ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ

**фЕДЕРАЛЬНОЕ ГОСУДАРСТвЕННОЕ Бюджетное ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ**

**ВЫСШЕГО ОБРАЗОВАНИЯ**

**«Санкт-Петербургский государственный университет» (СПбГУ)**

|  |  |
| --- | --- |
|  |  |

Выпускная квалификационная работа аспиранта на тему:

***Статистическая динамика лексических единиц***

***в повседневной речи***

***(на материале корпуса "Один речевой день")***

Образовательная программа «Прикладная и математическая лингвистика»

(специальность научных работников

10.02.21 «Прикладная и математическая лингвистика»)

 Автор:

 Косарева Екатерина Олеговна

 Научный руководитель:

 д.ф.н., профессор

 Мартыненко Григорий Яковлевич

Рецензент:

к.ф.н., доцент

Блинова Ольга Владимировна

Санкт-Петербург

2016

**Содержание**

Введение 3

Глава I. Предпосылки создания частотного списка 6

русской повседневной речи

1.1. Статистическая лексикография как раздел лингвистики 6

1.1.1. Основные принципы и понятия статистической лексикографии 6

1.1.2. Типология, структура и функции частотных словарей 7

1.1.3. Частотные словари русского языка 9

1.2. Особенности повседневной разговорной речи 16

1.2.1. Разговорная речь в системе функциональных стилей 16

1.2.2. Инвентарь языковых средств повседневной речи 21

1.3. Звуковой корпус «Один речевой день» 24

1.3.1. История создания корпуса «Один речевой день» 24

1.3.2. Основные принципы сбора и обработки речевого материала 25

1.3.3.Сбалансированность материала в Корпусе «Один речевой день»29

1.3.4. Области применения корпуса повседневного общения 32

Выводы к Главе I 35

Глава II. Разработка и анализ Частотника ОРД 36

2.1. Особенности процесса построения Частотника ОРД 36

2.2. Статистические характеристики Частотника ОРД 44

2.2.1. Выбор статистических параметров оценки 44

2.2.2. Расчет статистических рядов 46

2.3. Динамика статистических рядов в Частотнике ОРД 54

Выводы к Главе II 68

Заключение 69

Список использованной литературы 71

Приложение 1 Частотный список неоднословных единиц 81

Приложение 2 Верхушка Частотника ОРД 86

**Введение**

Настоящее исследование посвящено изучению статистической динамики лексических единиц в повседневной речи.

Математико-статистические методы становятся все более востребованными, в том числе в областях гуманитарного знания. В лингвистике до недавнего времени им была отведена преимущественно область статистической лексикографии, а именно разработка частотных словарей. Качественно новый уровень развития компьютерных технологий повлек за собой изменения в методах создания и использования частотных словарей.

Сегодня наиболее востребованы словари, составленные на основе лингвистических корпусов, так как большие объемы материала позволяют сделать словарь максимально репрезентативным и по возможности полным.

Разговорная речь в лингвистических корпусах традиционно была представлена записями радио- и телепрограмм, публичных выступлений, постановочных диалогов и монологов, субтитров к кинофильмам и телесериалам. Тексты драматургических произведений также рассматриваются в качестве зафиксированной устной речи. В таком виде представлена в «Частотный словарь живой устной речи» под редакцией О, Н. Ляшевской и С. А. Шарова. Этот словарь был создан по материалам Устного подкорпуса Национального корпуса русского языка (далее ‒ НКРЯ), крупнейшего корпуса русского языка на сегодняшний день.

Также разговорная речь представлена в «Частотном словаре словоформ русского языка» в виде подкорпуса драмы. Однако устная спонтанная речь по-прежнему остается сложным материалом для лексикографии.

Звуковой корпус «Один речевой день» (далее ‒ Корпус ОРД) дал лингвистам уникальная возможность анализа спонтанной речи, в том числе и с количественной точки зрения.

**Актуальность** исследования определяется слабой изученностью живой повседневной речи, в особенности с точки зрения количественных данных.

**Цель** работы ‒ описание динамики статистических показателей в лексике современной повседневной речи. Мы ставим перед собой следующие **задачи:**

1. Определить роль и место статистики в современной русской лексикографии;
2. Проанализировать особенности повседневной речи;
3. Охарактеризовать звуковой корпус «Один речевой день»;
4. Определить параметры для составления и статистического анализа частотного списка (далее - Частотник ОРД) повседневной речи.
5. Описать изменение статистических характеристик Частотника при изменении его объема.

Для решения поставленных задач применяются описательные и статистические **методы** (методы оценивания неизвестных параметров распределения и моделирования).

В качестве материала исследования были выбраны текстовые расшифровки звукозаписей 35 информантов в объеме 200 000 словоупотреблений из корпуса «Один речевой день». Текстовые примеры, приводимые в работе, также взяты из Корпуса ОРД.

Объект настоящего исследования ‒ повседневная устная речь. Его предметом стало изменение количественных показателей лексических единиц в частотном словнике.

Научная новизна работы обусловлена тем, что устная спонтанная речь впервые становится материалом подобного статистического исследования.

Теоретическая и практическая значимость полученных результатов заключается в том, что они позволили выявить наиболее устойчивые статистические параметры для частотного словаря устной спонтанной речи, что дает возможность провести в дальнейшем сравнение с другими частотными словарями, а также сделать прогноз изменения показателей при увеличении объема словаря.

Апробация результатов работы: доклады по вопросам, рассматриваемым в выпускной квалификационной работе, были представлены на Международной научной конференции «Голоса города: языковая вариативность и коммуникативное разнообразие» и на 45 Международной филологической научной конференции.

Работа состоит из введения, двух глав, заключения и двух приложений. Первая глава посвящена теоретическим проблемам, связанным с определением места повседневной речи в современной лингвистике, возможностей и инструментов ее статистического описания. Во второй главе проводится анализ количественных изменений в повседневной лексике при изменении объема выборки. В заключении подводятся итоги исследования.

**Глава 1. Предпосылки создания частотного списка русской повседневной речи**

**1.1. Статистическая лексикография как раздел лингвистики**

**1.1.1. Основные принципы и понятия статистической лексикографии**

Современной науке свойственны развитие тесных взаимосвязей между различными отраслями знания и междисциплинарный подход. Эта тенденция отразилась и в лингвистике. Она все чаще обращается к методам математико-статистических расчетов. Одним из наиболее тесно связанных со статистикой направлений стала лексикография.

Статистическая лексикография, занимающаяся теоретической разработкой и практической реализацией частотных словарей, оперирует набором специфических понятий. Среди базовых понятий можно назвать частоту и вероятность.

Частота ˗ это то количество раз, сколько слово встретилось в тексте. Вероятностью называется «идеальная, «истинная» частота, например слова, которая была бы получена, если бы удалось просчитать все тексты, порожденные на сегодняшний день, и все другие, которые могут быть созданы на данном языке»[[1]](#footnote-2). Таким образом, вероятность невозможно вычислить, но можно оценить. Увеличение числа наблюдений приближает частоту к вероятности. Различают нескольких видов частот. Так, абсолютная частота отображает общее количество данных единиц в рассматриваемой выборке, в то время как относительная частота более объективно указывает на соотношение частоты слова и объема выборки. Относительная частота получается при делении абсолютной частоты на длину текста. Сложение относительных частот по порядку списка дает накопленные относительные частоты для каждого из слов. Такие частоты дают представление о том, какую долю текста занимает любое количество самых частых слов.

Еще одно базовое понятие ‒ распределение (или ряд распределения) ‒ это ряд значений признака изучаемого явления, в случае со словарем, ряд частот. Каждое значение в этом ряду имеет свой порядковый номер ‒ ранг.

Общий объем корпуса текстов, рассматриваемых для составления частотного словаря, называется генеральной совокупностью. Она определяется в соответствии с задачами будущего словаря. По словам П. М. Алексеева, при составлении частотного словаря «вариант членения генеральной совокупности на равные части будет обладать ... несравненным преимуществом. Во-первых, он менее произволен, субъективизм составителя здесь проявляется только в перечислении областей генеральной совокупности, но не в приписывании им априорно заданных весов. Во-вторых, благодаря равноправному представительству областей генеральной совокупности обеспечивается статистическая надежность их сопоставления друг с другом»[[2]](#footnote-3) .

Современная лексикография развивается в нескольких направлениях. Создаются как масштабные словари, охватывающие большие объемы лексического материала языка, так и узкоспециализированные словари отдельных уровней языка, функциональных стилей, подъязыков.

**1.1.2. Типология, структура и функции частотных словарей**

Частотный словарь представляет собой упорядоченный список слов с данными о частоте их встречаемости в тексте или группе текстов.

Единицей частотного словаря может быть как лексема, так и словоформа, морфема, словосочетание и т. д.. Материал, в зависимости от целей, преследуемых составителем, может представлять собой целые тексты, фрагменты или группы текстов. При этом, чем больше объем выборки, тем отраженная в словаре информация достовернее.

Слова в частотном словаре могут быть расположены по алфавиту (алфавитно-частотные словари) или в порядке убывания частоты (частотные словари). Кроме основного частотного списка словарь может содержать дополнительные списки, например, по частям речи или разделам корпуса текстов.

Частотный словарь может быть опубликован в полном или неполном виде (в этом случае публикуется зона наиболее употребительных единиц) так как ценность представляет верхушка словника.

Частотные словари классифицируются по таким параметрам, как:

* язык вхождений

- словарь русского языка

- словарь английского языка

- словарь немецкого языка и т. д.

* форма текстов

- словари письменной речи

- словари устной речи

* содержание

- общие словари

- специальные словари (словари терминов, газетных текстов, автора, произведения и т.п.)

