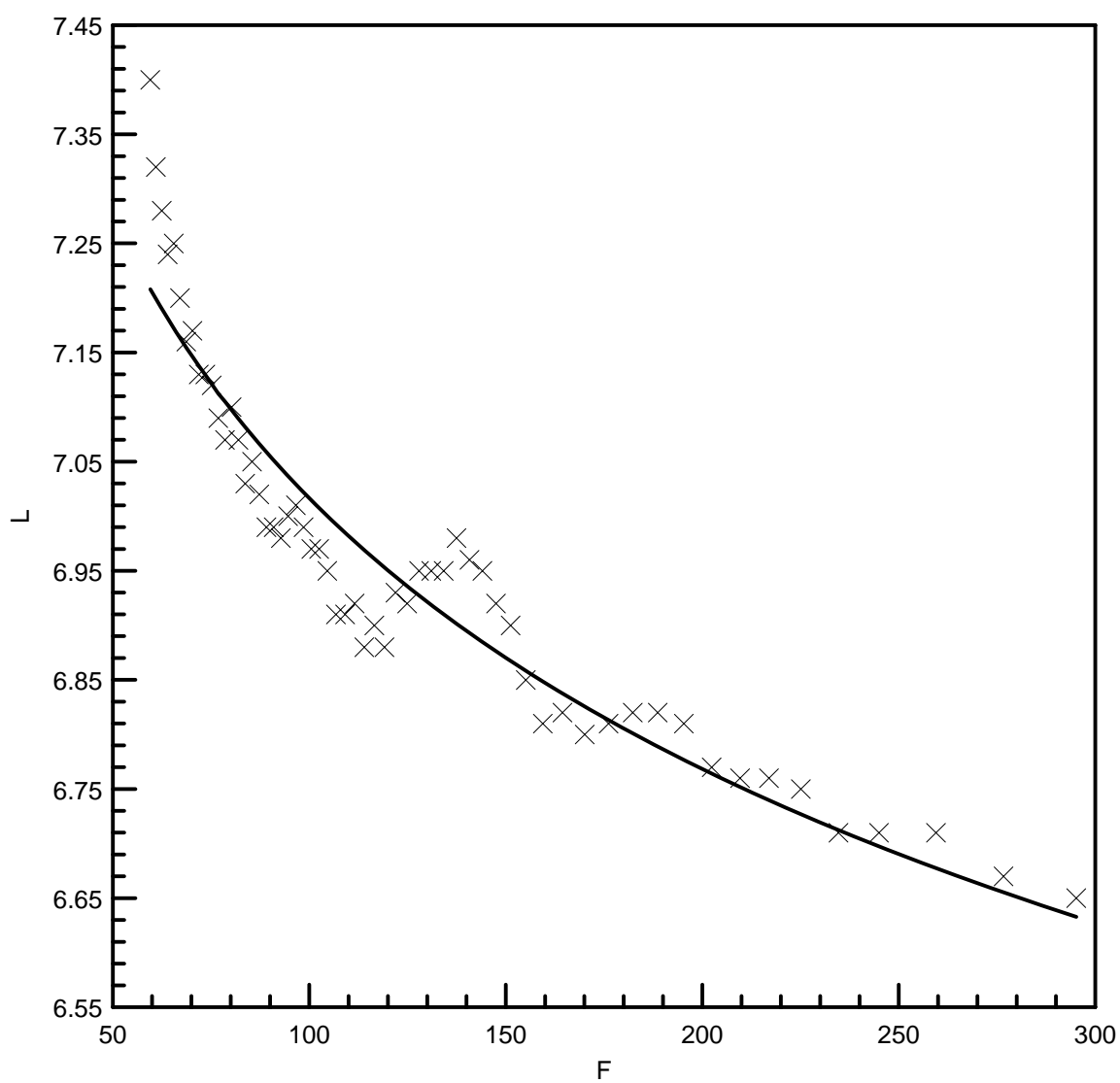


Рис. 33. Длина (L) в зависимости от частоты (F). Рассчитано с помощью скользящего среднего при интервале 100.

Таблица 19. Аппроксимация функции $L = AF^B$ при скользящих средних значениях (Интервал 100)

A	B	FG	F	$F_{0,01}$	$P(F)$
8,91	-0,052	1,58	276	6,75	$7 \cdot 10^{-12}$

Здесь мы возвращаемся к тому соображению, которое было высказано при конструировании модели, но не было тогда развито: структура модели в разделе о связи между длиной и частотой основана на предположении, что скорость изменения длины обратно-пропорциональна частоте лексической единицы. Это предположение мы выразили в виде дифференциального уравнения 1

$$\frac{y'}{y} = \frac{-N}{x}.$$

Этот подход мы хотим уточнить следующим дополнительным соображением: лексическая единица сокращается в тем большей степени, чем она длиннее. Выражение, уже короткое по причине частого употребления, не подвергается значительному дополнительному сокращению при очень высокой или даже растущей частоте. Так, нельзя ожидать, что такие русские слова, как *год*, *муж* или *плуг* могут стать еще короче. Напротив, вполне вероятно, что выражения *Katalysator* (катализатор) или *Umweltverschmutzung* (загрязнение окружающей среды) при увеличении частоты употребления в скором будущем подвергнутся такому сокращению или же будут заменены более короткими⁷⁵. Таким образом, сокращение слова может ускоряться или тормозиться в зависимости от его длины. Сформулируем в виде гипотезы:

Измерение скорости сокращения лексической единицы зависит от длины и частоты.

Для точного представления этого допущения в виде дифференциального уравнения имеется несколько очевидных возможностей; для однозначного выбора среди них у нас еще нет достаточных теоретических знаний. Однако, во всяком случае, утверждению соответствует дифференциальное уравнение второго порядка. Убедительно выглядит следующий подход:

$$L'' = -qL + ke^{-rF},$$

где L – длина, F – частота, q , r – константы. Это уравнение выражает линейную связь, согласно которой принимается, что:

⁷⁵ Слово *Bahn* вместо *Eisenbahn* (железная дорога), *Cafete* вместо *Cafeteria* (кафетерий) и *Micro* вместо *Microcomputer* – примеры, которые можно наблюдать в настоящее время. Эти процессы имеют различную скорость, которая определяется различным темпом роста частоты

а) при большой длине происходит небольшое изменение сокращения (= слабое торможение еще сильного процесса сокращения), при малой длине происходит сильное изменение сокращения (=сильное торможение уже незначительного сокращения);

б) существует экспоненциальная убывающая положительная зависимость от частоты.

Поэтому перед q, k, r ставится положительный знак. Корни характеристического уравнения – iq и $-iq$. Поскольку речь идет о комплексных числах, решение однородного уравнения (при $Q = q$)

$$L_h = a \cos QF + b \sin QF$$

является осциллирующей функцией от частоты. При частном решении $L_f = ce^{dF}$ получаем

$$L = a \cos QF + b \sin QF + ce^{dF} .$$

Эта функция является экспоненциальной функцией с модуляционным колебанием. В данном случае от попытки аппроксимации к данным придется отказаться. Здесь мы не будем проводить подробного решения этой проблемы (с точки зрения объема это потребовало бы дополнительного исследования), а попытаемся рассмотреть принципиальную возможность включения наблюдаемого явления в структуру модели. Осцилляцию длины в зависимости частоты можно объяснить, приняв допущение о переменной скорости сокращения, которая зависит от длины. Колебательная функция с уменьшающейся общей амплитудой получается при использовании линейного дифференциального уравнения второго порядка при соблюдении определенных условий:

а) изменение скорости сокращения прямо пропорционально длине

б) изменение скорости сокращения не зависит от сокращения; другими словами, коэффициент члена первого порядка меньше чем $1/4$ коэффициента нулевого порядка. Таким образом, если дифференциальное уравнение имеет общий вид

$$L'' + aL' + bL = f(F),$$

то в качестве решения получается осциллирующая функция в тех случаях, когда $4f < b$. Если $f = 0$, то колебание не затухает.

в) функция $f(F)$ монотонно убывает.

Первое условие, по-видимому, всегда выполняется для естественных языков, о втором сейчас мы еще ничего не можем сказать; из рассмотренных данных можно сделать только один вывод для немецкого языка. Третье условие непосредственно вытекает из модели.

8. ОТ БАЗОВОЙ МОДЕЛИ – К ТЕОРИИ ЯЗЫКА

8.1 ЗНАЧЕНИЕ РАЗРАБОТАННОЙ МОДЕЛИ

Основной целью настоящего исследования была разработка модели, объясняющей ряд до сих пор изолированных, обособленных языковых законов на основе единого подхода, и являющейся, кроме того, достаточно сложной для того, чтобы из нее можно было вывести до сих пор неизвестные законы. Описанная модель охватывает шесть лингвистических величин, десять связей с внешней средой в виде потребностей и двенадцать процессов, интерпретированных с лингвистической точки зрения⁷⁶. Функции, выводимые из предлагаемой модели, согласуются со связями переменных, известными до сих пор; они даже формально идентичны с законом Менцерата, открытым Альтманом (1980). Кроме того, из структуры модели был выведен ряд до сих пор неизвестных зависимостей и процессов. Еще одно следствие заключается в том, что распределения характеристик лексических единиц являются функциями друг друга. Как отдельные следствия, теоретически выведенные из структуры модели, так и модель в целом были подвергнуты эмпирической проверке на корпусе немецкоязычных текстов и нашли подтверждение.

Вытекающие из данных важные следствия⁷⁷ полностью согласуются с предположениями, лежащими в основе модели. Прямые и даже многократно опосредованные испытания предсказанных зависимостей переменных дают результаты на высоком уровне значимости.

Неожиданно обнаруженный факт колебания в лексическом пространстве по оси частоты легко был объяснен путем введения еще ранее предполагавшейся обратной связи. Основная мысль работы Ципфа, принцип наименьших усилий и его факторы унификации и диверсификации, нашли в модели подтверждение и новую интерпретацию на основе функционального анализа и синергетики. Существенной чертой предлагаемого подхода является структурная аксиома языковой саморегуляции, в том варианте, на котором настаивает Альтман (1981) для лингвистического функционального анализа. Модель построена таким образом, что ее можно включить в широкую коммуникативную модель на основе социологического и психофизического функционального анализа. В качестве связующих звеньев служат потребности и параметры функций в той степени, насколько не определены еще внутри модели. Это свойство модели позволяет выводить саму структуру системы из еще более высоких принципов.

8.2 НЕКОТОРЫЕ НАБРОСКИ ДАЛЬНЕЙШЕГО РАЗВИТИЯ МОДЕЛИ

По этим причинам модель можно охарактеризовать как первоначальную попытку создания языковой теории. Разработанное здесь ядро, безусловно, не является единственным и лучшим из возможных подходов к решению данной задачи; большинство других величин подходит для этого, несомненно, столь же удачно, поскольку при моделировании важен не исходный пункт, а способность модели к

⁷⁶ См. таблицы 1 – 3

⁷⁷ Ср. аппроксимацию зависимости частоты от политекстии, результат которой может указывать на существование еще не учтенной величины.

расширению. В этой связи следует рассмотреть в общем виде еще некоторые аспекты развития модели.

8.2.1 Расширение предметной области и уточнение структуры.

Специфическое лингвистическое значение развития модели будет заключаться в ее поэтапном распространении на более широкий круг явлений. При этом можно различать два взаимодополняющих вида действий: теоретически и эмпирически мотивированное расширение. В первом случае включение новых величин или моделирование дополнительных наборов правил осуществляется на основе уже утвердившегося лингвистического знания или обоснованных допущений. Так, например, предварительное моделирование фонологических, морфологических, синтаксических и лексотаксических связей, а также сопряжение получаемых подсистем относительно легко осуществить на основе понятийной системы подходящей грамматической модели⁷⁸. Так, отношение конституэнтности лингвистических единиц различных уровней в системе-модели находит выражение непосредственно в связи между величинами "объем инвентаря (единиц более низкого уровня)" и "длина (единиц вышестоящего уровня)". Синтаксические свойства определенного языка учитываются в параметре синтетичности (напр., в связи с потребностью в уточнении значения), в параметру класса слов и, естественно, в характеристиках связанного с предложением набора правил. Подобным образом можно построить большое число приблизительных структур; однако более точное моделирование подсистем и их взаимоотношений требует, безусловно, также эмпирического изучения. Иерархизация наборов правил и процессов была осуществлена тоже на основе теоретических соображений. В качестве примера очень высокого уровня саморегуляции здесь в модель следует также включить регулирующее влияние эффективности управления⁷⁹. Под этим термином мы понимаем меру того, насколько сильно импульсы из системного окружения могут воздействовать на модель в определенный момент времени. С одной стороны, это влияние потребностей (или корректировка системы в соответствии с потребностями) необходимо для приспособления к изменчивым обстоятельствам. С другой стороны, слишком быстрая реакция процессов на такие изменения может привести к катастрофе. Это связано с тем, что с различными элементами системы происходят изменения, которые нельзя связать между собой; такая ситуация может возникнуть, когда приспособление происходит быстрее, чем внутрисистемный поток сигналов. Для члена человеческого языкового коллектива такая катастрофа могла бы проявиться в том, что язык изменяется быстрее, чем он может к этому приспособиться, в результате чего язык функционально перестанет быть для него средством коммуникации. Поэтому потребности в способности приспособления системы противостоят потребности в стабильности. Она действует тормозящим и сдерживающим образом в целях сохранения системы. На рисунке 34 в схематическом и сильно упрощенном виде показано регулирующее влияние эффективности управления через равновесие между приспособлением и стабилизацией. При непосредственном включении этого уровня управления в модель следует учитывать, что связям между каждой потребностью и находящейся под ее влиянием системной величиной соответствуют индивидуальные проявления эффективности управления.

⁷⁸ Здесь в первую очередь имеется в виду стратификационная грамматика (Lamb: 1966; см. т. Lockwood: 1972 ; Schreyer: 1977)

⁷⁹ Ср. Köhler, Altmann (1983)

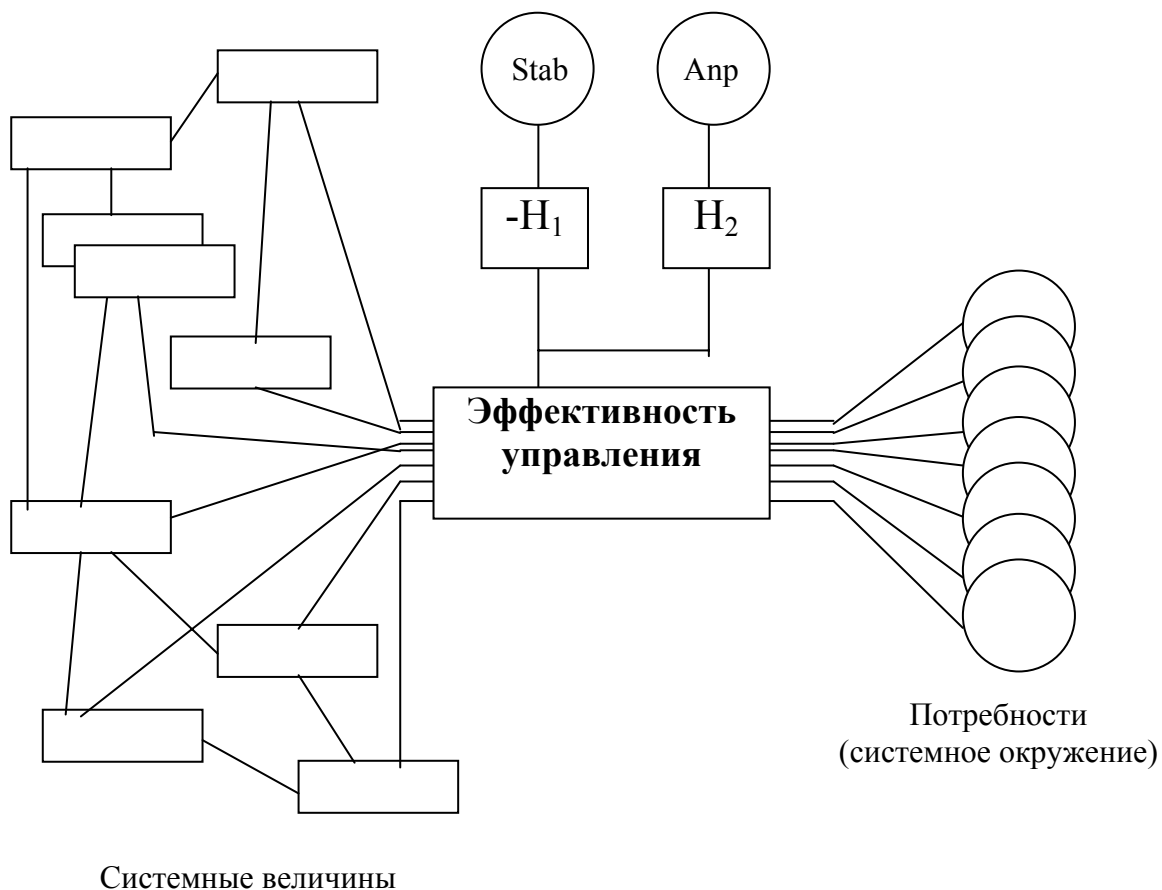


Рис. 34. Регулирующее влияние эффективности, с которой системное окружение действует на систему через потребности в стабильности и приспособлении.