* способ составления

- ручной

- машинный

- комбинированный

* вид частотной характеристики

- абсолютная частота

- ipm (instances per million words) ‒ частота на миллион словоформ

- количество источников, в котором данные единицы встречаются

* метод анализа материала

- сплошное описание

- выборочное описание (чаще всего применяется именно этот метод)

Первоначально большинство словарей создавалось с целью улучшения методики преподавания языка как иностранного. Затем стали появляться словари, предназначенные для лингвистических исследований.

**1.1.3. Частотные словари русского языка**

История статистической лексикографии русского языка начинается со словаря Г. Йоссельсона, изданного в 1953 году в Детройте преимущественно на материале языка дореволюционной России.

Позднее появились три наиболее известных в отечественной лексикографии прошлого века словаря: Э. А. Штейнфельдт[[3]](#footnote-4) (1963 г.), Л. Н. Засориной[[4]](#footnote-5) (1977 г.) и Л. Леннгрена[[5]](#footnote-6) (1993 г.).

«Частотный словарь современного русского литературного языка» Э. А. Штейнфельдт был составлен на материале литературы, прессы и радиопередач 50-60-х годов, общим объемом 400 000 словоупотреблений. Помимо частот слов, он содержит статистических характеристики отдельных морфологических категорий. Основным назначением словаря было применение в преподавании русского языка как иностранного.

«Частотный словарь русского языка» под редакцией Л. Н. Засориной, объемом 1 млн словоупотреблений, составлен на текстах четырех групп: художественной прозе, научных и публицистических текстах, газетных и журнальных текстах, драматургии. В ходе разработки словаря были использованы ЭВМ, что и позволило обработать такое количество материала. В дополнение к алфавитно-частотному словнику со статистическими характеристиками лексем, присутствует список омографов и грамматических омонимов. Словарь изначально предназначался для получения сведений о лексике с учетом ее жанровой принадлежности.

Указанные словари в большой степени отражают специфику русского языка советского периода. Так, например, частоты таких характерных для этой исторической эпохи слов как *партия* или *товарищ* в них сопоставимы с частотами служебных слов.

Отдельное внимание следует обратить на деятельность группы «Статистика речи» под руководством Р. Г. Пиотровского, объединившей коллективы ученых из нескольких республик СССР. Группа занималась статистическим описанием разных аспектов языка и разрабатывала частотные словари подъязыков, в том числе на иностранных текстах (например, частотные словари русского, английского, французского и румынского подъязыка электроники[[6]](#footnote-7)). При составлении словарей использовалась следующая методика, описанная П. М. Алексеевым[[7]](#footnote-8):

- специализация на анализе так называемых подъязыков, для чего выбирались тексты определенной жанрово-тематической направленности;

- минимальный размер анализируемого текста - 1000 словоупотреблений

- за словоупотребление принимается любая последовательность букв, ограниченная двумя пробелами, за исключением цифр и формул;

- за лексическую единицу основного частотного списка принимается одна из словоформ

- исходные формы слов входят в дополнительный частотный список

- учитывается лексическая, грамматическая, лексико-грамматическая омонимия

- для каждого подъязыка разрабатывается схема балансировки материала

- объем частотного словаря - 200 000 словоупотреблений

- для описания словаря, материал анализируется в виде выборок по 50 000, 100 000, 150 000 и 200 000 словоупотреблений

- частотный словарь представляется в виде частотного и алфавитного списков

- присутствуют различные списки с дополнительной информацией о словаре

С развитием компьютерных технологий труд лексикографов стал значительно легче с технической точки зрения, и количество словарей, в том числе и частотных, стремительно возросло.

Сегодня самым большим является «Частотный словарь современного русского языка» под ред. О. Н. Ляшевской и С. А. Шарова[[8]](#footnote-9). Этот словарь основан на материале корпуса НКРЯ объемом более 600 млн словоупотреблений[[9]](#footnote-10) и представляет собой систему из нескольких словарей. Кроме общего частотного и алфавитно-частотного списка лемм, в нем представлены частотные списки частей речи, алфавитно-частотный список имен собственных и аббревиатур, а также частотные словари отдельных функциональных стилей (художественной литературы, публицистики, другой нехудожественной литературы и живой устной речи). Форма представления лексических единиц ‒ лемма. Каждая лемма сопровождается информацией о ранге, общей частоте (в единицах ipm), коэффициенте вариации, распределении употребления в разные исторические периоды, о количестве текстов, в которых она встретилась. Также можно узнать сравнительную частоту слова в общем корпусе и подкорпусе. Кроме лемм, существует некоторое количество словоформ, собранных в алфавитно-частотный список.

 Другой крупный частотный словарь ‒ «Частотный словарь словоформ русского языка»[[10]](#footnote-11). Он составлен на основе Корпуса русского литературного языка объемом 1 млн словоупотреблений. Корпус создан в период 2002-2007 гг. в Лаборатории моделирования речевой деятельности кафедры общего языкознания Санкт-Петербургского государственного университета, под руководством В.Б. Касевича. В состав корпуса входит четыре относительно самостоятельных подкорпуса:

1. подкорпус художественной литературы (30% от общего объема);
2. подкорпус публицистики (30 %);
3. подкорпус научно-популярной литературы (20%);
4. подкорпус драмы (20%)

Последний рассматривается в качестве письменно зафиксированной разговорной речи. Главная особенность этого корпуса состоит в наличии акцентной разметки: каждая словоформа имеет символ ударения. Также восстановлена буква «ё» и неоднословные единицы (называемые «составными словами») размечены как единые словоформы.

Частотный словарь, как следует из названия, представлен словоформами. Создатели словаря считают это преимуществом, повышающим надежность и объективность полученных на его основе количественных данных, так как, по словам Л. Лённгрена, «лемма (исходная форма) для каждой словоформы должна указываться вручную. Это означает, что, прежде чем приступить к лемматизации, нужно установить принципы, по которым она будет проводиться. Эти последние могут отличаться от применявшихся в других работах принципов и дать результаты, которые невозможно будет полностью сравнить с уже существующими»[[11]](#footnote-12).

 Полная версия словаря включает 19 частотных и алфавитно-частотных списков, среди которых: общий список, списки по жанровым подкорпусам, списки составных слов, словари имен собственных, омографов, омонимов, словарь ритмических структур. Каждая словоформа снабжена информацией о ее ранге, абсолютной частоте в общем словаре и подкорпусе и данными о количестве текстов, в которых она встретилась.

В не меньшей степени, чем общие частотные словари, востребованы словари специализированные. Этот тип словарей может охватывать самые разные области знания, приведем здесь некоторые из них:

 - «Частотный тематический словарь “Городской общественный транспорт”» Г. А. Мартиновича[[12]](#footnote-13);

 - «Частотный словарь семантических множителей русского языка» Ю. Н. Караулова[[13]](#footnote-14);

 - «Частотный словарь языка массовой коммуникации» Б. В. Кривенко[[14]](#footnote-15);

 - «Статистический словарь русской газеты» А. Я. Шайкевича[[15]](#footnote-16).

 Отдельного упоминания заслуживают частотные словари языка писателей и литературных произведений, среди которых стоит отметить серию словарей, изданных на кафедре математической лингвистики Санкт-Петербургского государственного университета: частотные словари рассказов Чехов[[16]](#footnote-17), Андреева[[17]](#footnote-18), Бунина[[18]](#footnote-19) и Куприна[[19]](#footnote-20). Их особенность заключается в том, что при их составлении фиксировалось увеличение количества лексем при определенном равномерном увеличении объема выборки, а на основании полученных эмпирических данных был сделан прогноз теоретической возможности дальнейшего роста словаря.

 Среди писателей (поэтов), чей язык полностью описан частотными словарями ‒ Пушкин, Достоевский, Грибоедов, Цветаева.

 Для отдельных произведений русской литературы составлены такие словари, как:

 - «Частотный словарь романа Л. Н. Толстого “Война и мир”» З. Н. Великодворской;

 - «Частотный словарь романа Д. Н. Мамина-Сибиряка “Приваловские миллионы”» М. А. Генкель;

 - «Частотный словарь автобиографической трилогии М. Горького» П. М. Алексеева;

 - «Частотный грамматико-семантический словарь языка художественных произведений А. П. Чехова» под общей редакцией А. А. Поликарпова[[20]](#footnote-21) (созданный на базе электронного корпуса текстов 599 произведений А. П. Чехова, общим объемом 1 256 560 словоупотреблений);

 - частотные словари лирики А. Блока З. Г. Минц и О. А. Шишкиной и др[[21]](#footnote-22).

А. Я. Шайкевич предлагает использовать дифференциальный подход к частотному словарю, утверждая, что «частотные словари будут тем более полезны для анализа авторской речи, чем более дифференцированную информацию они будут содержать»[[22]](#footnote-23).

Так как наше исследование касается разговорной речи, необходимо сказать несколько слов и об этой группе частотных словарей. Их значительно меньше словарей письменной речи. Вот некоторые из них:

 - Словарь значимой лексики живой устной речи (часть частотного словаря на базе НКРЯ);

 - часть «Частотного словаря словоформ русского языка», посвященныя драме;

 - «Частотно-семантический словарь русской разговорной речи» Ю. Г. Овсиенко[[23]](#footnote-24) (содержит 400 000 словоупотреблений из живой русской речи; в качестве единицы выступает семантико-структурное целое; представлены лексико-семантические варианты слов)

Более полный список частотных словарей представлен в монографии В. А. Козырева, В. Д. Черняк[[24]](#footnote-25).