Другие шаги усовершенствования модели были осуществлены из эмпирически очевидных соображений. Так, например, для каждой связи переменных можно измерить, какая доля изменчивости не объяснена еще на основе модели и требует введения дополнительной величины.

8.2.2 Взаимодействие между системой и ее окружением

От этой области дальнейшего развития модели следует ожидать, что она в первую очередь будет связана с поиском потребностей. До сих пор лингвистические рассуждения с функционально-аналитической целевой установкой были еще очень далеки от достижения результатов, пригодных для реализации; формулировки "коммуникативных потребностей" остаются еще очень неконкретными и декларативными (ср. Kanngiesser 1977).

По сравнению с выявлением потребностей более трудной задачей будет отнесение функциональных эквивалентов к самой трудной – указание отношений эквивалентов между собой.

Чтобы прийти однажды к серьезной языковой модели, одновременно отражающей отношение языка и общества и психофизиологические предпосылки языка, следует сделать предметом исследования и обратное влияние системы на ее окружение. Так же, как к модели (не к системе) принадлежат входные величины (потребности), к ней

принадлежат и выходные. Они являются ключом к различным областям исследования, начиная с "языковой относительности" (в смысле Сэпира – Уорфа) и кончая "использованием языка в качестве инструмента политики".

8.2.3 Методика моделирования

Несомненно, с течением времени методы моделирования развиваются и совершенствуются. Уже сейчас можно предположить, что линейных операторов и алгебры графов не достаточно, даже когда почти все связи можно линеаризировать путем логарифмирования. Возможное решение может также состоять в том, чтобы осуществлять моделирование с помощью не дифференциальных уравнений, а стохастических процессов; это позволило бы также упростить вывод распределений лингвистических величин. При разработке методики большую пользу принесет, несомненно, междисциплинарное сотрудничество с учеными, которые в других областях занимаются саморегулирующимися системами – с синергетиками –, тем более что изучение кооперативных явлений в некоторых дисциплинах достигло значительных результатов.

8.2.4 Временной континуум системы

Начиная с достаточно высокого уровня сложности будет представлять интерес учет временной оси. Этот аспект динамики не рассматривался здесь до сих пор в первую очередь потому, что (как раз для немецкого языка) почти не имеется надежных лингвистических данных. В последующие годы станет возможным как проведение эмпирических исследований по проверке современного состояния модели также на языках, для которых имеются более качественные диахронные данные⁸⁰, так и разработка в конечном счете усовершенствованной комплексной модели, которую после тщательной эмпирической проверки можно будет использовать для имитационного моделирования.

Лингвистическое имитационное моделирование дает языкознанию возможность проводить квазиэкспериментальные исследования. Тем самым становится возможным в условиях, приближенных к лабораторным, изучать процессы, которые в действительности происходят один раз и не повторяются.

⁸⁰ Напр., для русского, английского или шведского языков

9. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

После прочтения этой книги у читателя не должно возникнуть впечатление, что намеченные выше шаги указывают путь к долгожданной теории языка. Осталось еще достаточно много открытых вопросов; мы ни разу не коснулись многих областей, для того чтобы контуры этой теории стали бы видны уже сегодня. Так, в предложенной модели почти не удалось задействовать такие области, как синтаксис, прагматика, и, в первую очередь, семантика. Были предложены только выходы на эти лингвистические аспекты, а не точные представления о том, как включить их в модель. С другой стороны, предложенный подход – особенно благодаря фундаментальной аксиоме саморегуляции – дает возможность объединить в одной модели все аспекты лингвистических исследований. Кроме того, с помощью этого подхода лингвистика вместе с целым рядом смежных современных естественных наук может внести вклад в междисциплинарную теорию синергетики и извлечь из этого значительную методологическую пользу для себя. В отношении собственного предмета исследования для лингвистики это означает также, что можно будет наконец снять или решить ряд мнимых вопросов.

Не в последнюю очередь эти мнимые вопросы касаются определения самого предмета лингвистических исследований. Никакая попытка отграничить определенный язык территориально, этнически или исторически не может быть удовлетворена с лингвистической точки зрения. Во всяком случае результат или слишком неоднороден для возможного единого описания, или слишком ограничен, чтобы назвать его языком: ”язык” имеет нечеткое определение.

В свете применяемого подхода становится ясно, что классификационное понятие ”язык” в смысле национального языка столь же недостаточно, как и другие категориальные понятия. Оно пригодно для определенных целей, однако злоупотребление им ведет к ошибке. Для пояснения можно сделать следующее сравнение: применение этого классификационного понятия похоже на попытку вычленивать и снабдить этикеткой одну отдельную волну в Атлантическом океане. Как и волна, отдельный язык, если язык понимать как саморегулирующуюся систему, является микроструктурой, которая выделена вниманием наблюдателя, ограничена во времени и возникает под действием кооперативных процессов внутри системы и внешних влияний. Различия между внешними влияниями, т.е. потребности, имеющие различия в географическом, культурном, социологическом и историческом плане, с помощью механизмов саморегуляции вызывают к жизни различные адаптивные процессы. Если стабильность по сравнению с адаптацией может быть сохранена лишь локально, если влияние приспособления в условиях несовместимости потребностей перевешивает, то возникают ”языки”, диалекты, социолекты и профессиональные языки. При соответствующих внешних условиях, т.е. при конвергенции системных потребностей языки и разновидности языков могут также ”сливаться”.

Столь же ясно проявляется при системно-теоретическом моделировании соотношение между лингвистическим описанием и лингвистическим объяснением. Описания, сформулированные с помощью грамматик, словарей, статистических данных и др., являются отображениями фрагментов моментальных снимков динамического развития языковой системы. О лингвистической теории можно будет говорить только тогда, когда из нее можно будет вывести т.е. объяснить и предсказать свойства лексики, грамматические закономерности (отклонения от нормы) и другие аспекты лингвистического описания.

Послесловие: новейшие исследования и дальнейшие перспективы

1. Над немецким оригиналом данного издания, вышедшим в 1986 г., автор работал в 1984–1985 гг. Вскоре Ю.А. Тулдава в одном из сборников "Квантитативная лингвистика и автоматический анализ текста" обратил внимание на данную книгу и опубликовал краткую рецензию на русском языке. Интерес к книге российских исследователей побудил руководителя исследовательского коллектива А.А. Поликарпова поручить своему сотруднику А.В. Малову выполнить перевод книги. Я хотел бы воспользоваться возможностью и еще раз выразить свою признательность им обоим. В работе над переработкой текста первого перевода книги автору оказали неоценимую услугу М. Нуртазина и В. Баскевич, которым я также глубоко благодарен. Ответственность за все не выявленные несоответствия текстов несую, разумеется, я. Русское издание полностью идентично немецкому, исправлены лишь выявленные ошибки и заново оформлены таблицы и диаграммы. Целесообразным оказалось также дать краткий обзор работ, опубликованных с момента выхода оригинального издания и привести полученные результаты. Я благодарен В.В. Кромеру за готовность перевести данное послесловие на русский язык и правку перевода книги.

2. Тематика данной книги ограничена исследованием и моделированием лексической системы естественных языков. Последующие работы автора книги и других исследователей раздвинули рамки применимости предложенной модели и затронули также области других дисциплин. В качестве примера можно привести семиотику (Köhler 1988) и музыковедение (Köhler 1995 и Köhler/Martináková 1998). Динамике языковых изменений как результату действия процессов саморегуляции и самоорганизации посвящена работа (Köhler 1987, 1990с). Представленная синергетическая модель подверглась развитию в ходе развернувшегося после опубликования немецкого оригинала обсуждения книги специалистами. В ходе дискуссии выявилось, что введенная терминология подлежит уточнению, а некоторые аспекты функционирования модели нуждаются в более тщательном описании. Дальнейшие уточнения и обобщения, например, введение понятия об иерархических структурах, были проведены в работах (Köhler 1990a, 1990b и 1990в).

В 1991 г. Рольф Хаммерль опубликовал свою докторскую диссертацию, где подробно описал приложение синергетической модели к лексике польского языка. В ряде выпускных работ (магистерских диссертаций), в основном студентов Трирского университета, была проведена проверка гипотез синергетической модели на материале различных языков. Была проверена возможность моделирования новых языковых подсистем и аспектов (например, морфологии производных слов и флексий, фонетики). Эти работы к настоящему времени не опубликованы. Некоторые из них (в сокращенном виде) публикуются в книге (Köhler 2002). Особого упоминания заслуживают перечисленные ниже работы.

Работа Кристины Хоффманн посвящена методу непосредственных составляющих Хокинса, объясняющему универсальные предпочтения в порядке слов. Андреа Кротт представила синергетическую модель словообразования. В модели введен ряд предположений, проверяемых на эмпирических данных трех языков (немецкого, английского, нидерландского).

Корнелия Менцель исследовала возможность приложения основных положений синергетической лингвистики (справедливых для лексики) к системе письменности китайского языка. Исследование показало, что свойства иероглифической системы аналогичны свойствам лексической системы, аналогии прослеживаются также во внутренних связях систем. На базе трех текстовых корпусов и одиночного текста Катрин Гизекинг исследовала возможность распространения основных понятий синергетической лингвистики на лексическую систему английского языка.

Синергетическая модель, увязанная с когнитивными представлениями в области лексической семантики, впервые описана Кёлером и Альманном (Köhler/Altmann 1993). Работа (Köhler/Altmann 1996) посвящена определенным аспектам математического моделирования, позволяющим объяснить некоторые языковые феномены

Главной темой докторской диссертации Эдды Леопольд (Leopold 1998) является математическое моделирование возникновения и распространения языковых изменений на уровне микропроцессов.

Харуко Санада-Його опубликовала в 1999 г. часть своей диссертации, в которой было проведено тщательное исследование возможности приложения синергетической лингвистики к лексике японского языка. В области синтаксиса возможности синергетического моделирования впервые исследовали Köhler (1999) и Köhler/Altmann (2000). Авторами сконструирована синтаксическая базовая модель, оперирующая системными величинами *сложность, частота, позиция, информация, уровень вложения, длина*, а также зависимостями между введенными переменными. Базовая модель увязывает новые гипотезы с известными в лингвистике и типологии закономерностями (методом непосредственных составляющих Хокинса, предпочтением правого ветвления Ингве и законом Менцерата-Альтманна). Смоделировано распределение вероятностей системных величин на основе выдвинутых предположений об их взаимосвязи.

2001

The distribution of some syntactic construction types in text blocks. In: Uhlířová, Ludmila; Wimmer, Gejza; Altmann, Gabriel; Köhler, Reinhard [Eds.]: Text as a linguistic paradigm: levels, constituents, constructs. Festschrift in honour of Luděk Hřebíček. Trier: Wissenschaftlicher Verlag Trier. 136-148

2000

Probability distributions of syntactic units and properties. Journal of Quantitative Linguistics 7,3, 189-200

1999

Syntactic structures. Properties and interrelations. Journal of Quantitative Linguistics 6,1, 46-57

1998

(вместе с Zuzana Martináková-Rendeková): A Systems theoretical approach to language and music. In: Altmann, G. und Koch, W.A. [Eds.]: Systems. New Paradigms for the Human Sciences, Berlin, New York: de Gruyter 1998, 513-546

1996

(вместе с Altmann, Gabriel): "Language Forces" and Synergetic Modelling of Language Phenomena. In: Schmidt, Peter [Ed.]: Glottometrika 15. Issues in General Linguistic Theory and The Theory of Word Length. Bochum: Brockmeyer, 62-76

1995

Der synergetische Ansatz in der Linguistik, seine Übertragbarkeit und die Rolle der Einheiten. Musikometrika 6(1995), 13-25

1994

Synergetic Linguistics. In: Asher, R.E. [Ed.]: The Encyclopedia of Language and Linguistics. Oxford, New York, Seoul, Tokyo: Pergamon Press, 4454-4455

1993

(вместе с Altmann, Gabriel): Begriffsdynamik und Lexikonstruktur. In: Beckmann, F. ; Heyer, G. [Hrsg.]: Theorie und Praxis des Lexikons. Berlin. New York de Gruyter, 1993, 173-190

Synergetic linguistics. In: Köhler, Reinhard ; Rieger, Burghard [Eds.]: Contributions to quantitative linguistics. Proceedings of the 1st Quantitative Linguistics Conference, QUALICO, University of Trier, 1991. Dordrecht: Kluwer, 1993, 41-51

1991

Основы синергетической лингвистики. В сб. Эврические возможности квантитативных методов исследования языка /под ред. Сильницкого Г.Г.; Тулдава Ю.А.; Поликарпова А.А./ Смоленск, 1991, 77-78

(вместе с Altmann, Gabriel): Synergetick aspekty jazykovedy. Jazykovedni Casopis 42 (1991), 3-14

1990

Elemente der synergetischen Linguistik. Glottometrika 12 (1990), 179-187.

Linguistische Analyseebenen, Hierarchisierung und Erklärung im Modell der sprachlichen Selbstregulation. *Glottometrika* 11 (1990), 1-18

Synergetik und sprachliche Dynamik. In: Koch, W. A. [Hrsg.]: *Natürlichkeit der Sprache und Kultur*. Bochum: Brockmeyer, 1990, 96-112

Zur Charakteristik dynamischer Modelle. Anmerkungen zu einem Beitrag von R. Hammerl und J. Maj. *Glottometrika* 11 (1990), 41-46

1988

Selbstregulation der Lexik. In: Bluhme, H. [Hrsg.]: *Beiträge zur quantitativen Linguistik*. Tübingen: Narr, 1988, 156-166

1987

Sprachliche Selbstregulation als Mechanismus des Sprachwandels. In: Boretzky, N. ; Enninger, W. ; Stolz, Th. [Hrsg.]: *Beiträge zum 3. Essener Kolloquium über Sprachwandel und seine bestimmenden Faktoren*. Bochum: Brockmeyer, 185-200

Systems theoretical linguistics. *Theoretical Linguistics* 14 (1987), 241-257

1988

Semiotik und Synergetik In: Posner, R.; Robering, K; Sebeok, Th. [Hrsg.]: *Semiotik. Ein Handbuch zu den Grundlagen von Natur und Kultur*. Berlin, New York: de Gruyter (erscheint).

1986

(вместе с Altmann, Gabriel): *Synergetische Aspekte der Linguistik*. *Zeitschrift für Sprachwissenschaft* 5 (1986), 253-265.

1983

(вместе с Altmann, Gabriel): *Systemtheorie und Semiotik*. *Zeitschrift für Semiotik* 5 (1983) 4, 424-431.

P.K. 2001

Литература

Altmann, G.

- 1980 Prolegomena to Menzerath's Law. In: Grotjahn, R. [Hrsg.]: Glottometrika 2. Bochum, 1-10
- 1981 Zur Funktionalanalyse in der Linguistik. In: Esser, J.; Hübler, A. [Hrsg.]: Forms and Functions. Tübingen, 25-32
- 1983 Das Piotrovski-Gesetz und seine Verallgemeinerungen. In: Best, K.-H.; Kohlhase, J. [Hrsg.]: Exakte Sprachwandelforschung. Göttingen, 59-90
- 1984a Vortrag auf der Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft in Bielefeld 1984
- 1984b Semantische Diversifikation. In: Folia Linguistica (erscheint)

Altmann, G.; von Buttlar, H.; Rott, W.; Strauss, U.

- 1983 A Law of Language Change. In: Brainerd, B. [Hrsg.]: Historical Linguistics. Bochum, 104-115

Altmann, G.; Beóthy, E.; Best, K.-H.

- 1982 Die Bedeutungskomplexität der Wörter und das Menzerathsche Gesetz. In: Zs. für Phonetik, Sprachwissenschaft und Kommunikationsforschung 35, 537-543

Altmann, G.; Lehfeldt, W.

- 1980 Einführung in die Quantitative Phonologie. Bochum

Altmann, G.; Schwibbe, M.

- 1989 Das Menzerathsche Gesetz in informationsverarbeitenden Systemen. Hildesheim. Zürich. New York

Arapov, M.V.; Cherc, M.M.

- 1983 Mathematische Methoden in der historischen Linguistik. Bochum

Beóthy, E.; Altmann, G.

- 1985 Semantic Diversification of Hungarian Verbal Prefixes III. "föl-", "el-", "be-". In: Rothe, U. [Hrsg.]: Glottometrika 7, Bochum, 45-56

Best, K.-H.

- 1983 Zum morphologischen Wandel einiger deutscher Verben. In: Best, Kohlhase 1983a, 107-118

Best, K.-H.; Kohlhase, J. [Hrsg.]

- 1983a Exakte Sprachwandelforschung. Göttingen

Best, K.-H.; Kohlhase, J.

- 1983b Der Wandel von ward zu wurde. In: Best, Kohlhase 1983a, 91-102

Boroda, M.G.

- 1982 Die melodische Elementareinheit. In : Orlov, Ju.K.; Boroda, M. G.; Nadarejšvili, I.Š.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 205-221
- 1982b Häufigkeitsstrukturen musikalischer Texte. In : Orlov, Ju.K.; Boroda, M. G.; Nadarejšvili, I.Š.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 231-262

1982c Zur Bestimmung einer phrasenähnlichen melodischen Informations-
einheit in der Musik. In: Orlov, Ju.K.; Boroda, M. G.; Nadarejšvili, I.Š.:
Sprache, Text, Kunst. Quantitative Analysen. Bochum, 222-232

Bunge, M.

1967 Scientific Research I,II. Berlin. Heidelberg. New York

Burghes, D.N.; Graham, A.

1980 Introduction to Control Theory, Including Optimal Control. Chichester

Carroll, J.B.

1967 On Sampling from a Lognormal Model of Word-Frequency Distribution.
In: Kučera, H.; Francis, W.N.: Computational Analysis of Present-Day
American English. Providence, Rhode Island, 406-424

1968 Word-Frequency Studies and the Lognormal Distribution. In: Zale, E.M.
[Hrsg.]: Proceedings of the Conference on Language and Language
Behavior. New York, 213-235

1969 A rationale for an asymptotic lognormal form of word-frequency
distribution. Princeton

1970/72 An Alternative to Juilland's Usage Coefficient for Lexical
Frequencies, and a Proposal for a Standard Frequency Index (SFI). In:
Computer Studies in the Humanities and Verbal Behavior 3, 61-65

Chomsky, N.

1957 Syntactic Structures. The Hague

Condon, E.V.

1923 Statistics of vocabulary. In: Science 67, 300

Cortés, F.; Przeworski, A.; Sprague, J.

1974 Systems Analysis for Social Scientists. New York. London. Sydney.
Toronto

Dym, C.L.; Ivey, E.S.

1980 Principles of mathematical modelling. London

Eigen, M.

1971 Selforganization of Matter and the Evolution of Biological Macro-
molecules. In: Die Naturwissenschaften 58, 465-523

Esser, H.; Klenovits, K.; Zehnpfennig, H.

1977 Wissenschaftstheorie 2, Stuttgart

Estoup, J.B.

1916 Gammes Stenographiques. Paris

Fickermann, I.; Markner-Jäger, B.; Rothe, U.

1984 Wortlänge und Bedeutungskomplexität. In: Boy, J.; Köhler, R. [Hrsg.]:
Glottometrika 6. Bochum, 115-126

Fónagy, I.; Magdics, K.

1960 Speed of utterance in phrases of different lengths. In: Language and
Speech 3, 179-192

Fucks, W.

1955a Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen.
Köln. Opladen

1955b Theorie der Wortbildung. In: Mathematisch-physikalische
Semesterberichte 4, 195-212

Galtung, J.

1967 Theory and Methods of Social Research. Oslo

Gerlach, R.

1982 Zur Überprüfung des Menzerathschen Gesetzes im Bereich der
Morphologie. In: Lehfeldt, W., Strauß, U. [Hrsg.]: Glottometrika 4.
Bochum, 95-102

Geršić, S., Altmann, G.

1980 Laut - Silbe - Wort und das Menzerathsche Gesetz. In: Forum
Phoneticum 21. Frankfurt/M., 115-123

Gigerenzer, G.

1981 Messung und Modellbildung in der Psychologie. München. Basel

Grotjahn, R.

1982 Ein statistisches Modell für die Verteilung der Wortlänge. In: Zeitschrift
für Sprachwissenschaft 1, 44-75

Güter, H.

1974 Les relations fréquence - longueur - sens des mots (langues romanes et
anglais). XIV Congresso Internazionale di linguistica e filologia
romanza. Napoli 1974, 15-20

Haight, F.A.

1969 Two probability distributions connected with Zipf's rank-size conjecture. In: *Applicationes Mathematicae* 10, 225-228

Haken, H.

1978 *Synergetics*. Berlin. Heidelberg. New York

Haken, H. [Hrsg.]

1981 *Chaos and order in nature*. Berlin

Haken, H.; Wagner, M. [Hrsg.]

1973 *Cooperative phenomena*. Berlin

Hempel, C.G.

1965a The logic of functional analysis. In: Hempel 1965b, 297-330

Hempel, C.G. [Hrsg.]

1965b *Aspects of scientific explanation*. New York

Hempel, C.G.; Oppenheim, P.

1948 *Studies in the Logic of Explanation*. In: *Philosophy of Science* 15, 135-175

Herdan, G.

1958 The Relation between the Dictionary Distribution and the Occurrence Distribution of Word length and its importance for the study of Quantitative Linguistics. In: *Biometrika* 45, 222-228

1966 *The advanced theory of language as choice and chance*. Berlin. Heidelberg. New York

Heups, G.

1983 Untersuchungen zum Verhältnis von Satzlänge zu Clauselänge am Beispiel deutscher Texte verschiedener Textklassen. In: Köhler, R., Boy, J. [Hrsg.]: *Glottometrika* 5. Bochum, 113-133

Hirsch, M.W.; Smale, S.

1974 *Differential equations, dynamical systems, and linear algebra*. London

Imsiepen, U.

1983 Die e-Epithese bei starken Verben im Deutschen. In: Best, Kohlhasse 1983a, 119-142

Job, U.; Altmann, G.

1985 Ein Modell für anstrengungsbedingte Lautveränderung. In: *Folia*

Linguistica Historica, 1985 (erscheint)

Kanngießer, S.

1977 Skizze des linguistischen Funktionalismus. In: Osnabrücker Beiträge zur Sprachtheorie 3, 132-187

Kohlhase, J.

1983 Die Entwicklung von ward zu wurde beim Nürnberger Chronisten Heinrich Deichsler. In: Best, Kohlhase 1983a, 103-106

Köhler, R.

1982 Das Menzerathsche Gesetz auf Satzebene. In: Lehfeldt, W., Strauß, U. [Hrsg.]: Glottometrika 4. Bochum, 103-113

Köhler, R.; Altmann, G.

1983 Systemtheorie und Semiotik. In: Zeitschrift für Semiotik 5/4, 424-431

Krylov, Ju.K.

1982a Eine Untersuchung statistischer Gesetzmäßigkeiten auf der paradigmatischen Ebene der Lexik natürlicher Sprachen. In: Guiter, H.; Arapov, M.V. [Hrsg.]: Studies on Zipf's Law. Bochum, 234-262

Крылов, Ю.К.

1982b Об одной парадигме лингвостатистических распределений. В сб. Лингвостатистика и вычислительная лингвистика / под ред. Я. Соонтак. Тарту, 80-102

Kuhn, T.S.

1967 Die Struktur wissenschaftlicher Revolutionen. Frankfurt/M.

Kulla, B.

1979 Angewandte Systemwissenschaft. Würzburg. Wien

Lakatos, I.

1974 Falsifikation und Methodologie wissenschaftlicher Forschungsprogramme. In: Lakatos, I.; Musgrave, A. [Hrsg.]: Kritik und Erkenntnisfortschritt. Braunschweig, 88-189

1975 Kritischer Rationalismus und die Methodologie wissenschaftlicher Forschungsprogramme. In: Weingart [Hrsg.]: Wissenschaftsforschung. Frankfurt, 91-132

Lamb, S.M.

1966 Outline of Stratificational Grammar. Washington

Lansky, P., Radil-Weiss, T.

1980 A Generalization of the Yule-Simon Model, With Special Reference to Word Association Tests and Neural Cell Assembly Formation. In: Journal of Mathematical Psychology 21, 53-65

Lehfeldt, W.

1980 Zur numerischen Erfassung der Schwierigkeit des Sprechbewegungsablaufs. In: Grotjahn, R. [Hrsg.]: Glottometrika 2. Bochum, 44-61

Levelt, W.J.M.

1974 Formal Grammars in Linguistics and Psycholinguistics I-III. The Hague. Paris

Lindblom, B.

1983 Economy of Speech Gestures. In: MacNeilage, P. [Hrsg.]: Speech Production. New York

Lockwood, D.G.

1972 Introduction to Stratificational Linguistics. New York

Lüdtke, H.

1980 Auf dem Weg zu einer Theorie des Sprachwandels. In: Lüdtke, H. [Hrsg.]: Kommunikationstheoretische Grundlagen des Sprachwandels. Berlin. New York, 182-252

Mandelbrot, B.

1953 An informational theory of statistical structure of language. In: Jackson, W. [Hrsg.]: Communication Theory. London, 486-502

1961 On the Theory of Word Frequencies and on Related Markovian Models of Discourse. In: Proceedings of Symposia in Applied Mathematics 12, 190-219

1965 Information Theory and Psycholinguistics. In: Wolman, B.B., Nagel, E. [Hrsg.]: Scientific Psychology. New York. London, 550-562

Menzerath, P.

1954 Die Architektonik des deutschen Wortschatzes. Bonn

Moulton, W.G.

1961 Lautwandel durch innere Kausalität: die ostschweizerische Vokalspaltung. In: Zeitschrift für Mundartforschung 28, 227-251

Müller-Hasemann, W.

1983 Das Eindringen englischer Wörter ins Deutsche ab 1945. In: Best, Kohlhasse 1983a, 143-160

Orth, B.

1974 Einführung in die Theorie des Messens. Stuttgart. Berlin. Köln. Mainz

Орлов Ю.К.

1970a Обобщение закона Ципфа-Мандельброта. В сб. Сообщения академии наук Грузинской ССР 57/1, 37-40

1970b О статистической структуре сообщений, оптимальных для человеческого восприятия. В сб. Научно-техническая информация, серия 2,8, 11-15

Orlov, Ju.K.

1982a Dynamik der Häufigkeitsstrukturen. In : Orlov, Ju.K.; Boroda, M.G.; Nadarejšvili, I.S.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 82-117

1982b Ein Modell der Häufigkeitsstruktur des Vokabulars. In: Orlov, Ju.K.; Boroda, M.G.; Nadarejšvili, I.Š.: Sprache, Text, Kunst. Quantitative Analysen. Bochum, 118-192

1982c Linguostatistik: Aufstellung von Sprachnormen oder Analyse des Redeprozesses? (Die Antinomie "Sprache – Rede" in der statistischen Linguistik). In: Orlov, Ju.K.; Boroda, M.G.; Nadarejšvili, I.Š.: Sprache, Text, Kunst. Bochum, 1-55

Orlov, Ju.K.; Boroda, M.G.; Nadarejšvili, I.Š.

1982 Sprache, Text, Kunst. Quantitative Analysen. Bochum

Osgood, C.E.

1963 On Understanding and Creating Sentences. In: Amer. Psychol. 18, 735-751

Paulre, E. [Hrsg.]

1980 System Dynamics and the Analysis of Change. Proceedings of the 6th International Conference on System Dynamics, University of Paris-Dauphine, November 1980. Amsterdam. New York. Oxford

Pieper, U.

1979 Über die Aussagekraft statistischer Methoden für die linguistische Stilanalyse. Tübingen

Пиотровский, Р.Г., Бектаев, К.Б.; Пиотровская, А.А.

1977 Математическая лингвистика. Москва

Пиотровская, А.А.; Пиотровский, Р.Г.

- 1974 Математические модели диахронии и текстообразования. В сб. Статистика речи и автоматический анализ текста / под ред. Пиотровского, Р.Г. / Ленинград, 361-400

Popper, K.R.

- 1972 Conjectures and Refutations. The Growth of Scientific Knowledge. London
- 1973 Objektive Erkenntnis. Hamburg

Prigogine, I.

- 1979 Vom Sein zum Werden. München

Rapoport, A.

- 1957 The Stochastic and the "Teleological" Rationales of Certain Distributions and the So-Called Principle of Least Effort. In: Behavioural Science, 147-161
- 1982 Zipf's Law Re-visited. In: Guiter, H.; Arapov, M.V. [Hrsg.]: Studies on Zipf's Law. Bochum, 1-28

Schaeder, B.

- 1976 Maschinenlesbare Textkorpora des Deutschen und des Englischen. In: Deutsche Sprache 4, 356-369

Scheuch, E.K., Zehnpfennig, H.

- 1974(3) Skalierungsverfahren in der Sozialforschung. In: Koenig, R. [Hrsg.]: Handbuch der empirischen Sozialforschung, Bd. 3a, 97-186. Stuttgart

Schreyer, R.

- 1977 Stratifikationsgrammatik. Tübingen

Schwibbe, M.

- 1984 Text- und wortstatistische Untersuchungen zur Validität der Menzerathschen Regel (MR). In: Boy, J., Köhler, R. [Hrsg.]: Glottometrika 6. Bochum, 152-176

Simon, H.A.

- 1955 On a Class of Skew Distribution Functions. In: Biometrika 42, 425-440
- 1960 Some Further Notes on a Class of Skew Distribution Functions. In: Information and Control 3, 80-88

Spinner, H.F.

- 1974 Pluralismus als Erkenntnismodell. Frankfurt/M. 1974

Stegmüller, W.

1969 Wissenschaftliche Erklärung und Begründung. Berlin

1970 Theorie und Erfahrung. Berlin. Heidelberg. New York

Steinthal, H.

1881 Abriss der Sprachwissenschaft. Berlin

Stoer, J.

1979 Einführung in die Numerische Mathematik I. Berlin. Heidelberg. New York

Strauß, U.

1980 Struktur und Leistung der Vokalsysteme. Bochum

Suppes, P.

1970 Probabilistic grammars for natural languages. In: Synthese 22, 95-116

Teupenhayn, R., Altmann, G.

1984 Clause length and Menzerath's law. In: Boy, J., Köhler, R. [Hrsg.]:
Glottometrika 6. Bochum, 127-138

Vološin, B.A.; Orlov, Ju.K.

1982 Das verallgemeinerte Zipf-Mandelbrot'sche Gesetz und die Verteilung
der Anteile von Farbflächen in der Malerei. In: Orlov, Ju.K.; Boroda,
M.G.; Nadarejšvili, I.Š.: Sprache, Text, Kunst. Bochum, 263-270

Wahrig, G.