Частотные словари имеют большую практическую ценность. Они дают материал для решения как теоретических, так и прикладных задач. С их помощью разрабатываются учебные словари и учебные пособия, словари-минимумы. Использование в автоматических информационно-поисковых системах и программах автоматической обработки текстов.

**1.2. Особенности повседневной разговорной речи**

**1.2.1. Разговорная речь в системе функциональных стилей**

Под разговорной речью традиционно понимается «разновидность устной литературной речи, обслуживающая повседневное обиходно-бытовое общение и выполняющая функции общения и воздействия»[[25]](#footnote-26). Однако в русистике существуют различные подходы к определению места разговорной речи по отношению к литературному языку. Наиболее разработаны следующие их них:

1. разговорная речь как разновидность литературного языка (Ю. М. Скребнев[[26]](#footnote-27));
2. разговорная речь как самостоятельное явление, противопоставляемое кодифицированному литературному языку (Е. А. Земская, М. В. Китайгородская, Е. Н. Ширяев[[27]](#footnote-28));
3. разговорная речь как особый стиль (Г. Г. Инфантова[[28]](#footnote-29), О. Б. Сиротинина[[29]](#footnote-30)).

Важным аргументом в определении сущности разговорной речи может стать характерная ее черта ‒ стилистическая неоднородность. Присутствие большого количества стилистически разных единиц, с одной стороны, в некоторой степени сближает разговорную речь с художественной литературой, допускающей использование широкого круга лингвистических средств. С другой, делает фактически невозможным определение ее как особого стиля из-за отсутствия набора маркеров.

 А. Н. Байкулова рассматривает разговорную речь в контексте неофициального общения. Неофициальное общение определяется как «социально-статусное речевое взаимодействие коммуникантов, реализующееся прежде всего в сфере повседневного (бытового, домашнего, дружеского, иногда делового) общения, не регламентированного жесткими рамками коммуникативных, этических и речевых норм, в наибольшей степени соответствующее волеизъявлению человека и отражающее особенности его языковой личности»[[30]](#footnote-31). Автор предлагает классифицировать следующим образом социальное окружение человека (по убыванию степени близости отношений)[[31]](#footnote-32):

1. Семья
2. Родственники
3. Друзья
4. Знакомые
5. Незнакомые

Все виды общения классифицируются в соответствии с таким распределением отношений. Преимущественная неофициальность и разговорная речь характерны для ближнего круга коммуникантов: они реализуются при общении родственников, семейном и дружеском общении. В общении знакомых, и тем более незнакомых возможно как официальное, так и неофициальное общение, а также смешение регистров.

 Таким образом, разговорную речь можно считать особой системой, своего рода инструментом бытовой повседневной коммуникации.

 Проблема фиксирования повседневной речи породила свою методологию сбора такого речевого материала, учитывая, что его ценность зависит от точности передачи не только содержания, но и формы.

 Существует несколько приемов фиксирования речи[[32]](#footnote-33):

1. Запись на материальном носителе (вручную, при помощи печатной машинки, компьютера и т. п.). Такой прием требует особой подготовки и специальных навыков.
2. Запись на диктофон. Позволяет получить объективные данные при условии, что исследователь минимизирует свое влияние при расшифровке звукозаписи.
3. Извлечение из письменных текстов фрагментов, наиболее близких к живой разговорной речи (дневниковые записи, частная переписка и т. п.). В этом приеме особенно перспективной нам видится работа с так называемыми текстами электронной коммуникации, такими как смс-сообщения и чат-диалоги. Общение этого вида максимально точно имитирует живое общение, вплоть до передачи эмоций знаками-смайлами и сведения текста до отдельных коротких предложений.

 Несмотря на возможность имитации разговорной речи в письменном виде, изначально она устная и существует в следующих формах, каждая из которых имеет свои особенности:

- диалог ‒ это основная форма существования разговорной речи;

- полилог ‒ встречается в повседневной речи чаще, чем в каком-либо другом виде коммуникации;

- монолог ‒ в отличие от монолога в других типах речи, он сочетает заготовленность с большой долей импровизации (к примеру, при пересказе одного и того же события несколько раз, возникает, соответственно, несколько вариантов рассказа). Значимость факта озвучивания в нем превалирует над содержанием. Стремление передать информацию о событиях как можно точнее, особенно с эмоциональной точки зрения, обуславливает наличие сюжета.

 Одна форма реализации разговорной речи может быть встроена в другую. Так, диалоговая реплика может превратиться в монолог (например, при желании собеседника услышать развернутый рассказ в ходе беседы), а в монологе может возникнуть диалог (например, при прерывании рассказа собеседником).

 Одной из главных отличительных черт повседневной устной речи является спонтанность, то есть неподготовленность. В. К. Харченко предлагает также категорию сиюминутности как «отражения в речи непосредственно происходящих событий»[[33]](#footnote-34).

 Помимо спонтанности, разговорная речь характеризуется такими чертами, как:

- линейность (линейное развертывание высказывания во времени);

- неподготовленность (порождение речи возникает одновременно с ситуацией коммуникации);

- непринужденность (разговорную речь используют в ситуациях общения, не ограниченный строгими формальными правилами);

- непосредственное участие коммуникантов (говорящие объединены одной речевой ситуацией и меняются ролями говорящего и слушающего в ходе общения);

- сопровождение ярко выраженными невербальными средствами коммуникации (мимика, жесты, положение тела, голосовые модуляции и т.п. заменяют в разговорной речи часть выраженного в словесной форме смысла);

- сильная связь с внеязыковой ситуацией (для понимания смысла высказывания крайне важен контекст ситуации в целом).

 Экстралингвистические признаки имеют большое влияние не только на успех устной коммуникации, но и на ее вид. Лингвисты выделяют следующие решающие факторы[[34]](#footnote-35):

1. число говорящих (один или более);
2. отнесенность коммуникативного акта ко времени;
3. спаянность коммуникативного акта с ситуацией;
4. равноправие партнеров в акте коммуникации;
5. степень подготовленности речи;
6. мена говорящих (нулевая, относительно малая или относительно большая);
7. фиксированность темы (тема задана заранее или не задана заранее);
8. степень публичности речи (публичная, полупубличная, не публичная или частная речь).

 В зависимости от соотношения совокупности этих признаков общение может принимать вид интервью, дискуссии, разговора, доклада, репортажа, рассказа и т. д..

Тесная связь с внеязыковыми факторами и важная роль всех обстоятельств, сопровождающих момент общения, заставляют обращать особое внимание на так называемую коммуникативную ситуацию.

 Коммуникативная ситуация включает в себя некоторый необходимый для ее возникновения набор компонентов, а именно: говорящего и слушающего, отношения между ними (степень близости, наличие субординации, хорошие или плохое взаимоотношения и т. д.), место, где происходит общение (общественное место, домашняя обстановка, рабочее место и т. п.), тональность общения (нейтральная, дружеская или формальная), способ и средства общения.

 Изменение одного из компонентов обычно приводит к изменению всей коммуникативной ситуации, при этом именно коммуникативная ситуация определяет правила, по которым осуществляется общение, и его форму.

 Речевые ситуации могут быть каноническими и неканоническими. В первых место и время говорящего и слушающего совпадают. Во вторых какая-либо из составляющих не совпадает (не совпадает времени говорящего и слушающего, их местонахождения, отсутствие конкретного адресата).

 О. А. Лаптева выделяет три типа ситуаций общения[[35]](#footnote-36):

1) стереотипные городские диалоги незнакомых лиц;

2) общение знакомых лиц в бытовой обстановке;

3) общение знакомых и незнакомых лиц в производственной и социально-культурной сфере (ситуации непубличного общения и ситуации публичного общения).

 В этой классификации учитываются три основные сферы, где среднестатистический горожанин проводит большую часть своей жизни: дом, работа, улицы города.

**1.2.2. Инвентарь языковых средств повседневной речи**

 Повседневная разговорная речь обладает обширным набором языковых средств на всех языковых уровнях.

 Так, в фонетике наблюдается вариативность реализации фонем, изобилие редуцированных форм. Отдельно стоит отметить важную роль интонации.

 На морфологическом уровне присутствует бóльшая подвижность морфем. Характерна специфическая парадигма имени (есть звательная форма, нет кратких прилагательных) и глагола (нет причастий и деепричастий).

 На синтаксическом уровне превалирует аналитизм (преимущественно семантические средства связи, а не формально-грамматические, распространенность полипредикативных предложений). Наблюдается особое членение высказывание на порции, удобные для восприятия. В отличие от письменной речи, в устной разговорной возможно свободное соединение, наложение, и использование специфических союзов.

 На семантическом уровне за одним означающим закрепляется большое количество означаемых. Это объясняется сильной связью с контекстом, а также постоянным приобретением новых значений.

 В словообразовании наблюдаются специфические суффиксы и способы словообразования, такие как усечение, универбация.

 На всех уровнях наблюдается редукция (так как чаще всего ситуация очевидна всем собеседникам), и в то же время избыточность, которую можно объяснить желанием воспроизвести картину событий как можно точнее.

 Наибольшее количество особенностей можно заметить на лексическом уровне, для которого характерна тематическая и стилевая неограниченность средств. В системе разговорной речи меньше так называемых «пустых клеток», чем в кодифицированном литературном языке. Это объясняется большим количеством неузуальных слов, либо созданных по законам словообразования русского языка на пустующих местах потенциальных слов и словоформ, либо с нарушением этих законов. Также возможна компенсация отсутствующих единиц несвойственными литературному языку употреблениями присутствующих. Словотворчество в этом типе речи развито более, чем в каком-либо другом.

 В повседневной речи очень часто встречаются ошибки сочетаемости и формообразования обусловленные спонтанностью, причем они воспринимаются как более допустимые, чем в письменном языке. Как следствие, именно в разговорной речи появляются отклонения от нормы.

 Как замечает В. К. Харченко, «неподготовленностью, спонтанностью речи легко объяснить не только ошибки и повторы, но также свободу нецензурного слова, и разговорный дискурс как раз та стихия, где недозволенное экспериментирует и процветает, особенно в эпоху отмены цензуры»[[36]](#footnote-37).

 Важное для нашего исследования наблюдение Е. А. Земской касается таких особенностей разговорной речи, как стремление говорящего к воспроизведению готовых формул и одновременно с этим к порождению новых единиц. С одной стороны, «разговорная речь - это система, в которой говорящий легче, свободнее и чаще, чем в КЛЯ, производит новые единицы (а не воспроизводит уже готовые), причем это касается единиц разных уровней: не только словосочетаний, высказываний, словоформ, но и слов»[[37]](#footnote-38). С другой стороны, «разговорная речь - это система, для которой в высшей степени характерно стремление к использованию готовых единиц, всякого рода готовых конструкций, что объясняется автоматизмом протекания речевого акта»[[38]](#footnote-39).

 По словам М. В. Панова, в разговорной речи наблюдается тенденция «инкрустировать литературную речь диалектными, жаргонными характеристиками, остро индивидуальными отклонениями от литературности, но на строгом фоне общелитературной, точно нормированной речи»[[39]](#footnote-40). Это наблюдение относилось к речи 60-х годов прошлого века, однако оно остается справедливым и по сей день. Вероятно, мы можем даже говорить о значительном усилении описанной тенденции. Современная повседневная речь редко обходится без иностранных заимствований (по большей части из английского языка), профессионализмов, жаргонизмов и других «инородных вкраплений».

**1.3. Корпус «Один речевой день»**

**1.3.1. История создания корпуса «Один речевой день»**

Корпусы спонтанной речи начали создаваться с 1990-х гг. Однако и сегодня их число все еще относительно невелико. Это объясняется высокой технической сложностью сбора материала и трудоемкостью его обработки. Иными словами, для создания подобного корпуса требуется привлечение большого научного коллектива, в состав которого должны входить специалисты различного профиля (технические специалисты, языковеды, прикладные лингвисты, социолингвисты, юристы и др.).

Первый опыт звукового корпуса был получен в рамках работы над Британским национальным корпусом. Для русской же речи Корпус ОРД стал первым в своем роде.

Проект Корпуса ОРД был начат в 2007 г. группой лингвистов на базе Филологического факультета Санкт-Петербургского государственного университета и первоначально был одним из блоков Звукового корпуса русского языка. На начальном этапе он характеризовался таким образом: «приоритетная задача данного исследования заключается в том, чтобы получить записи русской спонтанной речи в естественных условиях. Это значит, что, во-первых, ничто не должно влиять на особенности речевого поведения говорящего в конкретных речевых ситуациях. ˂...˃ Во-вторых, информант реализует свое речевое поведение в стандартных для него ситуациях, не меняя ради записи их репертуар и продолжительность»[[40]](#footnote-41). Выполнение поставленных задач стало возможным благодаря методике записи, впервые примененной британскими[[41]](#footnote-42) и японскими[[42]](#footnote-43) исследователями.

Корпус ОРД в настоящее время находится в стадии активной разработки и постоянно пополняется. На сегодняшний день его общий объем составляет около 500 000 словоупотреблений. Произведена запись более 1000 часов звукового материала, полученного от 110 информантов и нескольких сотен их коммуникантов. Возраст участников ‒ от 18 до 77 лет, средний возраст при этом ‒ 37 лет[[43]](#footnote-44).

**1.3.2. Основные принципы сбора и обработки речевого материала**

 В основе сбора материала для Корпуса ОРД лежит так называемый «принцип невода»[[44]](#footnote-45), который позволяет получить весь речевой материал, произнесенный информантом и его коммуникантами. При этом используется методика 24-часовой записи. Она «обеспечивает почти полную свободу говорящего от диктофона или от тех или иных коммуникативных заданий (чтение, пересказ, описание, рассказ на заданную тему), которые неизбежно усложняют его задачу и обеспечивают на «выходе» хоть и вполне спонтанную, но все же экспериментальную речь»[[45]](#footnote-46).

Суть методики состоит в том, что на теле информанта закрепляется диктофон, записывающий все речевые взаимодействия человека с окружающим миром в течение дня. Таким образом, появляется возможность наблюдать за изменением речевого поведения говорящего в зависимости от изменения ситуации и места коммуникации, социальной роли, количества собеседников и т. п..

Немаловажную роль в качестве полученного речевого материала играет подготовка информанта, выбор насыщенного общением, но при этом типичного дня записи. Широко распространено мнение, что осознание факта записи влияет на наше поведение, в том числе и речевое. Однако при таком длительном использовании диктофона, как правило, достаточно скоро происходит адаптация и говорящий забывает о происходящей в это время записи и ведет себя естественно. Это позволяет нам считать влияние данного фактора на речь минимальным.

По завершению записи достаточного количества материала, он проверяется на наличие дефектов записи, описывается и фрагментируется на отдельные эпизоды. Для расшифровки выбираются фрагменты надлежащего качества, отвечающие по коммуникативному сценарию и социологическим параметрам информанта требованиям сбалансированности материала в Корпусе.

Для описания фрагментов, или макроэпизодов, существует набор помет. Под макроэпизодами здесь понимаются крупные эпизоды событий речевого дня, в которых прослеживается единство места, действия и участников коммуникации. Согласно исследованиям Т. Ю. Шерстиновой[[46]](#footnote-47), в среднем за сутки в жизни человека сменяется от 20 до 50 таких эпизодов. Их описание в аннотации происходит по следующим параметрам[[47]](#footnote-48):

тип коммуникации (бытовой разговор, деловое общение, публичное выступление и т.д.)

условия коммуникации (телефонный разговор, «кухонный» разговор и т.п.)

социальная роль говорящего (родственники, друзья, коллеги и т. д.)

место коммуникации (дом, офис, улица, магазин и т. д.)

Для дальнейшей расшифровки и аннотирования полученных файлов используется программа ELAN.

ELAN представляет собой программу многоуровневого лингвистического аннотирования, разработанную в НИИ Психолингвистики им. Макса Планка в Неймегене (Голландия) «специально для архивирования записей языков, находящихся под угрозой исчезновения, в рамках проектов Фонда Фольксвагена»[[48]](#footnote-49). Программа позволяет создавать файлы с расширением .eaf, содержащие аудиоряд (или видеоряд) и синхронизированную с ним текстовую расшифровку. Среди преимуществ программы стоит отметить такие возможности:

возможность задавать неограниченное количество уровней аннотации со сложными связями;

импорт и экспорт файлов между ELAN и другими популярными лингвистическими программами;

экспорт в виде текстовых файлов;

возможность поиска по различным параметрам, а также получения статистической информации.

 На следующем этапе происходит заполнение уровней, специально разработанных для Корпуса ОРД в соответствии с его целями и задачами[[49]](#footnote-50):

Frase (членение записи на «боксы» с репликами говорящих и паузами);

Speaker (информация о говорящих)

Voice (информация о качестве голоса)

Events (неречевые звуковые события)

FonetComment (фонетические отклонения от нормы)

FraseComment (дополнительная информация о реплике)

Notes (общие наблюдения, которые могут быть полезны в дальнейшем)

Episode (обозначение мелких эпизодов)

При заполнении уровней происходит членение речевого потока на фрагменты, которые записываются в отдельные боксы. При этом используется особая система правил сегментации и знаков. Например, при заполнении уровня Frase разработчики отказались от запятых и точек. Вместо них используются показатели пауз, такие как */*, *//*, *\*П*, *()*, *(…)*. Учитывается различная длительность и природа пауз (интонационно-логическая, интонационно-синтаксическая, пустая, пауза хезитации и т. д.). Необходимость фиксировать большое количество паралингвистических явлений в потоке речи привела к появлению системы специальных помет. С ее помощью в расшифровках могут быть обозначены, например, смех (\*С), кашель (\*К), цыканье (\*Ц), вздох (\*О), зевок (\*З), отрицание с закрытым ртом (\*N), шмыганье носом (\*S) и т. п.

В ходе аннотирования также происходит анонимизация текста звукозаписи: изменяются имена, фамилии, адреса и любая другая информация, указывающая на личность говорящего.

Еще одним необходимым этапом является экспертное редактирование полученных в результате описанного выше процесса транскриптов. В ходе редактирования проверяется правильность расшифровки, как с содержательной точки зрения (соответствие текста расшифровки звуковому ряду), так и с формальной (точность и правильность границ боксов, отсутствие промежутков между ними, корректность использованных обозначений, правильность орфографии). При необходимости вносятся исправления.

Таким образом, основная работа сконцентрирована в настоящее время на уровне Frase. В дальнейшем планируется более детальное аннотирование, охватывающее все уровни языковой системы: фонетическую, лексическую, морфологическую, синтаксическую.

**1.3.3. Сбалансированность материала Корпуса ОРД**

Одно из главных требований, обуславливающих репрезентативность лингвистического корпуса, – сбалансированность материала. Для корпусов письменной речи она определяется соотношением текстов по жанрам, временным периодам и источникам. В случае с корпусом такого типа, как Корпус ОРД, добиться пропорционального соотношения материала можно обратившись к социологическим данным.

Корпус ОРД был сбалансирован в соответствии с данными Федеральной службы городской статистики[[50]](#footnote-51).

Для отбора информантов была разработана социологическая анкета, которая содержит данные о поле, возрасте, месте рождения, родном языке, знании других языков, национальности и социальном происхождении родителей, уровне образования, квалификации/специальности по диплому, прошлых и настоящей профессиях, а также основных местах проживания (основным считается место с проживанием больше года)[[51]](#footnote-52).