1981 Deutsches Wörterbuch. Gütersloh

Walker, H.M., Lev, J.

1953 Statistical Inference. New York

Watt, W.C.

1983 Grade der Systemhaftigkeit. Zur Homogenität der Alphabetschrift. In:
Zs. für Semiotik 5/4, 371-399

Weber, H.J.

1974 Mehrdeutige Wortformen im heutigen Deutsch. Tübingen

Willee, G.

1979 Lemma - ein Programmsystem zur automatischen Lemmatisierung deutscher Wortformen.
In: Sprache und Datenverarbeitung 1-2, 45-60

Yule, G.U.

1924 A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. In:
Philosophical Transactions, B, 213, 21

Zipf, G.K.

1949 Human Behaviour and the Principle of Least Effort. Reading, Mass.

1968(2) The psycho-biology of language. An introduction to dynamic philology. Cambridge,
Mass.

Таблицы

Таблица А1: Полилексия (PL) в зависимости от длины (L)

х	У_{эксп}	У_{теор}
3	5.0000	5.0485
4	4.6316	3.9779
5	4.2740	3.3066
6	3.6981	2.8430
7	2.6000	2.5022
8	1.8938	2.2402
9	1.5943	2.0319
10	1.7537	1.8621
11	1.4215	1.7207
12	1.3853	1.6010
13	1.2637	1.4983
14	1.2658	1.4091
15	1.1071	1.3308
16	1.2037	1.2615
17	1.0789	1.1998
18	1.0333	1.1443
19	1.0357	1.0941
20	1.0000	1.0486
21	1.1429	1.0071
22	1.1111	0.9690
23	1.0000	0.9340
24	1.2000	0.9016
25	1.0000	0.8716
26	1.0000	0.8438

Таблица А2: Политекстия (PT) в зависимости от полилексии (PL)

X	У_{эксп}	У_{теор}
1.0	3.1791	3.9158
2.0	10.5417	11.5634
3.0	19.3306	21.8456
4.0	28.2745	34.2652
5.0	69.6389	48.5833
6.0	89.3235	64.6220
7.0	77.3000	82.2491
8.0	112.8462	101.3609
9.0	230.6667	121.8732
10.0	187.6667	143.7157
11.0	80.0000	166.8287
13.0	365.0000	216.6655
14.0	286.6667	243.3037
16.0	193.0000	299.6389
19.0	373.0000	392.3432
21.0	285.0000	458.8559

Таблица А3: Частота (F) в зависимости от политекстии (PT)

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	1.3299	1.1158	41.0	72.3333	67.3411
2.0	2.7801	2.3986	42.0	62.3333	69.1568
3.0	4.4861	3.7530	44.0	102.0000	72.8018
4.0	5.5128	5.1561	45.0	57.0000	74.6308
5.0	6.4474	6.5966	46.0	75.0000	76.4640
6.0	8.6970	8.0677	47.0	67.5000	78.3014
7.0	10.6667	9.5646	48.0	71.0000	80.1429
8.0	10.6875	11.0840	49.0	61.0000	81.9884
9.0	17.0000	12.6233	50.0	58.0000	83.8378
10.0	14.9091	14.1806	51.0	82.2500	85.6910
11.0	15.4286	15.7542	52.0	68.5000	87.5480
12.0	19.6429	17.3428	54.0	81.5000	91.2732
13.0	23.0000	18.9453	55.0	70.0000	93.1412
14.0	20.9000	20.5607	56.0	69.0000	95.0128
15.0	28.2857	22.1681	60.0	76.5000	102.5333
16.0	27.4286	23.8269	62.0	109.0000	106.3134
17.0	22.8333	25.4764	63.0	77.3333	108.2083
18.0	24.6667	27.1360	64.0	133.0000	110.1063
19.0	26.0000	28.8052	66.0	93.0000	113.9115
20.0	28.2500	30.4837	67.0	118.6667	115.8186
21.0	38.7500	32.1709	68.0	115.0000	117.7287
22.0	35.2500	33.8664	69.0	132.0000	119.6417
23.0	27.0000	35.5701	70.0	131.0000	121.5576
24.0	35.0000	37.2841	74.0	159.0000	129.2494
26.0	48.0000	40.7262	76.0	141.5000	133.1117
27.0	43.5000	42.4591	77.0	94.0000	135.0469
28.0	50.2500	44.1987	78.0	120.0000	136.9846
29.0	43.6000	45.9448	81.0	159.0000	142.8133
30.0	71.2500	47.6972	82.0	248.0000	144.7613
31.0	40.5000	49.4556	85.0	124.0000	150.6199
32.0	86.0000	51.2200	87.0	144.0000	154.5376
33.0	63.0000	52.9901	89.0	171.0000	158.4648
34.0	61.5000	54.7659	90.0	123.0000	160.4318
35.0	52.6667	56.5470	94.0	197.0000	168.3224
36.0	62.2500	58.3335	96.0	154.0000	172.2810
37.0	48.0000	60.1251	97.0	192.0000	174.2635
38.0	77.0000	61.9218	100.0	176.5000	180.2238
39.0	55.0000	63.7235	102.0	137.0000	184.2076
40.0	56.0000	65.5299	105.0	167.0000	190.1987
110.0	152.0000	200.2232	199.0	369.0000	385.2839
111.0	172.5000	202.2339	200.0	368.0000	387.4221
116.0	182.0000	212.3153	209.0	353.0000	406.7158
117.0	187.0000	214.3370	215.0	348.0000	419.6266
118.0	145.0000	216.3606	217.0	416.0000	423.9386

122.0	184.0000	224.4726	222.0	390.0000	434.7367
123.0	342.0000	226.5050	257.0	657.0000	511.0062
124.0	234.0000	228.5391	277.0	611.0000	555.0876
129.0	179.0000	238.7350	278.0	483.0000	557.3006
130.0	175.0000	240.7791	285.0	548.0000	572.8145
132.0	201.0000	244.8724	296.0	762.0000	597.2735
133.0	314.0000	246.9214	299.0	667.0000	603.9608
137.0	221.0000	255.1335	300.0	791.0000	606.1914
140.0	200.0000	261.3091	301.0	685.0000	608.4228
144.0	205.0000	269.5645	304.0	778.0000	615.1216
149.0	220.0000	279.9174	319.0	868.0000	648.7180
150.0	203.0000	281.9924	339.0	719.0000	693.7685
154.0	210.0000	290.3066	365.0	1014.0000	752.7473
167.0	358.5000	317.4805	373.0	1090.0000	770.9843
173.0	287.0000	330.0979	447.0	1912.0000	941.5159
175.0	367.0000	334.3139	449.0	1513.0000	946.1682
180.0	277.0000	344.8758	458.0	1781.0000	967.1301
193.0	331.0000	372.4781			

**Таблица А4: Частота (F) в зависимости от политекстии (PT);
уточненная модель**

X	У_{эсп}	У_{теор}	X	У_{эсп}	У_{теор}
1.0	1.3299	1.6231	17.0	22.8333	25.6990
2.0	2.7801	3.1724	18.0	24.6667	27.2073
3.0	4.4861	4.6987	19.0	26.0000	28.7183
5.0	6.4474	7.7181	20.0	28.2500	30.2323
6.0	8.6970	9.2190	21.0	38.7500	31.7495
7.0	10.6667	10.7166	22.0	35.2500	33.2698
8.0	10.6875	12.2124	23.0	27.0000	34.7936
9.0	17.0000	13.7072	24.0	35.0000	36.3208
10.0	14.9091	15.2019	26.0	48.0000	39.3860
11.0	15.4286	16.6970	27.0	43.5000	40.9243
12.0	19.6429	18.1930	28.0	50.2500	42.4663
13.0	23.0000	19.6903	29.0	43.6000	44.0123
14.0	20.9000	21.1893	30.0	71.2500	45.5623
15.0	28.2857	22.6903	31.0	40.5000	47.1163
16.0	27.4286	24.1934	32.0	86.0000	48.6745
33.0	63.0000	50.2369	106.0	173.0000	178.1385
34.0	61.5000	51.8035	110.0	152.0000	186.0451
35.0	52.6667	53.3744	111.0	172.5000	188.0379
36.0	62.2500	54.9497	116.0	182.0000	198.0990
37.0	48.0000	56.5294	117.0	187.0000	200.1310
38.0	77.0000	58.1136	118.0	245.0000	202.1695
39.0	55.0000	59.7022	122.0	184.0000	210.3899
40.0	56.0000	61.2955	123.0	342.0000	212.4617
41.0	72.3333	62.8933	124.0	234.0000	214.5403

42.0	62.3333	64.4958	129.0	179.0000	225.0348
44.0	102.0000	67.7148	130.0	175.0000	227.1543
45.0	57.0000	69.3315	132.0	201.0000	231.4138
46.0	79.0000	70.9529	133.0	314.0000	233.5538
47.0	67.5000	72.5792	137.0	221.0000	242.1840
48.0	71.0000	74.2104	140.0	200.0000	248.7303
49.0	61.0000	75.8464	144.0	205.0000	257.5583
50.0	58.0000	77.4875	149.0	220.0000	268.7551
51.0	82.2500	79.1335	150.0	203.0000	271.0162
52.0	68.5000	80.7845	154.0	210.0000	280.1341
54.0	81.5000	84.1017	167.0	358.5000	310.5926
55.0	70.0000	85.7680	173.0	287.0000	325.0859
56.0	69.0000	87.4393	175.0	367.0000	329.9792
60.0	76.5000	94.1768	180.0	277.0000	342.3510
62.0	109.0000	97.5772	193.0	331.0000	375.4590
63.0	77.3333	99.2853	199.0	369.0000	391.2091
64.0	133.0000	100.9989	200.0	368.0000	393.8634
66.0	93.0000	104.4422	209.0	353.0000	418.1359
67.0	118.6667	106.1720	215.0	348.0000	434.7065
68.0	115.0000	107.9072	217.0	416.0000	440.3003
69.0	132.0000	109.6480	222.0	390.0000	454.4406
70.0	131.0000	111.3942	257.0	657.0000	559.8929
74.0	159.0000	118.4351	277.0	611.0000	625.5237
76.0	141.5000	121.9893	278.0	483.0000	628.9130
77.0	94.0000	123.7749	285.0	548.0000	652.9338
78.0	120.0000	125.5663	296.0	762.0000	691.7418
81.0	159.0000	130.9748	299.0	667.0000	702.5552
82.0	248.0000	132.7893	300.0	791.0000	706.1817
85.0	124.0000	138.2677	301.0	685.0000	709.8194
87.0	144.0000	141.9495	304.0	778.0000	720.7994
89.0	171.0000	145.6551	319.0	868.0000	777.2331
90.0	123.0000	147.5169	339.0	719.0000	856.5908
94.0	197.0000	155.0244	365.0	1014.0000	967.1730
96.0	154.0000	158.8147	373.0	1090.0000	1002.9668
97.0	192.0000	160.7190	447.0	1912.0000	1377.4685
100.0	176.5000	166.4692	449.0	1513.0000	1388.7594
102.0	137.0000	170.3338	458.0	1781.0000	1440.3884
105.0	167.0000	176.1778			

Таблица А5: Длина (L) в зависимости от частоты (F)

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	12.9497	10.9000	43.0	8.0000	7.2613
2.0	12.4107	10.1138	44.0	11.5000	7.2433
3.0	11.9375	9.6805	46.0	8.5000	7.2086
4.0	10.0893	9.3843	47.0	5.0000	7.1918
5.0	9.8140	9.1609	48.0	7.7333	7.1755
6.0	10.3750	8.9823	49.0	4.0000	7.1595
7.0	8.8095	8.8340	51.0	5.0000	7.1287
8.0	9.4444	8.7075	52.0	5.0000	7.1137
9.0	9.3077	8.5974	53.0	7.0000	7.0991
10.0	6.6667	8.5001	54.0	9.0000	7.0848
11.0	7.5625	8.4131	55.0	5.5000	7.0708
12.0	7.0000	8.3344	56.0	4.0000	7.0570
13.0	9.2222	8.2627	57.0	4.0000	7.0436
14.0	6.1000	8.1968	58.0	7.5000	7.0303
15.0	9.3333	8.1360	59.0	5.5000	7.0174
16.0	8.2857	8.0794	61.0	6.0000	6.9922
17.0	8.2000	8.0267	62.0	8.6000	6.9799
18.0	8.0000	7.9773	64.0	6.6667	6.9560
19.0	8.1429	7.9309	65.0	8.5000	6.9444
20.0	8.5000	7.8871	66.0	8.0000	6.9329
21.0	8.0000	7.8456	68.0	7.0000	6.9106
22.0	7.0000	7.8063	69.0	4.0000	6.8997
23.0	7.1667	7.7689	70.0	11.0000	6.8890
24.0	7.7500	7.7333	71.0	6.0000	6.8785
25.0	6.0000	7.6993	72.0	8.0000	6.8681
26.0	7.6000	7.6667	73.0	6.5000	6.8578
27.0	8.3333	7.6355	74.0	7.0000	6.8478
28.0	7.3333	7.6056	75.0	10.0000	6.8379
29.0	6.0000	7.5768	77.0	5.0000	6.8184
30.0	8.6667	7.5491	79.0	8.6667	6.7996
31.0	9.0000	7.5225	80.0	11.0000	6.7904
32.0	6.3333	7.4967	82.0	5.0000	6.7723
33.0	8.6667	7.4718	86.0	10.0000	6.7375
34.0	7.0000	7.4478	87.0	11.0000	6.7291
35.0	7.3333	7.4245	89.0	8.0000	6.7126
36.0	6.5000	7.4019	91.0	10.0000	6.6965
37.0	5.5000	7.3801	93.0	9.0000	6.6808
38.0	7.2500	7.3588	94.0	6.0000	6.6731
39.0	4.5000	7.3382	98.0	7.0000	6.6432
40.0	7.0000	7.3182	99.0	5.0000	6.6359
41.0	7.5000	7.2987	101.0	6.0000	6.6216
42.0	9.0000	7.2797	104.0	6.0000	6.6007
106.0	8.0000	6.5871	219.0	8.0000	6.0906
109.0	5.0000	6.5673	220.0	6.0000	6.0876
113.0	6.0000	6.5418	221.0	4.0000	6.0846
115.0	7.0000	6.5294	225.0	8.0000	6.0728

118.0	7.0000	6.5112	234.0	6.0000	6.0472
120.0	6.0000	6.4994	248.0	3.0000	6.0093
123.0	4.0000	6.4821	265.0	6.0000	5.9664
124.0	5.0000	6.4764	270.0	10.0000	5.9544
125.0	6.5000	6.4708	277.0	5.0000	5.9380
130.0	6.0000	6.4435	287.0	8.0000	5.9153
131.0	9.0000	6.4382	290.0	7.0000	5.9086
132.0	4.5000	6.4329	314.0	12.0000	5.8581
133.0	6.0000	6.4276	331.0	6.0000	5.8249
134.0	7.0000	6.4224	342.0	10.0000	5.8043
137.0	5.0000	6.4071	348.0	5.0000	5.7934
141.0	7.0000	6.3872	353.0	8.0000	5.7845
142.0	10.0000	6.3823	367.0	7.0000	5.7603
144.0	5.0000	6.3727	368.0	5.0000	5.7586
152.0	7.0000	6.3356	369.0	7.0000	5.7569
155.0	7.0000	6.3222	390.0	6.0000	5.7226
159.0	10.5000	6.3049	416.0	6.0000	5.6828
163.0	5.0000	6.2880	427.0	6.0000	5.6668
167.0	5.0000	6.2715	483.0	5.0000	5.5919
168.0	10.0000	6.2675	548.0	6.0000	5.5162
171.0	6.0000	6.2555	611.0	6.0000	5.4517
173.0	6.0000	6.2477	657.0	5.0000	5.4092
175.0	7.0000	6.2399	667.0	6.0000	5.4003
179.0	4.5000	6.2247	685.0	5.0000	5.3848
182.0	7.0000	6.2135	719.0	6.0000	5.3567
184.0	9.0000	6.2062	762.0	5.0000	5.3232
187.0	6.0000	6.1954	778.0	4.0000	5.3113
190.0	4.0000	6.1847	791.0	4.0000	5.3018
192.0	11.0000	6.1777	868.0	3.0000	5.2489
200.0	5.0000	6.1506	1014.0	4.0000	5.1615
201.0	5.0000	6.1473	1090.0	5.0000	5.1213
203.0	8.0000	6.1407	1513.0	6.0000	4.9432
205.0	7.0000	6.1342	1781.0	5.0000	4.8569
210.0	9.0000	6.1182	1912.0	4.0000	4.8198