Таким образом, Корпус ОРД содержит примерно равное количество записей мужчин и женщин. По возрасту информанты были разделены на следующие группы: 18-24 года, 25-34 года, 35-44 года, 45-59 лет, 60 лет и старше «Такая классификация отражает определенные социальные этапы жизни горожанина: до 24 лет - молодость, это время студенчества и первые годы работы, когда человек отрывается от родительского дома и находит свое место в социальной структуре общества; 25-34 и 35-44 - десятилетия, когда человек активно работает, а также заводит семью и из категории «ребенка» окончательно переходит в категорию «родителя»; 45-59 лет - время от социальной зрелости до пенсионного возраста; выше 60 лет - старшее поколение, менее активно принимающее участие в социальной жизни»[[52]](#footnote-53). Наличие такой информации позволяет учитывать возможное региональное влияние, межъязыковую и внутриязыковую интерференцию.

Балансировка по социальным группам может дать информацию для изучения социолектов. В Корпусе представлены следующие группы[[53]](#footnote-54):

* работники, занятые на производстве, строительстве, транспорте, а также представители профессий, связанных с физическим трудом
* служащие силовых структур
* работники сферы услуг
* специалисты по экономической деятельности
* специалисты по информационным технологиям
* специалисты по связям с общественностью
* специалисты, занятые в спорте
* представители творческих профессий
* представители гуманитарных наук
* представители естественных наук
* работники образования
* учащиеся
* неработающие пенсионеры

Важным аспектом представляется степень владения языком. Следует различать, например, владение языком как средством коммуникации и изучение языка филологами. Кроме того, уровень речевой компетенции говорящих также влияет на конечное качество материала в корпусе.

В соответствии с разработанной Н. В. Богдановой-Бегларян классификацией, информантов можно разделить на 4 группы по принципу «профессионального или непрофессионального отношения говорящего к речи»[[54]](#footnote-55):

1. язык - только средство коммуникации (все нефилологи, а также люди, профессионально не работающие с языком)
2. язык - средство коммуникации и объект изучения (все филологи - не преподаватели, в том числе студенты)
3. язык - средство коммуникации и орудие труда (нефилологи, чья профессия связана с речью - преподаватели, актеры, дикторы, лекторы и т.п.)
4. язык - средство коммуникации, объект изучения и орудие труда (преподаватели-филологи)

Уровень речевой компетенции – это «степень свободы говорящего в выборе речевых средств, уровень владения этими возможностями, его способность решать те или иные коммуникативные задачи»[[55]](#footnote-56). Выделяют три уровня речевой компетенции:

* высокий (высшее образование и профессиональное отношение к речи);
* средний (высшее образование и непрофессиональное отношение к речи);
* низкий (отсутствие высшего образования и профессионального отношения к речи).

Подробный анализ факторов, влияющих на состав Корпуса, позволяет регулировать степень их присутствия в и поддерживать пропорциональное их соотношение в речевом материале.

**1.3.4. Области применения корпуса повседневного общения**

Лингвистические корпуса дают большое количество возможностей для исследований в самых разных областях.

1. Составление словарей

Разговорная речь как отдельный объект лексикографического описания сравнительно недавно заинтересовала разработчиков словарей. Во многом это стало возможным благодаря появлению корпусов разговорной речи. Именно поэтому Корпус ОРД становится важным источником материала для подобных работ.

С использованием Корпуса ОРД были разработаны следующие словари:

Словарь русской повседневной разговорной речи (в том числе частотный словарь словоформ, словарь антрополексем);

Словарь редуцированных форм русской речи;

Словарь дискурсивных единиц (вербальных хезитативов)

(подробнее о словарях в монографии Звуковой корпус как материал для анализа русской речи[[56]](#footnote-57)).

В настоящее время планируется разработка частотного словаря повседневной речи.

2. База для исследований.

Корпус может быть использован в качестве базы для исследований в различных областях лингвистики и смежных дисциплин:

прагматические иследования (исследование Ермоловой О. Б. прагматики повседневной речи на примере изменений в течение речевого дня одного из информантов[[57]](#footnote-58))

статистические исследования (исследование Шерстиновой Т. Ю. наиболее употребительных слов в мужской и женской речи в самых частых типах повседневной коммуникации: бытовое, деловое, учебное общение, общение вида «клиент-сервис»[[58]](#footnote-59)).

социолингвистические исследования (см. Н. В. Богданова-Бегларян, Т.Ю. Шерстинова и др.[[59]](#footnote-60))

3. Использование в преподавании русского языка (например, работа Е. А. Зобниной[[60]](#footnote-61))

Звукозаписи Корпуса ОРД обеспечивают возможность тренировать навыки восприятия живой устной речи и порождения речи в повседневной коммуникации. Это дает неоспоримое преимущество по сравнению с другими корпусами при обучении русскому языку как иностранному. Многообразие диалогов и монологов различной бытовой тематики, возможность выбора среди информантов с разным темпа речи позволяет разрабатывать задания любой сложности для работы с учащимися разных уровней подготовки. Среди возможных форм заданий могут быть аудирование, ролевые игры по максимально приближенным к жизни ситуациям, диалоги в виде телефонных разговоров, упражнения на тренировку лексических и грамматических навыков.

4. Разработка программного обеспечения

 Материал Корпуса ОРД может быть использован при разработке программ сегментации, распознавания и синтеза речи, для автоматического снятия омонимии, а также в программах автоматического перевода.

**Выводы к Главе I**

Проанализировав предпосылки для создания частотника повседневной разговорной речи, мы пришли к следующим выводам:

1. Современная статистическая лексикография выходит за рамки создания частотных словарей и предлагает новые методы решения теоретических и практических задач. В настоящее время появляется большое количество частотных словарей самой разной направленности.
2. Повседневная речь изучается сравнительно недавно и не имеет единой системы взглядов на свои проблемы и особенности. Ряд характеристик отличает повседневную речь от других типов речи и литературного языка. Среди важнейших ‒ спонтанность, индивидуальность и отсутствие стилистического однообразия.
3. Звуковой корпус «Один речевой день» наилучшим образом подходит для статистического анализа повседневной речи, так как является на сегодняшний день единственным звуковым корпусом с уникальной методикой спонтанной речи.

**Глава 2. Разработка и анализ Частотника ОРД**

**2.1. Особенности процесса построения Частотника ОРД**

Наше исследование основано на применении статистических методов. «Статистический метод (или просто статистика) ‒ это комплекс приемов и принципов, согласно которым производятся сбор, систематизация, обработка и интерпретация статистических данных с целью получения научных и практических выводов. Статистическими данными называют при этом сведения о некоем числе объектов в какой-либо совокупности, единицы которой обладают теми или иными общими признаками, способны изменять свое качественное и количественное состояние»[[61]](#footnote-62).

Для получения данных об изменении статистических показателей лексики при увеличении объема выборки, нами был составлен частотный список из 200 000 словоупотреблений. Материал для него был взят из расшифровок звукозаписей Корпуса ОРД, представленных в виде списка словоформ. В выборку вошли записи 35 информантов и их коммуникантов. Для наблюдения за динамикой в Частотнике ОРД, список был разбит на 20 «порций» по 10 000 словоупотреблений каждая, так как методика исследования предполагает подсчет статистических параметров при постоянном приросте на 10 000 единиц объема выборки.

Чтобы по возможности исключить предопределенность и обеспечить бóльшую объективность результатов, границы фрагментов не соотносились с началом и концом расшифровок, записями конкретных информантов или конкретных речевых ситуаций. Выборка производилась из общего списка словоформ текстов расшифровок.

Понятие словоупотребления в данной работе включает в себя помимо графических слов (последовательности букв, ограниченной пробелами с двух сторон), неодносложные лексические единицы и имена собственные, состоящие из более чем одного слова. Последнее обусловлено правилами записи, принятыми для расшифровок Корпуса «Один речевой день», предполагающими соединение слов нижним подчеркиванием в названиях.

Что касается неоднословных единиц, под ними мы понимаем устойчивые сочетания, эквивалентные слову, обладающие семантической, грамматической и фонетической неделимостью, которые выполняют следующие функции:

- предлог

- союз / союзное слово

- частица

- вводный оборот

- наречие / предикатив

- формула вежливости

- междометие

При определении списка неоднословных единиц мы опирались на следующие источники:

1. Словарь Р. П. Рогожниковой[[62]](#footnote-63)
2. Словарь Т. Ф. Ефремовой[[63]](#footnote-64)
3. Малый академический словарь[[64]](#footnote-65)
4. Корпусной словарь неоднословных лексических единиц (оборотов) НКРЯ[[65]](#footnote-66)
5. Частотный словарь словоформ русского языка на основе Корпуса русского литературного языка[[66]](#footnote-67)

Список наиболее употребительных неоднословных оборотов и их частот представлен в Таблице 1.