Таблица А6: Опосредованная зависимость полилексии (PL) от частоты (F)

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	1.2114	1.0300	42.0	3.0000	2.9663
2.0	1.3988	1.2532	43.0	1.5000	2.9861
3.0	1.5250	1.4056	44.0	1.5000	3.0056
4.0	1.5893	1.5248	46.0	2.5000	3.0437
5.0	1.8140	1.6242	47.0	5.5000	3.0623
6.0	1.7168	1.7102	48.0	2.8333	3.0806
7.0	2.0476	1.7865	49.0	3.0000	3.0986
8.0	2.0556	1.8553	51.0	8.0000	3.1339
9.0	2.4615	1.9182	52.0	2.5000	3.1511
10.0	2.8333	1.9762	53.0	2.5000	3.1682
11.0	2.6250	2.0303	54.0	2.0000	3.1850
12.0	2.4286	2.0809	55.0	6.5000	3.2015
13.0	2.0000	2.1285	56.0	5.0000	3.2179
14.0	2.6000	2.1737	57.0	3.0000	3.2341
15.0	1.6667	2.2165	58.0	4.5000	3.2500
16.0	3.0000	2.2574	59.0	3.0000	3.2658
17.0	3.6000	2.2964	61.0	4.0000	3.2967
18.0	2.5000	2.3339	62.0	2.8000	3.3120
19.0	3.0000	2.3699	64.0	2.0000	3.3418
20.0	2.5000	2.4045	65.0	5.0000	3.3565
21.0	3.0000	2.4380	66.0	1.0000	3.3711
22.0	1.5000	2.4703	68.0	4.0000	3.3997
23.0	3.6667	2.5015	69.0	8.0000	3.4137
24.0	3.0000	2.5318	70.0	1.0000	3.4277
25.0	4.5000	2.5613	71.0	5.0000	3.4415
26.0	2.4000	2.5899	72.0	4.0000	3.4551
27.0	3.6667	2.6177	73.0	3.0000	3.4686
28.0	2.0000	2.6447	74.0	2.0000	3.4820
29.0	2.5000	2.6711	75.0	2.0000	3.4953
30.0	2.3333	2.6969	77.0	4.5000	3.5214
31.0	3.3333	2.7220	79.0	3.3333	3.5470
32.0	4.3333	2.7466	80.0	3.0000	3.5597
33.0	2.0000	2.7706	82.0	4.0000	3.5846
34.0	3.0000	2.7941	86.0	3.0000	3.6333
35.0	3.0000	2.8171	87.0	5.0000	3.6452
36.0	2.2500	2.8397	89.0	1.0000	3.6687
37.0	3.0000	2.8618	91.0	2.0000	3.6919
38.0	4.7500	2.8835	93.0	2.5000	3.7146
39.0	3.5000	2.9047	94.0	2.0000	3.7259
40.0	2.5000	2.9256	98.0	5.0000	3.7701
41.0	2.5000	2.9462	99.0	2.0000	3.7810
101.0	6.0000	3.8024	210.0	5.0000	4.6776
104.0	5.0000	3.8340	219.0	8.0000	4.7335
106.0	1.0000	3.8548	220.0	6.0000	4.7396
109.0	4.0000	3.8853	221.0	8.0000	4.7457

113.0	3.0000	3.9252	225.0	5.0000	4.7699
115.0	6.0000	3.9447	234.0	1.0000	4.8231
118.0	6.0000	3.9736	248.0	4.0000	4.9031
120.0	1.0000	3.9925	265.0	6.0000	4.9959
123.0	8.0000	4.0205	270.0	3.0000	5.0224
124.0	4.5000	4.0297	277.0	14.0000	5.0589
125.0	4.5000	4.0389	287.0	6.0000	5.1100
130.0	5.0000	4.0840	290.0	5.0000	5.1250
131.0	2.0000	4.0928	314.0	6.0000	5.2417
132.0	6.5000	4.1017	331.0	16.0000	5.3205
133.0	6.0000	4.1104	342.0	6.0000	5.3699
134.0	10.0000	4.1192	346.0	3.0000	5.3876
137.0	4.0000	4.1450	353.0	5.0000	5.4182
141.0	2.0000	4.1789	367.0	5.0000	5.4782
142.0	1.0000	4.1873	368.0	5.0000	5.4824
144.0	5.0000	4.2039	369.0	4.0000	5.4866
152.0	4.0000	4.2687	390.0	14.0000	5.5733
155.0	3.0000	4.2924	416.0	6.0000	5.6760
159.0	6.0000	4.3235	427.0	6.0000	5.7181
163.0	6.0000	4.3540	483.0	6.0000	5.9210
167.0	8.0000	4.3840	548.0	21.0000	6.1364
168.0	6.0000	4.3914	611.0	6.0000	6.3283
171.0	8.0000	4.4134	657.0	8.0000	6.4596
173.0	2.0000	4.4280	667.0	5.0000	6.4873
175.0	2.0000	4.4424	685.0	7.0000	6.5364
179.0	5.0000	4.4709	719.0	9.0000	6.6266
182.0	1.0000	4.4920	762.0	5.0000	6.7364
184.0	3.0000	4.5059	778.0	6.0000	6.7762
187.0	5.0000	4.5266	791.0	6.0000	6.8080
190.0	7.0000	4.5470	868.0	9.0000	6.9894
192.0	5.0000	4.5605	1014.0	13.0000	7.3038
200.0	3.0000	4.6135	1090.0	19.0000	7.4547
201.0	11.0000	4.6200	1513.0	8.0000	8.1796
203.0	1.0000	4.6330	1781.0	14.0000	8.5659
205.0	2.0000	4.6458	1912.0	10.0000	8.7397

Таблица А7: Опосредованная зависимость политекстии (PT) от длины (L)

X	У_{эксп}	У_{теор}
3.0	54.2222	123.7874
4.0	67.5789	62.2405
5.0	60.0000	36.5139
6.0	43.0000	23.6164
7.0	26.8125	16.3385
8.0	14.9646	11.8744
9.0	8.6415	8.9610
10.0	9.0075	6.9662
11.0	5.6364	5.5471
12.0	4.3303	4.5056
13.0	2.0989	3.7211
14.0	1.5190	3.1171
15.0	1.6786	2.6432
16.0	2.9444	2.2654
17.0	1.1316	1.9599
18.0	1.7333	1.7096
19.0	1.1071	1.5024
20.0	1.1667	1.3290
21.0	1.2143	1.1827
22.0	1.2222	1.0583
23.0	1.1667	0.9516
24.0	1.0000	0.8596
25.0	1.0000	0.7797
26.0	1.0000	0.7099

Таблица А8: Опосредованная зависимость частоты (F) от полилексии (PL)

X	У_{эсп}	У_{теор}
1.0	4.7881	5.5000
2.0	15.5046	18.6284
3.0	30.6048	38.0274
4.0	46.3333	63.0940
5.0	132.0000	93.4437
6.0	184.5294	128.7979
7.0	145.5000	168.9411
8.0	268.7692	213.6983
9.0	547.3333	262.9236
10.0	688.3333	316.4920
11.0	119.5000	374.2949
13.0	1014.0000	502.2303
14.0	816.0000	572.2003
16.0	331.0000	723.7921
19.0	1090.0000	979.4201
21.0	548.0000	1168.0696

Таблица А9: Опосредованная зависимость длины (L) от политектсии (PT)

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	12.8321	11.5000	42.0	7.8000	7.5665
2.0	11.5065	10.6410	43.0	7.7500	7.5466
3.0	10.5072	10.1686	44.0	8.7000	7.5272
4.0	9.6329	9.8462	45.0	6.1667	7.5082
5.0	9.7745	9.6031	46.0	8.2000	7.4898
6.0	9.1667	9.4090	47.0	7.1667	7.4718
7.0	8.8561	9.2480	48.0	6.6667	7.4542
8.0	9.2165	9.1107	49.0	8.7143	7.4370
9.0	8.8913	8.9913	50.0	7.7500	7.4202
10.0	8.1688	8.8858	51.0	7.1579	7.4037
11.0	8.5593	8.7915	52.0	7.1382	7.3876
12.0	8.9111	8.7062	53.0	7.1188	7.3719
13.0	7.4000	8.6285	54.0	7.0997	7.3565
14.0	7.5556	8.5572	55.0	7.0809	7.3414
15.0	8.6216	8.4913	56.0	7.0625	7.3266
16.0	7.8519	8.4302	57.0	7.0443	7.3121
17.0	9.2800	8.3731	58.0	7.0265	7.2978
18.0	8.3333	8.3197	59.0	7.0089	7.2839
19.0	8.1364	8.2695	60.0	6.9916	7.2702
20.0	7.1765	8.2221	61.0	6.9745	7.2567
21.0	8.5625	8.1773	62.0	6.9577	7.2435
22.0	8.3846	8.1348	63.0	6.9411	7.2306
23.0	8.2000	8.0944	64.0	6.9248	7.2178
24.0	7.1875	8.0559	65.0	6.9087	7.2053
25.0	8.1333	8.0192	66.0	6.8928	7.1930
26.0	7.7000	7.9840	67.0	6.8772	7.1809
27.0	8.7059	7.9503	68.0	6.8617	7.1690
28.0	6.5455	7.9180	69.0	6.8465	7.1573
29.0	8.0000	7.8870	70.0	6.8314	7.1457
30.0	6.0000	7.8571	72.0	6.8018	7.1232
31.0	8.6364	7.8283	73.0	6.7873	7.1122
32.0	8.7143	7.8005	74.0	6.7730	7.1014
33.0	7.0000	7.7736	75.0	6.7588	7.0907
34.0	6.6250	7.7477	76.0	6.7448	7.0802
35.0	6.2857	7.7226	77.0	6.7309	7.0699
36.0	8.2308	7.6982	78.0	6.7172	7.0597
37.0	7.5000	7.6747	79.0	6.7037	7.0496
38.0	5.5000	7.6518	81.0	6.6770	7.0299
39.0	8.6000	7.6295	82.0	6.6639	7.0202
40.0	6.9167	7.6079	85.0	6.6254	6.9920
41.0	8.0000	7.5869	87.0	6.6003	6.9738
89.0	6.5758	6.9561	169.0	6.0000	6.4740
90.0	6.5637	6.9474	173.0	8.0000	6.4571
92.0	6.5398	6.9303	174.0	5.0000	6.4529
94.0	6.5163	6.9137	175.0	7.0000	6.4488

95.0	6.5048	6.9055	180.0	5.0000	6.4285
96.0	6.4933	6.8974	182.0	3.0000	6.4205
97.0	6.4319	6.8894	185.0	5.0000	6.4088
99.0	6.4595	6.8736	192.0	7.0000	6.3822
100.0	6.4484	6.8659	193.0	5.5000	6.3785
101.0	6.4374	6.8583	194.0	3.5000	6.3748
102.0	6.4265	6.8507	197.0	4.0000	6.3638
105.0	6.3943	6.8285	199.0	7.0000	6.3566
106.0	6.3837	6.8212	200.0	4.5000	6.3531
107.0	6.3732	6.8141	209.0	8.0000	6.3218
108.0	6.3628	6.8070	211.0	6.0000	6.3151
109.0	6.3525	6.8000	215.0	5.0000	6.3018
110.0	6.3422	6.7930	217.0	6.5000	6.2953
111.0	6.3321	6.7861	219.0	5.0000	6.2888
115.0	7.5000	6.7593	220.0	7.0000	6.2856
116.0	6.5000	6.7527	222.0	6.0000	6.2792
117.0	6.0000	6.7462	227.0	6.0000	6.2636
118.0	6.6667	6.7398	232.0	4.0000	6.2483
121.0	6.0000	6.7209	238.0	9.0000	6.2305
122.0	8.6667	6.7147	243.0	6.0000	6.2160
123.0	10.0000	6.7085	257.0	5.0000	6.1771
124.0	6.0000	6.7025	267.0	4.0000	6.1508
126.0	5.0000	6.6905	272.0	6.0000	6.1380
128.0	5.0000	6.6787	277.0	6.0000	6.1255
130.0	7.3333	6.6671	278.0	5.0000	6.1230
132.0	5.0000	6.6557	285.0	6.0000	6.1060
133.0	12.0000	6.6501	287.0	5.0000	6.1012
134.0	4.0000	6.6445	295.0	4.0000	6.0824
137.0	4.0000	6.6280	296.0	5.0000	6.0801
140.0	5.0000	6.6120	299.0	6.0000	6.0733
141.0	9.0000	6.6067	300.0	5.5000	6.0710
142.0	8.0000	6.6015	301.0	5.0000	6.0687
144.0	5.0000	6.5912	304.0	4.0000	6.0620
147.0	7.0000	6.5759	308.0	5.0000	6.0531
149.0	6.6667	6.5660	319.0	3.0000	6.0294
150.0	6.5000	6.5611	323.0	4.0000	6.0210
152.0	12.0000	6.5514	333.0	5.0000	6.0005
154.0	9.0000	6.5418	334.0	6.0000	5.9984
157.0	6.0000	6.5277	339.0	6.0000	5.9885
161.0	6.0000	6.5093	365.0	4.0000	5.9391
162.0	5.5000	6.5048	373.0	5.0000	5.9247
167.0	6.3333	6.4827	393.0	4.0000	5.8901
402.0	3.5000	5.8752	490.0	3.0000	5.7464
416.0	4.0000	5.8527	493.0	5.0000	5.7425
447.0	4.0000	5.8058	489.0	6.0000	5.7477
449.0	6.0000	5.8029	500.0	4.0000	5.7334
458.0	5.0000	5.7900			

Таблица А10: Дважды опосредованная зависимость полилексии (L) от политекстии (PT)

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	1.2180	1.1100	31.0	4.0000	3.2184
2.0	1.4326	1.3761	32.0	3.0000	3.2503
3.0	1.6667	1.5604	33.0	2.0000	3.2814
4.0	2.0256	1.7059	34.0	6.5000	3.3119
5.0	2.1842	1.8281	35.0	4.3333	3.3418
6.0	2.1818	1.9344	36.0	3.2500	3.3711
7.0	2.6667	2.0291	37.0	3.0000	3.3999
8.0	1.8750	2.1149	38.0	8.0000	3.4281
9.0	2.6667	2.1935	39.0	4.0000	3.4558
10.0	3.0000	2.2663	40.0	5.0000	3.4831
11.0	2.5714	2.3343	41.0	2.0000	3.5098
12.0	2.3571	2.3981	42.0	4.0000	3.5361
13.0	3.8571	2.4584	44.0	3.5000	3.5875
14.0	2.9000	2.5155	45.0	2.5000	3.6126
15.0	2.7143	2.5699	46.0	3.5000	3.6373
16.0	3.8571	2.6218	47.0	3.5000	3.6616
17.0	3.0000	2.6715	48.0	4.0000	3.6856
18.0	3.1111	2.7193	49.0	5.0000	3.7092
19.0	4.3333	2.7653	50.0	4.0000	3.7325
20.0	2.0000	2.8096	51.0	2.0000	3.7555
21.0	2.0000	2.8524	52.0	3.0000	3.7782
22.0	2.5000	2.8938	54.0	3.0000	3.8227
23.0	2.3333	2.9340	55.0	2.0000	3.8445
24.0	3.5556	2.9730	56.0	8.0000	3.8660
26.0	1.3333	3.0476	60.0	1.5000	3.9496
27.0	3.5000	3.0835	62.0	4.0000	3.9899
28.0	4.5000	3.1185	63.0	2.3333	4.0098
29.0	2.8000	3.1526	64.0	6.0000	4.0294
30.0	2.2500	3.1859	66.0	3.5000	4.0680
67.0	6.0000	4.0870	140.0	3.0000	5.1360
68.0	5.5000	4.1058	144.0	2.0000	5.1810
69.0	7.0000	4.1245	149.0	6.0000	5.2361
70.0	2.0000	4.1429	150.0	1.0000	5.2470
74.0	5.5000	4.2149	154.0	5.0000	5.2900
76.0	5.0000	4.2499	167.0	5.5000	5.4246
77.0	2.0000	4.2671	173.0	6.0000	5.4843
78.0	1.0000	4.2842	175.0	5.0000	5.5038
81.0	8.0000	4.3346	180.0	14.0000	5.5521

82.0	4.0000	4.3512	193.0	16.0000	5.6734
85.0	3.6667	4.3999	199.0	4.0000	5.7275
90.0	8.0000	4.4786	200.0	5.0000	5.7364
94.0	3.0000	4.5393	209.0	5.0000	5.8152
96.0	5.0000	4.5691	215.0	3.0000	5.8665
97.0	5.0000	4.5838	217.0	6.0000	5.8834
100.0	9.0000	4.6273	222.0	14.0000	5.9251
102.0	4.0000	4.6557	257.0	8.0000	6.2002
105.0	8.0000	4.6978	277.0	6.0000	6.3459
106.0	2.0000	4.7116	276.0	6.0000	6.3388
110.0	4.0000	4.7660	285.0	21.0000	6.4021
111.0	4.0000	4.7794	296.0	5.0000	6.4777
116.0	1.0000	4.8451	299.0	5.0000	6.4980
117.0	5.0000	4.8580	300.0	6.0000	6.5047
118.0	5.5000	4.8709	301.0	7.0000	6.5115
122.0	3.0000	4.9215	304.0	6.0000	6.5315
123.0	6.0000	4.9339	319.0	9.0000	6.6298
124.0	1.0000	4.9463	339.0	9.0000	6.7559
125.0	7.0000	4.9587	365.0	13.0000	6.9125
130.0	2.0000	5.0193	373.0	19.0000	6.9591
132.0	11.0000	5.0431	447.0	10.0000	7.3607
133.0	6.0000	5.0550	449.0	8.0000	7.3709
137.0	8.0000	5.1016	458.0	14.0000	7.4164

**Таблица А11: Зависимость частоты (F) от политекстии (PT):
словоформы в совокупности**

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	1.3865	1.7727	42.0	61.6667	85.3195
2.0	2.7702	3.6002	43.0	75.3333	87.5232
3.0	4.4235	5.4524	44.0	78.6000	89.7343
4.0	6.0548	7.3229	45.0	82.0000	91.9528
5.0	8.0157	9.2083	46.0	59.0000	94.1787
6.0	8.6530	11.1070	47.0	103.6000	96.4119
7.0	10.9950	13.0177	48.0	110.2500	98.6526
8.0	12.8392	14.9395	49.0	208.8333	100.9006
9.0	14.5639	16.8719	50.0	93.5000	103.1561
10.0	15.6514	18.8143	51.0	79.6250	105.4189
11.0	18.7267	20.7663	52.0	93.1250	107.6891
12.0	23.0149	22.7277	53.0	94.6667	109.9667
13.0	22.9383	24.6981	54.0	98.4000	112.2517
14.0	20.9333	26.6774	55.0	2372.6000	114.5441
15.0	23.3043	28.6654	57.0	89.4286	119.1510
16.0	26.1522	30.6619	58.0	182.0000	121.4656
17.0	35.1316	32.6667	59.0	124.8750	123.7877
18.0	30.0222	34.6798	60.0	85.8000	126.1171
19.0	32.9048	36.7011	61.0	109.0000	128.4540
20.0	48.2143	38.7305	62.0	76.0000	130.7983
21.0	35.0000	40.7678	63.0	244.0000	133.1501
22.0	39.7222	42.8130	64.0	126.2500	135.5093
23.0	498.9048	44.8662	65.0	108.1667	137.8760
24.0	41.3846	46.9271	66.0	110.8000	140.2501
25.0	372.4091	48.9957	67.0	81.0000	142.6318
26.0	35.2222	51.0721	68.0	692.5000	145.0208
27.0	48.6250	53.1561	69.0	133.0000	147.4174
28.0	930.2857	55.2478	70.0	102.7500	149.8215
29.0	60.7500	57.3470	71.0	185.5000	152.2331
30.0	49.6500	59.4539	72.0	103.0000	154.6522
31.0	55.5000	61.5683	73.0	116.5000	157.0789
32.0	69.7692	63.6902	74.0	250.0000	159.5131
33.0	52.6667	65.8196	75.0	106.0000	161.9548
34.0	74.6250	67.9565	76.0	550.0000	164.4041
35.0	47.7143	70.1008	77.0	185.5000	166.8610
36.0	56.8235	72.2526	78.0	125.0000	169.3255
37.0	64.5556	74.4119	79.0	157.0000	171.7975
38.0	162.6000	76.5786	80.0	114.0000	174.2771
39.0	68.8000	78.7527	81.0	609.0000	176.7644
40.0	73.0000	80.9342	82.0	133.0000	179.2593
41.0	69.6000	83.1231	83.0	146.0000	181.7617
84.0	346.5000	184.2719	157.0	243.0000	389.1054
85.0	169.2000	186.7898	158.0	303.0000	392.2213
86.0	121.0000	189.3153	159.0	1328.0000	395.3463

87.0	144.0000	191.8485	160.0	276.0000	398.4798
89.0	130.0000	196.9380	161.0	412.0000	401.6222
91.0	137.0000	202.0583	162.0	1073.0000	404.7736
94.0	134.2000	209.7970	165.0	304.0000	414.2808
95.0	131.0000	212.3921	166.0	428.0000	417.4676
96.0	182.0000	214.9950	174.0	20079.5000	443.2863
100.0	137.0000	225.4849	176.0	262.0000	449.8316
101.0	127.0000	228.1270	180.0	278.0000	463.0317
104.0	233.6667	236.1007	182.0	252.0000	469.6869
105.0	216.0000	238.7744	186.0	645.0000	483.1084
106.0	171.0000	241.4561	190.0	336.0000	496.6791
107.0	313.0000	244.1456	192.0	302.0000	503.5207
108.0	168.5000	246.8432	193.0	13375.3300	506.9556
109.0	137.0000	249.5488	194.0	530.0000	510.4000
110.0	231.4000	252.2623	199.0	2817.0000	527.7647
111.0	174.0000	254.9839	202.0	431.0000	538.2981
113.0	1120.5000	260.4511	211.0	311.0000	570.4211
114.0	178.0000	263.1968	215.0	348.0000	584.9522
115.0	166.0000	265.9505	217.0	411.0000	592.2771
116.0	367.5000	268.7123	219.0	320.0000	599.6421
117.0	470.0000	271.4824	225.0	741.0000	621.9763
118.0	222.5000	274.2604	232.0	437.0000	648.4932
119.0	216.0000	277.0465	237.0	423.5000	667.7413
121.0	237.0000	282.6434	264.0	585.0000	776.2291
126.0	243.0000	296.7788	271.0	434.0000	805.6431
128.0	282.0000	302.4907	284.0	786.0000	861.7226
129.0	179.0000	305.3590	295.0	535.0000	910.6827
135.0	203.0000	322.7442	300.0	791.0000	933.4037
136.0	317.0000	325.6710	308.0	905.0000	970.3737
137.0	255.0000	328.6062	323.0	1028.0000	1041.7787
140.0	200.0000	337.4626	333.0	729.0000	1090.9292
144.0	228.0000	349.3901	393.0	1183.0000	1413.4170
147.0	352.0000	358.4255	402.0	1137.0000	1466.0990
148.0	255.0000	361.4545	416.0	1772.0000	1550.4162
154.0	253.0000	379.8100	485.0	3863.0000	2010.8470
156.0	385.0000	385.9983	487.0	6091.0000	2025.3721

Таблица А12: Зависимость частоты (F) от политекстии (PT): леммы в совокупности

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	1.3829	2.1685	42.0	74.1000	60.5744
2.0	2.7768	3.9185	43.0	87.2500	62.0373
3.0	4.2178	5.5481	44.0	76.8000	63.5064
4.0	5.9233	7.1087	45.0	63.3333	64.9819
5.0	7.3018	8.6232	46.0	68.4000	66.4637
6.0	8.8889	10.1045	47.0	76.0000	67.9521
7.0	10.4892	11.5607	48.0	77.3333	69.4472
8.0	11.3402	12.9975	49.0	71.4286	70.9490
9.0	14.5870	14.4190	50.0	77.7500	72.4578
10.0	14.6494	15.8283	51.0	77.0000	73.9735
11.0	15.8814	17.2280	52.0	75.0000	75.4962
12.0	20.1333	18.6199	53.0	80.7500	77.0262
13.0	21.3429	20.0057	54.0	86.7500	78.5635
14.0	19.9333	21.3867	55.0	75.1250	80.1081
15.0	23.6216	22.7641	56.0	72.3333	81.6602
16.0	26.8148	24.1389	57.0	115.3333	83.2198
17.0	24.2400	25.5119	58.0	88.0000	84.7871
18.0	28.1667	26.8838	59.0	99.0000	86.3622
19.0	31.1818	28.2553	60.0	72.8333	87.9451
20.0	27.5294	29.6271	61.0	86.5000	89.5359
21.0	35.0000	30.9996	62.0	92.6667	91.1347
22.0	34.5769	32.3732	63.0	114.0000	92.7415
23.0	34.2000	33.7485	64.0	113.0000	94.3566
24.0	40.2500	35.1259	65.0	222.0000	95.9799
25.0	34.0667	36.5056	66.0	90.8000	97.6115
26.0	43.7000	37.8880	67.0	111.8333	99.2516
27.0	38.2941	39.2734	68.0	115.0000	100.9001
28.0	42.8182	40.6622	69.0	163.0000	102.5572
29.0	41.0000	42.0544	70.0	111.0000	104.2229
30.0	60.3333	43.4506	72.0	106.5000	107.5807
31.0	44.3636	44.8507	73.0	88.0000	109.2729
32.0	47.4286	46.2551	74.0	159.0000	110.9740
33.0	51.9000	47.6640	75.0	112.5000	112.6841
34.0	52.6250	49.0776	76.0	116.5000	114.4033
35.0	52.5714	50.4960	77.0	105.0000	116.1317
36.0	66.3077	51.9194	78.0	120.0000	117.8694
37.0	48.8750	53.3480	79.0	110.0000	119.6164
38.0	68.7500	54.7820	81.0	159.0000	123.1387
39.0	60.4000	56.2215	82.0	166.0000	124.9141
40.0	72.5000	57.6666	85.0	168.6667	130.2983
41.0	68.5000	59.1175	87.0	151.3333	133.9368
89.0	139.7500	137.6153	169.0	355.0000	324.1000
90.0	123.5000	139.4696	173.0	287.0000	335.8046
92.0	138.0000	143.2090	174.0	273.0000	338.7719

94.0	176.0333	146.9897	175.0	367.0000	341.7559
95.0	124.0000	148.8957	180.0	277.0000	356.9286
96.0	174.7500	150.8122	182.0	252.0000	363.1173
97.0	188.5000	152.7394	185.0	289.0000	372.5309
99.0	191.0000	156.6259	192.0	302.5000	395.1179
100.0	177.0000	158.5853	193.0	367.0000	398.4171
101.0	167.0000	160.5556	194.0	304.5000	401.7347
102.0	137.0000	162.5369	197.0	437.0000	411.7988
105.0	167.0000	168.5473	199.0	369.0000	418.6021
106.0	551.0000	170.5732	200.0	524.5000	422.0321
107.0	167.0000	172.6104	209.0	353.0000	453.7733
108.0	151.0000	174.6590	211.0	330.5000	461.0442
109.0	248.0000	176.7191	215.0	348.0000	475.8293
110.0	157.0000	178.7907	217.0	414.0000	483.3451
111.0	168.6667	180.8739	219.0	320.0000	490.9446
115.0	173.0000	189.3242	220.0	1024.0000	494.7757
116.0	160.5000	191.4665	222.0	390.0000	502.5016
117.0	187.0000	193.6209	227.0	542.0000	522.1923
118.0	226.0000	195.7873	232.0	437.0000	542.4313
121.0	164.0000	202.3601	238.0	439.0000	567.4619
122.0	207.0000	204.5757	243.0	422.0000	588.9577
123.0	342.0000	206.8038	257.0	657.0000	652.3588
124.0	220.6667	209.0445	267.0	593.0000	700.6897
126.0	236.0000	213.5637	272.0	435.0000	725.8536
129.0	179.0000	220.4386	277.0	611.0000	751.7060
130.0	192.6667	222.7561	278.0	483.0000	756.9604
132.0	201.0000	227.4304	285.0	548.0000	794.5445
133.0	314.0000	229.7873	287.0	792.0000	805.5449
134.0	463.0000	232.1576	295.0	535.0000	850.7458
137.0	260.5000	239.3489	296.0	762.0000	856.5333
140.0	200.0000	246.6625	299.0	667.0000	874.0823
141.0	257.0000	249.1280	300.0	701.0000	879.9947
142.0	206.0000	251.6073	301.0	685.0000	885.9388
144.0	216.5000	256.6081	304.0	778.0000	903.9623
147.0	204.0000	264.2155	308.0	905.0000	928.4465
149.0	291.3333	269.3589	319.0	868.0000	998.5261
150.0	233.5000	271.9524	323.0	1028.0000	1025.0404
152.0	234.0000	277.1833	333.0	729.0000	1093.8329
154.0	210.0000	282.4733	334.0	818.0000	1100.9138
157.0	248.0000	290.5207	339.0	719.0000	1136.8839
161.0	245.0000	301.4635	365.0	1014.0000	1339.9283
162.0	225.5000	304.2379	373.0	1090.0000	1408.1565
167.0	370.6667	318.3454	393.0	1235.0000	1591.6538
402.0	1137.0000	1680.6223	490.0	6098.0000	2804.1735
416.0	1772.0000	1827.4731	493.0	3884.0000	2851.9833
447.0	1912.0000	2192.6101	498.0	7584.0000	2933.2812
449.0	1513.0000	2218.2049	500.0	64275.0000	2966.3751
458.0	1781.0000	2336.6494			

**Таблица А13: Зависимость частоты (F) от политекстии (РТ):
словоформы существительных**

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	1.0000	1.3500	33.0	56.0000	55.9179
2.0	2.0000	2.8244	34.0	86.0000	57.7243
3.0	4.0000	4.3498	35.0	47.0000	59.5342
4.0	6.0000	5.9092	36.0	62.0000	61.3474
5.0	8.0000	7.4944	37.0	80.0000	63.1638
6.0	8.0000	9.1005	38.0	67.0000	64.9835
7.0	11.0000	10.7242	39.0	51.0000	66.8063
8.0	14.0000	12.3630	40.0	80.0000	68.6322
9.0	15.0000	14.0153	41.0	97.0000	70.4610
10.0	17.0000	15.6796	42.0	72.0000	72.2927
11.0	22.0000	17.3547	44.0	104.0000	75.9645
12.0	21.0000	19.0398	45.0	97.0000	77.8046
13.0	20.0000	20.7340	47.0	56.0000	81.4926
14.0	27.0000	22.4368	48.0	152.0000	83.3404
15.0	27.0000	24.1474	49.0	107.0000	85.1908
16.0	27.0000	25.8655	50.0	63.0000	87.0436
17.0	35.0000	27.5906	51.0	96.0000	88.8988
18.0	31.0000	29.3224	52.0	90.0000	90.7564
19.0	35.0000	31.0603	53.0	94.0000	92.6163
20.0	40.0000	32.8043	54.0	91.0000	94.4785
21.0	35.0000	34.5539	55.0	144.0000	96.3430
22.0	38.0000	36.3090	57.0	83.0000	100.0784
23.0	46.0000	38.0692	59.0	132.0000	103.8224
24.0	42.0000	39.8344	60.0	114.0000	105.6975
25.0	44.0000	41.6045	61.0	86.0000	107.5747
26.0	33.0000	43.3791	63.0	195.0000	111.3349
27.0	67.0000	45.1582	64.0	151.0000	113.2180
28.0	51.0000	46.9415	65.0	113.0000	115.1030
29.0	54.0000	48.7290	66.0	126.0000	116.9898
30.0	58.0000	50.5205	69.0	184.0000	122.6614
31.0	61.0000	52.3159	70.0	113.0000	124.5556
32.0	73.0000	54.1151	73.0	125.0000	130.2485
74.0	250.0000	132.1495	105.0	216.0000	191.8228
76.0	153.0000	135.9566	107.0	313.0000	195.7165
77.0	160.0000	137.8626	110.0	243.0000	201.5659
79.0	157.0000	141.6793	111.0	174.0000	203.5180
82.0	152.0000	147.4163	119.0	216.0000	219.1751
85.0	169.0000	153.1669	129.0	179.0000	238.8426
86.0	126.0000	155.0867	136.0	317.0000	252.6694
87.0	144.0000	157.0080	157.0	243.0000	294.4197