Таблица 1. Наиболее употребительные неоднословные единицы.

|  |  |
| --- | --- |
| Словосочетание | Абсолютная частота |
| как бы Part | 476 |
| потому что | 403 |
| то есть Conj | 360 |
| в общем | 157 |
| может быть | 146 |
| что ли | 140 |
| вот тАк | 135 |
| а вот | 123 |
| в принципе | 115 |
| ну ладно | 111 |
| на самом деле | 101 |
| на хуй | 100 |
| то есть Part | 98 |
| и всё | 96 |
| ну и | 90 |
| да нет | 82 |
| вот так вот | 77 |
| всё равно Adv |
| тАк вот | 75 |
| всё время | 72 |
| если … то |
| всё равно Part | 61 |
| так и | 59 |
| как раз | 56 |
| так что | 55 |
| и так далее | 50 |
| а то Conj | 47 |
| мне кажется | 43 |
| хотя бы Part | 41 |
| что за |
| и вот | 40 |
| ну что |
| ну вот | 38 |
| в смысле | 36 |
| а что | 34 |
| ну всё |
| то ли |
| если бы Conj | 33 |
| один раз |
| до свидания | 32 |
| ещё раз |
| да и | 30 |
| ну и что | 28 |
| ну как |
| в течение | 25 |
| просто так |

Спонтанная речь имеет ряд особенностей, которые безусловно должны быть отражены в частотном словаре и, соответственно, в нашем частотном списке. Одна из важнейших задач – зафиксировать разнообразие устной повседневной речи, поэтому наряду с общей нарицательной лексикой решено было включить в частотник следующие единицы:

1. Имена собственные и их разговорные варианты
* топонимы

Так как проект ОРД территориально охватывает город Санкт-Петербург и Ленинградскую область, наибольшее разнообразие разговорных образований наблюдается именно в топонимике города: *Санкт-Петербург – Питер, Васильевский остров – Васька, Технологический институт – Техноложка* и т. п.. Также присутствуют характерные названия граждан сопредельных государств, например, *финик* (вместо *финн*). Географические названия, состоящие из нескольких слов, приводятся в виде одной единицы.

* имена людей, художественных персонажей и клички

 животных

Разговорные варианты записываются с сохранением особенностей произношения (*Дмитрий – Дима – Димочка – Димон – Димчик – Димуха; Константин – Костя – Костян – Костянка – Костёныч; Дыма – Дымка – Дымочка – Дымуля – Дымулька – Дымулечка – Дымуленька*)

* названия литературных произведений
* коммерческие названия (название компании, лекарства и т.д.)
1. Аббревиатуры и сокращения
2. Иностранные слова

Сохраняется написание латиницей в случае, когда слова произносятся на иностранном языке либо если это общеизвестное иностранное наименование:

* *good bye / my love / good bye !*
* *Windows, DSL, TDK* и т. п.

 В случае если иностранное слово произносится явно по-русски, произношение искажено, а тем более слово изменяется говорящим в соответствии с правилами русского языка, оно записывается кириллицей (*Три минуты и афидерзейн; сидиэрка; ресетом* и т. п.)

1. Окказионализмы

 В эту группу входят слова, которые не представлены в словарях, не являются нормой языка, появились сравнительно недавно и главным образом в контексте неформального общения в сети Интернет (*пока - пукасики, нырять - ныркать; магичить*)

*что ты там* ***магичишь****\* ? · \*П · что делаю ? · \*П · магичишь\* // · \*П · что такое* ***магичишь****\* ? · \*П · знаешь что такое магия ? · \*П · знаю // · \*П · ну вот // · \*П · но такого глагола /* ***магичишь****\* / нету // · \*П · ну хорошо / замени чтобы / \*В (...) с корнем (...) магия была // ну как ? · \*П · давай замени ! · \*П · \*Н // · \*П · я жду // · \*В · есть в русском языке / нормальное слово / колдовство // · \*П · а мне нравиться / магия // · \*П · тогда бы ты еще сказала волшебство //*

1. Бранная лексика

Вошла в словарь как показатель особой эмоциональности спонтанной речи, а также многообразия словотворчества на её основе.

1. Звукоподражание (*тыт тыр-ры-ры ры-ры ры*)
2. Ошибки и оговорки

Эта группа иллюстрирует типичные, часто повторяющиеся ошибки в словах (например, *еёный, евошний, ихний*). Возможно позволит изучить наиболее часто встречающиеся оговорки и объяснить их причины:

 *между прочим у французов тоже простая кухня / просто () это нам теперь здесь её () подают как это самое / как будто* ***пре-парте*** *какое-то //*

*а не надо* ***дорешивать*** */ я говорю* ***прорешивать*** *учебник*

1. Звуки-заполнители пауз хезитации (*э*, *а*, *м* и т. п.)

После отбора словарный материал прошел несколько этапов обработки для составления частотного списка.

Прежде всего было определено, что слова в Частотнике ОРД представляются в начальной форме (в виде лексемы), с пометами в случае снятой частеречной омонимии.

Форма представления во многом обусловлена стандартом расшифровки звукозаписей, принятым составителями Корпуса ОРД. В неоднозначных случаях были приняты следующие правила:

* Неоднословные единицы, такие как названия, состоящие из нескольких слов (за исключением топонимов), составные предлоги и союзы представлены в виде единых неоднословных единиц.
* Слова с дефисом в большинстве случаев записываются через дефис в соответствии с правилами орфографии. Исключение составляют слова, компоненты которых обладают достаточно большой автономностью (так называемые «дублеты», частицы «то», «таки», «ка», «де»).
* Междометия (а-а-а, о-о-о и т.д.) записываются черед дефис, с разным количеством гласных, в зависимости от длительности. Длительность определяется расшифровщиком в ходе работы со звуковым файлом.
* Предлоги и частицы, имеющие фонетические варианты, представлены в виде отдельных лемм.
* Слова с обрывом конца слова по возможности восстанавливаются до первоначального, исходя из контекста (*спрашива... – спрашивать, зарегистриро... – зарегистрировать, представля... – представлять, сейча... – сейчас, мануальна... – мануальный, дореволюцион... – дореволюционный, пародонто... – пародонтоз, копей... – копейка*).

В потоке спонтанной речи довольно часто встречаются слова с оборванной конечной частью. Такие единицы бывает трудно идентифицировать по частеречному признаку и привести к начальной лемме, в таком случае приходится отказаться от занесения их в частотник.. Подобные случаи иллюстрируют ситуации перебивания собеседниками друг друга, прерывания на полуслове, внезапной смены хода мыслей, свойственные потоку живой речи.

Противоположное явление – отсутствие начала слова – также не вошло в частотник. Как и предыдущее, оно может быть связано с наложением друг на друга реплик говорящих. Однако возможны и случаи умышленного «приглушения» говорящим начала слова или его артикуляция без голоса, как своего рода цензура табуированных выражений и тем разговора. Это явление еще предстоит изучить более детально.

* Редуцированные формы слов, такие как *щас, чё* и т.д., заменяются полными формами и записываются в стандартной орфогрфии.
* Омографы различаются выделением ударного гласного.
* Омонимы сопровождаются частеречной пометой (лексическая омонимия не снята).

Далее, в соответствии с описанными принципами, была осуществлена лемматизация. Процедура производилась вручную. При этом необходимо было собрать в цельные единицы неоднословные сочетания, так как в Корпусе ОРД они представлены отдельными словами.

После этого была снята частеречная омонимия и проставлены соответствующие пометы. Омонимия снималась вручную, так как специфический синтаксис спонтанной речи создает определенные трудности при обработке текстов Корпуса ОРД.

После обработки материала и уточнения частот лексем возникла необходимость дополнить первоначальные частотные списки новыми единицами, с тем чтобы количество словоупотреблений составляло 10 000 в каждом, как было задумано изначально.

На следующем этапе работы необходимо произвести расчеты статистических рядов для каждого объема выборки с нарастанием в 10 000 слов в каждом из частотных списков. Другими словами, отдельно для частотного списка в 10 000 словоупотреблений, затем для списка в 20 000 словоупотреблений (первые 10 000 плюс последующие 10 000) и т. д.. Таблица 2 дает представление о наиболее частотных словах в общем Частотнике ОРД.

Таблица 2. Верхушка частотного списка.

|  |  |  |
| --- | --- | --- |
| Ранг | Лексема | Частота |
| 1 | я | 7975 |
| 2 | не | 4863 |
| 3 | вот | 4344 |
| 4 | да Part | 4259 |
| 5 | ну Part | 4256 |
| 6 | и Conj | 3343 |
| 7 | а Conj | 3275 |
| 8 | быть | 2991 |
| 9 | в | 2968 |
| 10 | это Spro | 2879 |

**2.2. Статистические характеристики Частотника ОРД**

**2.2.1. Выбор статистических параметров оценки**

Для описания какого-либо явления в терминах статистики, необходимо прежде всего определить систему параметров, которую целесообразно применить к данному явлению.

В статистике для измерения традиционно применяют три шкалы: номинальную, количественную и порядковую. Измерение с помощью номинальной шкалы, строго говоря, измерением не является. Это скорее классификация. Самая популярная характеристика этой шкалы ‒ мода. Также часто используются объем словаря и максимальная частота.

Самой разработанной методикой вычислений обладает количественная шкала: «здесь на основании теории моментов была сформирована развитая система средних, показателей вариации, характеристик формы распределения и т.п.»[[67]](#footnote-68).

Порядковая, называемая также ординальной, шкала предполагает измерение разной степени интенсивности признака и используется для ранжировки данных. Чаще других исследователями используется такой показатель этой шкалы как медиана.

Наиболее полный перечень параметров описывается в работе Г. Я. Мартыненко и Г. А. Мартиновича[[68]](#footnote-69).

Для статистического описания динамики лексики повседневного общения были выбраны следующие девять параметров:

1. Номинальные:

* Объем словаря (*n*) ;
* Частота самого частотного слова (*F*max);

2. Количественные:

* Средняя частота (*Fср*);
* Коэффициент разнообразия (*K*);

3. Порядковые:

* Медианный ранг (*Meᵣ*);
* Ранговое среднее (*Rср*);
* Индекс концентрации (*γ*);
* Индекс Хирша (*h*);
* Золотое сечение (*Gᵣ*).

Выбранные параметры позволят описать основные статистические характеристики нашей выборки.

**2.2.2. Расчёт статистических рядов**

Рассмотрим такой параметр как объем словаря (*n*). Он показывает количество разноименных единиц (в данном исследовании лексем) в частотном словаре.

Общий объем Частотника ОРД составил 13 200 лемм. Для сравнения, объем словаря рассказов А. П. Чехова составляет 13 736 слов (при объеме выборки 198 066 словоупотреблений), Л. Н. Андреева ‒ 14 132 (при выборке 198 592 словоупотребления), а А. И. Куприна ‒ 20 075 (при выборке 288 260 словоуптреблений). Объем словаря в словаре устной непубличной речи НКРЯ ‒ 4927 лемм при объеме выборки 1 017 568 словоупотреблений.

П. М. Алексеев говорит о том, что «первая тысяча самых употребительных слов может обеспечить покрытие 80% всех словоупотреблений текста, а первые две тысячи слов - до 90%»[[69]](#footnote-70). Это актуально и для нашего Частотника, в котором первая тысяча лемм покрывает около 82% всего объема.

Интересен также тот факт, что первое имя собственное ‒ *Настя* ‒ появляется только с рангом 267 и частотой 87. Также только 12 неоднословных единиц имеют частоту 100 и выше.

Перейдем к частоте самого частотного слова (*F*max). Как мы уже видели в Таблице 2, таковым является местоимение *я*. Здесь стоит отметить, что по данным большинства частотных словарей наиболее употребительным является лексема *и*. В Частотном словаре живой устной речи НКРЯ это место занимает лексема *ну*. Полученные нами результаты, подтвержденные также исследованиями Т. Ю. Шерстиновой на материале того же корпуса[[70]](#footnote-71), позволяют предположить, что такая частность лексемы *я* характерна именно для повседневной речи.

Также стоит обратить внимание на тот факт, что статус самого частотного слова сохраняется на протяжении всего наращивания объема словаря. То же верно и для второго по частотности слова - *не* (Таблица 3). При этом остальные слова ведут себя по-разному: одни постепенно переходят преимущественно на более высокие уровни; другие, напротив, на более низкие; у третьих наблюдаются неустойчивые колебания в разных направлениях. В целом, слова в первой десятке демонстрируют большую долю случайности в перемещении между уровенями.

Таблица 3. Изменение позиций слов в верхушке частотного списка.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **1** | **2** | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 10000 | **я** | **не** | а | ну | он | да | быть | в | вот | и |
| 20000 | **я** | **не** | а | ну | вот | да | быть | это | в | он |
| 30000 | **я** | **не** | а | вот | да | ну | это | быть | и | ты |
| 40000 | **я** | **не** | да | вот | а | ну | это | быть | и | у |
| 50000 | **я** | **не** | да | а | вот | ну | это | быть | и | у |
| 60000 | **я** | **не** | да | а | вот | ну | это | быть | и | у |
| 70000 | **я** | **не** | да | а | ну | вот | это | быть | и | в |
| 80000 | **я** | **не** | да | вот | ну | а | это | и | быть | в |
| 90000 | **я** | **не** | да | вот | ну | а | и | это | в | быть |
| 100000 | **я** | **не** | да | вот | ну | а | и | быть | это | у |
| 110000 | **я** | **не** | да | вот | ну | а | и | быть | это | у |
| 120000 | **я** | **не** | да | вот | ну | а | и | быть | у | это |
| 130000 | **я** | **не** | да | вот | ну | и | а | быть | у | это |
| 140000 | **я** | **не** | да | вот | ну | и | а | быть | у | в |
| 150000 | **я** | **не** | да | вот | ну | и | а | быть | у | в |
| 160000 | **я** | **не** | да | ну | вот | и | а | быть | у | ты |
| 170000 | **я** | **не** | да | ну | вот | а | и | быть | в | у |
| 180000 | **я** | **не** | да | ну | вот | и | а | быть | в | ты |
| 190000 | **я** | **не** | вот | ну | да | и | а | быть | в | это |
| 200000 | **я** | **не** | вот | ну | да | и | а | быть | в | это |

Следующий параметр ‒ средняя частота (*Fср*), или средняя арифметическая, которая вычисляется по формуле:

$$F\_{ср.}=\frac{\sum\_{i=1}^{n}f\_{i}×n\_{i}}{n}$$

где *n* - объем словаря. Полученный статистический ряд представлен в Таблице 3.

Далее обратимся к коэффициенту разнообразия (*K*). Он определяет долю слов с однократным употреблением в словаре и вычисляется по формуле:

$$K=\frac{t}{n}$$

где *t* - число однократных слов.

Коэффициент разнообразия показывает богатство частотного словаря. Г. Я. Мартыненко[[71]](#footnote-72) связывает его с эстетической теорией Айзенка[[72]](#footnote-73) в контексте статистического анализа художественной литературы. По его наблюдениям, коэффициент К в рассказах А. П. Чехова (0,383), Л. Н. Андреева (0,413) и А. И. Куприна (0,389) стремится к гармонии золотого сечения (0,382). В нашем Частотнике этот коэффициент несколько выше.

Стоит отметить, что в словаре «Частотном словаре словоформ русского языка», к примеру, число однократных единиц по сравнению с остальными составляет больше 50%[[73]](#footnote-74), в то время как в Частотнике ОРД ‒ 45%.

Следующий параметр ‒ медианный ранг (*Meᵣ*), то есть «величина ранга, соответствующая накопленной половине объема выборки (числа словоупотреблений) в кумулятивном убывающем ранговом распределении»[[74]](#footnote-75). В случае с Частотником ОРД мы принимаем за медиану конец интервала, включающего в себя ранг накопленной половины, так как имеем дело с большими величинами. Данные о медиане в каждом из исследуемых частотных списков содержатся в Таблице 3.

Отдельно следует рассмотреть ранговое среднее (*Rср*). Этот параметр был впервые предложен и теоретически обоснован лингвистом Г. Я. Мартыненко[[75]](#footnote-76)[[76]](#footnote-77). Он рассчитывается по следующей формуле:

$$R\_{ср}=\frac{\sum\_{i=1}^{n}r\_{i}×f\_{i}}{\sum\_{i=1}^{n}f\_{i}}$$

где *ri* ‒ это ранг единицы в ранговом распределении.

Показатель *Rср* изменяется в диапазоне

$$1\leq R\_{ср}\leq \frac{n+1}{2}$$

достигая максимум при равномерном распределении.

Еще одним значимым параметром является индекс концентрации (*γ*), для расчета которого применяется формула:

$$γ=1-\frac{R\_{у}}{R\_{в}}$$

*Rу* в этой формуле представляет собой ранговое среднее убывающего ранжированного ряда, то есть описанный выше показатель *Rср*. *Rв,* соответственно, ранговое среднее возрастающего ранжированного ряда, вычисляемое по формуле:

$$R\_{в}=n-R\_{у}$$

«Чем больше перепад между «головой» и «хвостом» рангового распределения, тем больше уровень концентрации в этом распределении»[[77]](#footnote-78).

Рассмотрим индекс Хирша (*h*). Этот показатель был предложен в 2005 г. физиком Х. Хиршем[[78]](#footnote-79) и предназначался изначально для оценки научного вклада физиков, а затем и ученых в целом (например, по нему оценивается вклад в науку российских учёных[[79]](#footnote-80)). В контексте лингвистического исследования этот индекс отображает активность слова в словаре и находится в частотном списке по соответствию показателей ранга и частоты.

 Еще один параметр в нашем исследовании золотое сечение (*Gᵣ*) ‒ числовая гармония, достигаемая при накоплении объема выборки в 38,2 %. Оно рассматривается в качестве статистического распределения в работах Г. Я. Мартыненко. В одной из них он выводит из классических уравнений золотого сечения

$$\left\{\begin{array}{c}a+b=c;\\\frac{a}{c}=\frac{b}{a},\end{array}\right.$$

где *c* ˗ целое, *a* ˗ большая часть, а *b* ˗ меньшая,

квадратное уравнение, после решения которого «устанавливается константное долевое участие “меньшего” и “большего” в “ целом”»[[80]](#footnote-81). Эти доли равны 0,618 и 0,382 соответственно.

Идея числовой гармонии применялся в исследованиях поэтических текстов (например, в работах О. Н. Гринбаума[[81]](#footnote-82), Н. А. Васютинский[[82]](#footnote-83)), однако интересно увидеть, как ведет себя этот показатель в повседневной речи.

По наблюдениям Г. Я. Мартыненко, «самобытность поведения ранговых статистик определяется тем, что они реагируют не столько на распределение частот в совокупности, сколько на концентрацию частот в верхней части частотного словаря, указывая на критическую точку, разделяющую словарь на две качественно однородные зоны»[[83]](#footnote-84).

Все статистические данные, полученные в результате подсчетов, представлены в Таблице 3. Также так указано число однократных слов (*t*) для каждой «порции».