91.0	137.0000	164.7073	165.0	304.0000	310.4232
94.0	163.0000	170.4963	180.0	278.0000	340.5642
104.0	204.0000	189.8779			

Таблица А14: Зависимость частоты (F) от политекстии (PT): леммы существительных

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	1.4227	1.5300	27.0	47.0000	45.6033
2.0	2.9231	3.1243	28.0	4225.0000	47.3439
3.0	4.5571	4.7438	29.0	47.0000	49.0864
4.0	6.4224	6.3799	30.0	78.3333	50.8307
5.0	8.1412	8.0284	31.0	52.6000	52.5768
6.0	9.7217	9.6870	32.0	62.3333	54.3245
7.0	11.8442	11.3538	33.0	54.0000	56.0739
8.0	12.1957	13.0279	34.0	55.6000	57.8249
9.0	14.0656	14.7083	35.0	53.0000	59.5774
10.0	15.8108	16.3942	36.0	65.5000	61.3314
11.0	17.8333	18.0853	37.0	55.0000	63.0869
12.0	22.1071	19.7810	38.0	76.3333	64.8438
13.0	24.7143	21.4809	40.0	70.1667	68.3618
14.0	22.1538	23.1848	41.0	72.0000	70.1227
15.0	27.0000	24.8923	42.0	63.8000	71.8850
16.0	31.7059	26.6033	43.0	104.0000	73.6485
17.0	30.5000	28.3174	44.0	86.7500	75.4133
18.0	37.6364	30.0346	45.0	75.0000	77.1792
19.0	34.2727	31.7547	46.0	63.0000	78.9463
20.0	29.5833	33.4774	47.0	102.5000	80.7146
21.0	35.7143	35.2028	48.0	80.5000	82.4840
22.0	38.5000	36.9306	49.0	111.0000	84.2545
23.0	37.5000	38.6608	50.0	73.0000	86.0262
24.0	56.4286	40.3933	51.0	82.2500	87.7988
25.0	36.5000	42.1279	53.0	73.0000	91.3473
26.0	47.4000	43.8646	54.0	98.0000	93.1230
55.0	79.6667	94.8997	97.0	188.5000	170.2418
56.0	69.0000	96.6774	99.0	191.0000	173.8584
57.0	137.0000	98.4561	100.0	198.5000	175.6675
58.0	85.5000	100.2357	107.0	161.0000	188.3461
60.0	89.0000	103.7976	109.0	248.0000	191.9733
61.0	94.0000	105.5799	111.0	161.0000	195.6024
62.0	102.0000	107.3631	116.0	182.0000	204.6836
63.0	140.3333	109.1471	118.0	245.0000	208.3195
65.0	222.0000	112.7177	122.0	202.5000	215.5966
66.0	93.0000	114.5043	123.0	342.0000	217.4171
67.0	127.5000	116.2917	129.0	179.0000	228.3488

68.0	115.0000	118.0798	130.0	201.5000	230.1724
69.0	163.0000	119.8688	133.0	314.0000	235.6452
70.0	131.0000	121.6585	134.0	463.0000	237.4704
72.0	114.0000	125.2402	137.0	300.0000	242.9482
74.0	159.0000	128.8250	149.0	400.0000	264.8947
75.0	131.0000	130.6185	162.0	232.0000	288.7300
77.0	122.0000	134.2075	167.0	411.0000	297.9130
81.0	159.0000	141.3940	185.0	289.0000	331.0384
82.0	137.0000	143.1923	194.0	304.0000	347.6379
85.0	264.0000	148.5911	197.0	437.0000	353.1763
87.0	158.0000	152.1935	200.0	681.0000	358.7173
89.0	116.0000	155.7985	278.0	483.0000	503.5674
90.0	124.0000	157.6018	296.0	862.0000	537.1826
96.0	154.0000	168.4344			

**Таблица А15: Зависимость частоты (F) от политекстии (PT):
словоформы прилагательных**

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	1.2235	1.5200	13.0	22.1111	24.2606
2.0	2.6479	3.2133	14.0	17.1765	26.2822
3.0	4.8000	4.9789	15.0	19.7500	28.3154
4.0	5.2612	6.7931	16.0	24.3636	30.3594
5.0	7.3614	8.6443	17.0	27.4615	32.4137
6.0	8.7288	10.5256	18.0	30.1176	34.4777
7.0	11.1702	12.4322	19.0	34.4545	36.5508
8.0	10.9412	14.3609	20.0	39.2500	38.6328
9.0	16.3103	16.3089	21.0	41.3636	40.7231
10.0	15.1875	18.2744	22.0	44.3571	42.8213
11.0	18.2778	20.2557	23.0	52.8333	44.9272
12.0	23.4444	22.2515	24.0	48.1111	47.0405
25.0	53.7500	49.1608	78.0	125.0000	167.9987
26.0	35.5000	51.2879	82.0	140.0000	177.3220
27.0	41.0000	53.4215	83.0	146.0000	179.6585
28.0	41.0000	55.5615	84.0	386.0000	181.9974
29.0	62.5714	57.7076	94.0	124.0000	205.5047
30.0	55.5714	59.8597	95.0	128.0000	207.8668
31.0	60.6250	62.0175	104.0	267.0000	229.2132
32.0	74.1667	64.1809	108.0	168.5000	238.7488
33.0	55.2500	66.3497	110.0	162.0000	243.5273
34.0	74.0000	68.5237	114.0	178.0000	253.1051
35.0	47.3333	70.7029	115.0	166.0000	255.5037
36.0	59.5000	72.8870	116.0	367.5000	257.9041
37.0	64.5000	75.0760	117.0	470.0000	260.3061
38.0	79.0000	77.2698	118.0	262.0000	262.7098

39.0	77.5000	79.4682	121.0	237.0000	269.9304
40.0	113.0000	81.6711	126.0	243.0000	281.9965
41.0	79.6667	83.8784	135.0	203.0000	303.8114
42.0	56.2500	86.0900	137.0	255.0000	308.6753
43.0	95.5000	88.3059	140.0	200.0000	315.9816
44.0	56.0000	90.5258	144.0	228.0000	325.7430
45.0	70.6667	92.7498	148.0	255.0000	335.5261
46.0	53.0000	94.9778	154.0	253.0000	350.2401
47.0	117.0000	97.2096	161.0	412.0000	367.4646
48.0	76.0000	99.4453	162.0	218.0000	369.9302
50.0	153.0000	103.9277	174.0	273.0000	399.6104
52.0	111.5000	108.4244	182.0	252.0000	419.4891
54.0	152.0000	112.9351	190.0	386.0000	439.4379
57.0	91.0000	119.7260	192.0	302.0000	444.4357
58.0	261.0000	121.9961	193.0	403.0000	446.9361
59.0	75.0000	124.2693	194.0	305.0000	449.4376
60.0	67.0000	126.5456	211.0	311.0000	492.1173
61.0	164.0000	128.8249	215.0	348.0000	502.2004
63.0	94.5000	133.3925	219.0	320.0000	512.2988
64.0	129.6667	135.6807	225.0	741.0000	527.4736
65.0	97.5000	137.9717	232.0	437.0000	545.2185
66.0	87.5000	140.2656	237.0	410.0000	557.9198
67.0	81.0000	142.5622	271.0	434.0000	644.8378
69.0	113.0000	147.1637	284.0	786.0000	678.3088
70.0	96.5000	149.4684	295.0	535.0000	706.7266
71.0	185.5000	151.7759	323.0	1028.0000	779.4393
72.0	103.0000	154.0858	333.0	729.0000	805.5332
73.0	108.0000	156.3984	393.0	1183.0000	963.3577
75.0	106.0000	161.0311	402.0	1282.0000	987.2056
77.0	155.5000	165.6737	416.0	1772.0000	1024.3880

Таблица А16: Зависимость частоты (F) от политекстии (PT): леммы прилагательных

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
1.0	1.2856	1.48	42.0	52.3333	58.68
2.0	2.5472	2.88	43.0	83.0	60.16
3.0	3.7739	4.26	44.0	83.0	61.65
4.0	5.1625	5.62	45.0	59.0	63.14
5.0	6.1695	6.98	46.0	55.0	64.63
6.0	7.8421	8.33	47.0	70.0	66.14
7.0	9.6765	9.68	48.0	71.0	67.65
8.0	11.3333	11.04	49.0	65.5	69.16
9.0	18.6316	12.39	50.0	122.0	70.68
10.0	15.1429	13.74	52.0	75.0	73.73
11.0	14.3880	15.09	53.0	83.3333	75.27
12.0	19.25	16.44	55.0	76.0	78.36
13.0	17.3333	17.79	56.0	67.0	79.91
14.0	16.9286	19.15	58.0	109.0	83.04
15.0	21.2727	20.51	60.0	69.25	86.19
16.0	18.3333	21.87	63.0	92.0	90.96
17.0	22.2	23.23	66.0	87.5	95.78
18.0	19.7778	24.16	67.0	92.5	97.40
19.0	27.1429	25.97	70.0	91.0	102.30
20.0	22.0	27.35	72.0	99.0	105.59
21.0	45.0	28.72	75.0	94.0	110.59
22.0	30.7778	30.11	76.0	141.5	112.26
23.0	30.75	31.49	77.0	94.0	113.95
24.0	27.5	32.88	79.0	110.0	117.33
25.0	32.0	34.28	82.0	248.0	122.46
26.0	43.0	35.67	85.0	129.5	127.65
27.0	33.0	37.08	88.0	138.0	131.15
28.0	45.2	38.49	89.0	155.0	134.67
29.0	39.5	39.89	90.0	123.0	136.44
30.0	42.3333	41.31	92.0	133.0	140.01
31.0	38.0	42.73	94.0	176.3333	143.60
32.0	35.6667	44.16	95.0	128.0	145.4
33.0	57.75	45.59	101.0	167.0	156.39
34.0	42.0	47.02	102.0	137.0	158.24
35.0	45.0	48.46	106.0	173.0	165.74
36.0	74.125	49.90	107.0	173.0	167.63
37.0	50.5	51.35	108.0	151.0	169.53
38.0	46.0	52.81	110.0	162.0	173.35
39.0	63.0	54.27	111.0	190.0	175.27
40.0	85.0	55.74	115.0	173.0	183.04
41.0	68.6667	57.20	116.0	139.0	185.0
118.0	188.0	188.94	215.0	348.0	421.77
121.0	164.0	194.91	219.0	320.0	433.34
124.0	234.0	200.96	220.0	1024.0	436.26

137.0	221.0	227.98	222.0	390.0	442.13
140.0	200.0	234.40	232.0	437.0	472.17
141.0	257.0	236.56	243.0	422.0	506.52
144.0	228.0	243.09	272.0	435.0	604.08
150.0	264.0	256.39	287.0	792.0	658.77
152.0	234.0	260.88	295.0	535.0	689.18
161.0	245.0	281.57	300.0	611.0	708.64
162.0	218.0	283.91	304.0	778.0	724.47
167.0	290.0	295.76	319.0	868.0	785.88
169.0	355.0	300.56	323.0	1028.0	802.83
174.0	273.0	312.73	333.0	729.0	846.27
182.0	252.0	332.70	334.0	818.0	850.70
192.0	302.0	358.52	365.0	1014.0	996.13
193.0	403.0	361.16	393.0	1235.0	1141.87
194.0	305.0	363.81	402.0	1282.0	1191.83
200.0	368.0	379.90	416.0	1772.0	1272.71
209.0	353.0	404.74	447.0	1912.0	1466.34
211.0	330.5	410.37			

**Таблица А17: Зависимость частоты (F) от политекстии (PT):
словоформы глаголов**

Х	У_{эксп}	У_{теор}	Х	У_{эксп}	У_{теор}
1.0	1.1604	1.4600	42.0	63.5000	107.4211
2.0	2.4731	3.2399	43.0	62.2500	110.3676
3.0	3.7283	5.1647	44.0	71.0000	113.3244
4.0	5.6890	7.1899	45.0	83.3333	116.2913
5.0	8.3419	9.2933	46.0	66.0000	119.2681
6.0	8.3871	11.4611	47.0	114.0000	122.2546
7.0	9.1667	13.6841	48.0	60.0000	125.2507
8.0	12.2083	15.9554	50.0	95.0000	131.2709
9.0	12.0465	18.2697	51.0	65.0000	134.2946
10.0	14.1880	20.6230	52.0	85.5000	137.3272
11.0	15.3939	23.0120	54.0	66.0000	143.4187
12.0	17.9000	25.4338	55.0	90.0000	146.4772
13.0	20.0690	27.8861	57.0	94.3333	152.6191
14.0	17.2000	30.3669	58.0	103.0000	155.7023
15.0	22.2000	32.8744	59.0	131.5000	158.7935
16.0	24.1250	35.4071	61.0	93.0000	164.9993
17.0	42.4000	37.9638	62.0	76.0000	168.1138
18.0	28.6429	40.5430	63.0	105.5000	171.2357
19.0	29.0625	43.1439	64.0	106.0000	174.3652
20.0	68.5000	45.7654	66.0	147.0000	180.6460
21.0	29.3846	48.4067	67.0	81.0000	183.7972
22.0	35.5000	51.0668	68.0	692.5000	186.9554
23.0	32.1429	53.7452	69.0	125.2500	190.1207
24.0	34.8000	56.4411	70.0	105.0000	193.2927
25.0	129.8333	59.1540	77.0	214.0000	215.6836
26.0	36.2500	61.8831	80.0	114.0000	225.3753
27.0	45.6000	64.6281	82.0	107.0000	231.8669
28.0	35.0000	67.3883	84.0	307.0000	238.3823
29.0	63.2000	70.1634	86.0	116.0000	244.9210
30.0	38.7500	72.9529	89.0	130.0000	254.7718
31.0	48.0000	75.7563	94.0	110.0000	271.3001
32.0	48.6667	78.5734	95.0	134.0000	274.6218
33.0	48.4000	81.4037	96.0	182.0000	277.9488
34.0	58.6667	84.2468	100.0	137.0000	291.3083
35.0	48.3333	87.1026	101.0	127.0000	294.6608
36.0	48.4000	89.9706	104.0	230.0000	304.7483
37.0	52.5000	92.8506	106.0	171.0000	311.4976
38.0	53.5000	95.7423	109.0	137.0000	321.6573
39.0	69.0000	98.6455	110.0	265.0000	325.0532
40.0	61.2500	101.5598	113.0	211.0000	335.2688
41.0	48.2500	104.4851	118.0	183.0000	352.3848
128.0	282.0000	386.9406	176.0	262.0000	558.0749
147.0	352.0000	453.6990	186.0	645.0000	594.6929
156.0	525.0000	485.7874	190.0	286.0000	609.4239

158.0	303.0000	492.9563	193.0	455.0000	620.5027
159.0	229.0000	496.5462	194.0	755.0000	624.2015
160.0	276.0000	500.1392	199.0	2817.0000	642.7379
162.0	1928.0000	507.3354	202.0	431.0000	653.8933
166.0	428.0000	521.7675			

Таблица А18: Зависимость частоты (F) от политекстии (PT): леммы глаголов

Х	У _{эсп}	У _{теор}	Х	У _{эсп}	У _{теор}
1.0	1.1186	1.26	31.0	35.5	38.42
2.0	2.2067	2.47	32.0	38.0	39.74
3.0	3.2527	3.67	33.0	36.5	41.07
4.0	4.8627	4.86	34.0	59.0	42.4
5.0	5.6364	6.05	35.0	63.5	43.75
6.0	6.8077	7.24	36.0	47.0	45.1
7.0	7.7692	8.44	37.0	41.6667	46.45
8.0	9.2353	9.63	39.0	56.5	49.19
9.0	10.9091	10.83	40.0	64.6667	50.56
10.0	11.9444	12.03	41.0	61.0	51.95
11.0	12.8	13.24	42.0	58.0	53.34
12.0	14.75	14.44	43.0	58.0	54.74
13.0	14.40	15.66	44.0	57.3333	56.14
14.0	16.8	16.87	45.0	53.0	57.56
15.0	19.875	18.1	46.0	80.5	58.98
16.0	18.75	19.32	47.0	55.5	60.41
17.0	19.6667	20.56	49.0	60.6667	63.29
18.0	20.6667	21.79	50.0	58.0	64.74
19.0	29.75	23.04	52.0	75.0	67.66
20.0	23.5	24.29	54.0	83.0	70.61
21.0	29.1667	25.54	55.0	67.0	72.1
22.0	27.6667	26.8	56.0	81.0	73.6
23.0	30.0	28.07	58.0	72.0	76.62
24.0	27.8	29.34	59.0	99.0	78.14
25.0	27.0	30.62	60.0	71.0	79.67
26.0	31.75	31.9	61.0	79.0	81.2
27.0	34.5	33.19	62.0	74.0	82.75
28.0	38.0	34.49	63.0	79.0	84.3
29.0	36.0	35.79	64.0	133.0	85.86
67.0	115.5	90.6	149.0	237.0	253.2
73.0	88.0	100.29	150.0	203.0	255.63
76.0	91.5	105.25	154.0	210.0	265.5
77.0	99.0	106.92	173.0	287.0	315.15
78.0	120.0	108.6	175.0	367.0	320.65
82.0	113.0	115.41	180.0	277.0	334.66

85.0	112.5	120.61	193.0	331.0	372.73
89.0	144.0	127.67	199.0	369.0	391.14
92.0	143.0	133.07	217.0	416.0	449.7
95.0	123.0	138.55	227.0	542.0	484.51
96.0	141.5	140.4	257.0	657.0	599.52
100.0	134.0	147.88	277.0	611.0	685.78
110.0	152.0	167.27	285.0	548.0	722.6
117.0	187.0	181.46	299.0	667.0	790.44
122.0	216.0	191.91	301.0	685.0	800.5
124.0	231.0	196.17	339.0	719.0	1010.17
126.0	236.0	200.47	373.0	1090.0	1230.98
130.0	175.0	209.20	449.0	1513.0	1863.37
132.0	201.0	213.63	458.0	1781.0	1953.12
144.0	205.0	241.21	500.0	7673.0	2421.18
147.0	204.0	248.37			

Таблица А19: Длина (L) в зависимости от частоты (F). Рассчитано с помощью скользящего среднего при интервале 20

X	Уэксп	Утеор	X	Уэксп	Утеор
11.5	8.86	7.9756	26.5	7.58	7.5524
12.5	8.59	7.9323	27.5	7.44	7.5341
13.5	8.35	7.8925	28.5	7.40	7.5166
14.5	8.23	7.8558	29.5	7.22	7.4997
15.5	8.04	7.8216	30.5	7.15	7.4834
16.5	7.90	7.7898	31.5	7.12	7.4676
17.5	7.88	7.7599	32.5	7.22	7.4524
18.5	7.78	7.7318	33.5	7.26	7.4376
19.5	7.61	7.7052	34.5	7.45	7.4234
20.5	7.71	7.6801	35.55	7.58	7.4089
21.5	7.78	7.6562	36.6	7.45	7.3948
22.5	7.75	7.6335	37.65	7.42	7.3811
23.5	7.72	7.6119	38.7	7.25	7.3679
24.5	7.77	7.5912	39.8	7.2	7.3544
25.5	7.67	7.5714	40.9	7.02	7.3413
42.0	6.92	7.3286	112.95	6.45	6.8700
43.1	7.05	7.3162	115.1	6.3	6.8616
44.2	6.9	7.3042	117.3	6.1	6.8531
45.3	6.75	7.2925	119.65	6.15	6.8442
46.4	6.58	7.2811	121.85	6.3	6.8361
47.5	6.63	7.2699	124.1	6.3	6.8279
48.6	6.63	7.2591	126.65	6.35	6.8189
49.75	6.57	7.2480	129.2	6.4	6.8100
50.9	6.77	7.2372	131.85	6.53	6.8010
52.1	6.76	7.2262	134.55	6.53	6.7920

53.3	6.81	7.2154	137.25	6.48	6.7832
54.5	6.76	7.2050	139.9	6.63	6.7747
55.75	6.71	7.1943	142.55	6.58	6.7664
57.0	6.33	7.1839	145.2	6.58	6.7583
58.2	6.46	7.1741	147.8	6.73	6.7504
59.4	6.51	7.1645	150.55	6.7	6.7423
60.6	6.51	7.1552	153.4	6.73	6.7340
61.8	6.64	7.1460	156.1	6.88	6.764
62.95	6.74	7.1374	158.9	6.73	6.7186
64.1	6.99	7.1290	161.8	6.7	6.7106
65.3	6.89	7.1204	164.75	6.95	6.7027
66.55	6.87	7.1116	168.05	6.85	6.6940
67.8	7.15	7.1029	171.25	6.85	6.6858
69.1	7.2	7.0941	174.35	6.9	6.6780
70.55	7.5	7.0845	177.5	6.75	6.6702
72.0	7.67	7.0751	180.8	6.95	6.6621
73.5	7.8	7.0656	184.15	7.0	6.6542
75.0	8.0	7.0562	187.4	6.95	6.6466
76.55	8.02	7.0468	190.5	6.63	6.6394
78.05	7.99	7.0379	193.6	6.78	6.6324
79.7	7.91	7.0283	196.95	6.83	6.6250
81.35	7.76	7.0189	200.95	6.48	6.6163
83.3	7.71	7.0097	205.65	6.48	6.6063
84.75	7.81	7.0001	210.5	6.68	6.5963
86.55	7.66	6.9905	215.6	6.58	6.5860
88.45	7.61	6.9806	221.0	6.75	6.5754
90.5	7.51	6.9702	226.4	6.75	6.5650
92.6	7.53	6.9598	232.9	6.9	6.5529
94.8	7.53	6.9491	240.1	6.9	6.5398
97.05	7.33	6.9385	247.7	7.2	6.5265
99.35	7.28	6.9276	255.5	6.9	6.5133
101.6	7.1	6.9177	263.15	7.05	6.5008
103.85	6.88	6.9078	271.45	7.15	6.4876
106.25	6.93	6.8975	279.7	7.0	6.4750
108.5	6.88	6.8881	287.9	7.0	6.4628
110.75	6.55	6.8789	296.9	6.85	6.4498
306.75	6.75	6.4360	472.35	6.45	6.2571
317.1	6.75	6.4221	496.75	6.3	6.2365
330.2	6.8	6.4051	520.6	5.9	6.2175
346.35	6.7	6.3852	547.45	5.75	6.1971
365.2	6.7	6.3631	581.05	5.45	6.1730
385.65	6.8	6.3405	618.15	5.45	6.1481
405.75	6.8	6.3195	676.15	5.35	6.1122
426.5	6.55	6.2990	746.85	5.25	6.0726
448.6	6.6	6.2782	824.05	5.2	6.0337

Таблица А20: Длина (L) в зависимости от частоты (F). Рассчитано с помощью скользящего среднего при интервале 50

X	У_{эксп}	У_{теор}	X	У_{эксп}	У_{теор}
26.7	7.81	7.4280	56.6	7.27	7.1527
27.74	7.74	7.4138	57.88	7.33	7.1446
27.78	7.61	7.4001	59.16	7.28	7.1368
29.82	7.49	7.3869	60.5	7.24	7.1288
30.86	7.37	7.3742	61.84	7.21	7.1209
31.9	7.32	7.3619	63.2	7.16	7.1131
32.94	7.25	7.3500	64.6	7.14	7.1053
34.0	7.18	7.3383	66.02	7.15	7.0975
35.06	7.17	7.3270	67.48	7.12	7.0897
36.14	7.17	7.3158	69.0	7.13	7.0818
37.22	7.17	7.3050	70.54	7.13	7.0739
38.3	7.21	7.2945	72.12	7.18	7.0616
39.4	7.16	7.2841	73.72	7.16	7.0583
40.5	7.12	7.2741	75.36	7.09	7.0505
41.6	7.15	7.2643	77.0	7.01	7.0428
42.7	7.11	7.2547	78.64	6.98	7.0354
43.8	7.1	7.2455	80.36	6.87	7.0277
44.9	7.07	7.2364	82.06	6.88	7.0203
46.0	7.05	7.2276	83.76	6.87	7.0131
47.1	7.08	7.2191	85.46	6.83	7.0060
48.22	7.02	7.2105	87.16	6.89	6.9991
49.36	7.05	7.2021	88.88	6.89	6.9922
50.5	7.13	7.1938	90.66	6.93	6.9853
51.66	7.08	7.1856	92.44	6.99	6.9784
52.88	7.16	7.1772	94.24	6.91	6.9717
54.1	7.22	7.1689	96.18	6.94	6.9645
55.34	7.22	7.1608	98.16	7.0	6.9574
100.2	7.13	6.9502	175.06	6.6	6.7579
102.3	7.08	6.9430	179.26	6.72	6.7499
104.46	7.07	6.9357	183.76	6.68	6.7415
106.6	7.15	6.9286	188.42	6.78	6.7330
108.78	7.01	6.9216	193.12	6.76	6.7247
110.96	7.08	6.9147	197.88	6.78	6.7164
113.16	7.05	6.9078	202.86	6.78	6.7080
115.42	6.98	6.9010	207.82	6.76	6.6999
117.7	6.98	6.8942	212.74	6.82	6.6920
120.0	7.08	6.8875	218.06	6.84	6.6837
122.34	6.98	6.8808	223.88	6.83	6.6749
124.72	6.94	6.8741	229.82	6.83	6.6661
127.12	7.0	6.8675	236.86	6.75	6.6560

129.66	6.98	6.8607	245.18	6.78	6.6444
132.2	6.93	6.8540	254.74	6.78	6.6317
134.76	6.89	6.8474	265.2	6.74	6.6183
137.32	6.93	6.8409	275.8	6.76	6.6053
139.94	6.94	6.8344	286.68	6.72	6.5924
142.72	6.88	6.8277	298.22	6.64	6.5793
145.45	6.9	6.8211	310.58	6.64	6.5659
148.18	6.78	6.8148	323.1	6.58	6.5529
150.94	6.72	6.8108	335.82	6.52	6.5402
153.84	6.68	6.8020	350.0	6.37	6.5266
156.98	6.54	6.7951	367.02	6.35	6.5110
160.42	6.48	6.7877	385.48	6.35	6.4950
163.94	6.56	6.7803	412.38	6.27	6.4730
167.52	6.52	6.7729	444.58	6.25	6.4486
171.28	6.58	6.7654	479.36	6.21	6.4242

Таблица А21: Длина (L) в зависимости от частоты (F). Рассчитано с помощью скользящего среднего при интервале 100

X	У_{эсп}	У_{теор}	X	У_{эсп}	У_{теор}
59.57	7.4	7.2081	114.01	6.88	6.9690
60.99	7.32	7.1992	116.56	6.9	6.9610
62.48	7.28	7.1902	119.13	6.88	6.9531
63.99	7.24	7.1813	121.93	6.93	6.9447
65.53	7.25	7.1724	124.89	6.92	6.9361
67.1	7.2	7.1636	127.95	6.95	6.9274
68.7	7.16	7.1545	131.06	6.95	6.9187
70.3	7.17	7.1463	134.21	6.95	6.9102
71.92	7.13	7.1378	137.49	6.98	6.9015
73.55	7.13	7.1295	140.77	6.96	6.8931
75.19	7.12	7.1214	144.05	6.95	6.8848
76.86	7.09	7.1123	147.53	6.92	6.8763
78.55	7.07	7.1052	151.26	6.9	6.8674
80.25	7.1	7.0973	155.09	6.85	6.8585
81.97	7.07	7.0895	159.46	6.81	6.8486
83.71	7.03	7.0818	164.45	6.82	6.8376
85.46	7.05	7.0742	170.1	6.8	6.8256
87.28	7.02	7.0664	176.18	6.81	6.8132
89.1	6.99	7.0588	182.34	6.82	6.8010
90.93	6.99	7.0514	188.67	6.82	6.7890
92.77	6.98	7.0441	195.33	6.81	6.7767
94.65	7.0	7.0367	202.41	6.77	6.7642
96.61	7.01	7.0292	209.64	6.76	6.7519
98.57	6.99	7.0219	216.99	6.76	6.7396
100.53	6.97	7.0147	225.1	6.75	6.7270
102.52	6.97	7.0076	234.66	6.71	6.7125
104.59	6.95	7.0003	244.97	6.71	6.6975
106.79	6.91	6.9927	259.49	6.71	6.6775
109.15	6.91	6.9848	276.68	6.67	6.6553
111.55	6.92	6.9769	295.16	6.65	6.6329

Содержание

ВВЕДЕНИЕ	1
ИЗМЕРЕНИЕ И МОДЕЛИРОВАНИЕ В ЛИНГВИСТИКЕ	3
Распределения фундаментальных лингвистических характеристик	5
Зависимости между лингвистическими величинами	5
Структурные законы	6
Законы языковой динамики	7
ПОСТАНОВКА ЗАДАЧИ И МЕТОДИКА ИССЛЕДОВАНИЯ	8
ПОСТРОЕНИЕ ЯЗЫКОВОЙ МОДЕЛИ	10
Введение: элементы на уровне фонем / графем	10
Язык как саморегулирующаяся система	13
Лингвистика и синергетика	17
К вопросу об определении понятий	19
Системно-теоретические средства: алгебра графов и линейные операторы	23
Операторы и графы	23
Структура и функция	24
Упрощение графов	25
Элементы, структура и функция лексической системы	26
Инвентарь фонем	26
Объем словаря	27
Длина	28
Полилексия	30
Политекстия	34
Частота	36
Длина (продолжение)	38
Функции подструктур	43
Потребности, параметры и процессы	43
ПОСЛЕДСТВИЯ, ВЫТЕКАЕМЫЕ ИЗ МОДЕЛИ	44
Прямые связи	44
Опосредованные связи	45
Распределения структурных характеристик	46
Динамика	48
ЭМПИРИЧЕСКАЯ ПРОВЕРКА МОДЕЛИ	49
Определение рабочих понятий	49
Получение и обработка данных	52
Методика эксперимента	54
Полилексия	55
Политекстия	56
Частота	58
Длина	61
Проверка опосредованных зависимостей	63
Сравнение возможных рабочих понятий	70
Обобщение результатов проверки	72

ФЕНОМЕН ОСЦИЛЛЯЦИИ ЛЕКСИЧЕСКИХ ХАРАКТЕРИСТИК	80
ОТ БАЗОВОЙ МОДЕЛИ К ТЕОРИИ ЯЗЫКА	87
Значение разработанной модели	87
Некоторые наброски дальнейшего развития модели	87
Расширение предметной области и уточнение структуры модели	88
Взаимодействие между системой и ее окружением	89
Методика моделирования	90
Временной континуум системы	90
ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ	91
Послесловие: новейшие исследования и дальнейшие перспективы	92
ЛИТЕРАТУРА	96
ПРИЛОЖЕНИЕ: ТАБЛИЦЫ	107