Таблица 3. Статистические параметры Частотника ОРД.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Объем выборки | Объем словаря | Частотасамого частотного слова | Средняя частота | Коэффициент разнообразия | Коли-чество одно-кратных слов | Медиана по *R* | Ранговое среднее | Индекс концентра-ции | Индекс Хирша | Золотое сечение |
| *N* | *n* | *F* max | *Fср* | *K* | *t* | *Meᵣ* | *Rср* | *γ* | *h* | *Gᵣ* |
| 10 000 | 2146 | 391 | 4,660 | 0,572 | 1228 | 70 | 323 | 0,823 | 41 | 33 |
| 20 000 | 3254 | 844 | 6,146 | 0,533 | 1735 | 69 | 404 | 0,858 | 54 | 31 |
| 30 000 | 4085 | 1186 | 7,344 | 0,504 | 2059 | 68 | 453 | 0,875 | 66 | 30 |
| 40 000 | 4867 | 1538 | 8,219 | 0,490 | 2384 | 71 | 501 | 0,885 | 75 | 31 |
| 50 000 | 5556 | 1953 | 8,999 | 0,481 | 2671 | 69 | 537 | 0,893 | 83 | 31 |
| 60 000 | 6151 | 2416 | 9,755 | 0,476 | 2929 | 67 | 560 | 0,900 | 92 | 30 |
| 70 000 | 6755 | 2841 | 10,363 | 0,473 | 3198 | 69 | 587 | 0,905 | 101 | 30 |
| 80 000 | 7244 | 3377 | 11,044 | 0,469 | 3395 | 67 | 600 | 0,910 | 105 | 30 |
| 90 000 | 7864 | 3683 | 11,445 | 0,465 | 3657 | 67 | 634 | 0,912 | 112 | 30 |
| 100 000 | 8554 | 4069 | 11,690 | 0,470 | 4018 | 70 | 674 | 0,915 | 116 | 31 |
| 110 000 | 9007 | 4450 | 12,213 | 0,468 | 4211 | 69 | 689 | 0,917 | 122 | 31 |
| 120 000 | 9562 | 4882 | 12,550 | 0,466 | 4455 | 70 | 716 | 0,919 | 129 | 31 |
| 130 000 | 10023 | 5321 | 12,970 | 0,457 | 4577 | 71 | 744 | 0,920 | 135 | 31 |
| 140 000 | 10531 | 5720 | 13,294 | 0,454 | 4783 | 71 | 766 | 0,922 | 138 | 32 |
| 150 000 | 10976 | 6133 | 13,666 | 0,453 | 4976 | 71 | 782 | 0,923 | 143 | 31 |
| 160 000 | 11421 | 6624 | 14,009 | 0,455 | 5195 | 71 | 796 | 0,925 | 147 | 31 |
| 170 000 | 11909 | 6934 | 14,275 | 0,453 | 5395 | 72 | 819 | 0,926 | 151 | 32 |
| 180 000 | 12364 | 7255 | 14,558 | 0,451 | 5572 | 72 | 839 | 0,927 | 154 | 32 |
| 190 000 | 12742 | 7644 | 14,911 | 0,447 | 5702 | 71 | 852 | 0,928 | 159 | 32 |
| 200 000 | 13200 | 7975 | 15,150 | 0,447 | 5898 | 71 | 869 | 0,930 | 163 | 32 |

**2.3. Динамика статистических рядов в Частотнике ОРД**

Статистические, или динамические ряды ‒ это «ряды показателей, характеризующую величину явления по состоянию на определенные моменты (моментные ряды) или за определенные периоды (интервальные ряды)»[[84]](#footnote-85).

В результате вычислений мы получили 9 статистических рядов. В числовом выражении эти данные сложны для восприятия и тем более анализа закономерностей, поэтому представим их в виде графиков. «Статистический график представляет собой чертеж, на котором с помощью условных геометрических образов или знаков описываются в различных разрезах те или иные статистические совокупности»[[85]](#footnote-86). Графические методы, как неотъемлемый инструмент статистики, позволяют визуально представить массив данных таблицы.

Полученные эмпирические ряды распределений «выражают количественные отношения частей качественно определенной совокупности, сложившиеся в силу действия объективных законов развития данного общественного явления, характеризуемого изучаемой совокупностью»[[86]](#footnote-87). Для уменьшения влияния на график данных с большей степенью случайности, необходимо произвести выравнивание (или сглаживание) динамических рядов.

Выравнивание позволяет найти плавную теоретическую кривую, которая наилучшим образом описывала бы статистическое распределение. Оно может преследовать две цели: для изучения общей тенденции ряда, так называемого тренда («очищенную» от случайных колебаний, не имеющих принципиального значения) или, напротив, для выявления резких колебаний (которые могут оказаться периодическими (циклическими) отклонениями) на фоне общей динамики ряда. Цель исследователя обуславливает выбор метода, применяемого для процедуры выравнивания.

Для выравнивания возможно использование следующих методов:

1. Механическое выравнивание

основано на нахождении усредненных значений элементов ряда с учетом значений соседних уровней:

* метод усреднения по двум половинам ряда (деление ряда на две половины и вычисление среднего значения для каждой из них, используется для определения тренда динамического ряда)
* метод укрупнения интервалов (увеличение временных интервалов первоначального динамического ряда до размеров, при которых тенденции его развития приобретают более ярко выраженные черты, и последующий расчёт новых значений)
* метод скользящей средней (построение второго динамического ряда на основе средних арифметических показателей первого, эмпирического, ряда)
* метод экспоненциального сглаживания (схож с предыдущим, однако «здесь более старым наблюдениям приписываются экспоненциально убывающие веса, при этом, в отличие от скользящего среднего, учитываются все предшествующие наблюдения ряда, а не те, что попали в определенное окно»[[87]](#footnote-88))
1. Аналитическое выравнивание

«заключается в том, что, ориентируясь на общую конфигурацию эмпирического («натурального») ряда, нанесенного на диаграмму, выбирают желаемую форму плавной кривой, удовлетворяющей одной из аналитических функций (прямой линии, параболы 2-го или 3-го порядка), и подбирают такие параметры избранной функции, при которой плавная кривая пройдет ближе всего ко всем точкам выравниваемой кривой»[[88]](#footnote-89). То есть предполагает использование аналитических формул:

* линейная зависимость
* параболическая зависимость
* экспоненциальная зависимость

Целям нашего исследования удовлетворяет метод скользящей средней. Далее рассмотрим его более подробно.

Суть метода скользящей средней заключается в том, что «первоначальные значения элементов ряда заменяются средней арифметической величиной внутри выбранного интервала»[[89]](#footnote-90). Иными словами, определенное количество значений эмпирического ряда, начиная с первого, выбирается в качестве интервала. В нашем случае это 3 значения. Для выбранных элементов рассчитывается среднее арифметическое. Далее интервал сдвигается на одно значение и действие повторяется. Так как данный интервал содержит нечетное количество элементов, полученные значения соответствуют среднему из них.

После применения метода скользящей средней мы получили ряды, отличающиеся количеством значений от первоначального. Для заполнения недостающих значений на концах последовательностей используются различные методы. Так, «Н. С. Четвериков выравнивал скользящие средние по параболе второго порядка и экстраполировал ее на недостающие уровни; Я. П. Герчук заполнял пропущенные концы динамического ряда путем вычитания сезонной волны из фактических уровней и сглаживания результатов с уменьшающимся периодом сглаживания»[[90]](#footnote-91). Мы, в свою очередь, применяем метод взвешенных скользящих средних[[91]](#footnote-92). Он позволит добавить по уровню с обоих концов ряда. Необходимое значение для начала ряда получаем по следующей формуле:

$$y\_{+1}=\frac{2y\_{1}+ y\_{2}-y\_{4} }{2}$$

Значение конца ряда, учитывая тот факт, что уровней в нашем распределении 20, рассчитываем по формуле:

$$y\_{-1}=\frac{2y\_{20}+ y\_{19}-y\_{17} }{2}$$

Некоторые исследователи называют одним из главных недостатков метода скользящей средней такую «условность определения сглаженных уровней для точек в начале и конце ряда»[[92]](#footnote-93).

Нанесем обе кривые, как эмпирическую, так и сглаженную, на графики. «Первое нужно для того, чтобы наметить тип и порядок интерполируемой кривой, второе – для того, чтобы глазомерно оценить, в какой мере достигнуты цели выравнивания»[[93]](#footnote-94).

Опираясь на данные графиков, можно проследить характер изменений в Частотнике ОРД в зависимости от увеличения объема выборки.

Рисунок 1. График зависимости объема словаря от объема выборки.

 Рисунок 1а. График зависимости объема словаря от объема выборки при сглаживании ряда.

 На Рисунках 1 и 1а видно, что объем словаря Частотника ОРД постоянно и достаточно равномерно растет с увеличением объема выборки и не демонстрирует тенденции к постепенному замедлению на данном этапе.

Рисунок 2. График зависимости частоты самого частотного слова от объема выборки.

Рисунок 2а. График зависимости частоты самого частотного слова от объема выборки при сглаживании ряда.

 Частота самого частотного слова является одной из самых устойчивых характеристик Частотника ОРД, так как ее кривая постоянна и растет линейно, что подтверждают Рисунки 2 и 2а.

Рисунок 4. График зависимости средней частоты от объема выборки.

Рисунок 4а. График зависимости средней частоты от объема выборки при сглаживании ряда.

 Рисунки 4 и 4а указывают на то, что средняя частота имеет тенденцию к медленному росту с постепенным затуханием при увеличении объема выборки.

Рисунок 5. График зависимости коэффициента разнообразия от объема выборки.

Рисунок 5а. График зависимости коэффициента разнообразия от объема выборки при сглаживании ряда.

Коэффициент разнообразия, в отличие от всех остальных статистик, имеет тенденцию к очень плавному и постепенно замедляющемуся убыванию, что видно по Рисункам 5 и 5а.

Рисунок 3. График зависимости медианы от объема выборки.

Рисунок 3а. График зависимости медианы от объема выборки при сглаживании ряда.

Как видно на Рисунках 3 и 3а, значение медианы Частотника ОРД колеблется в небольшом диапазоне и остается практически на одном уровне, независимо он объема выборки.

Рисунок 6. График зависимости рангового среднего от объема выборки.

Рисунок 6а. График зависимости рангового среднего от объема выборки при сглаживании ряда.

По Рисункам 6 и 6а можно сделать вывод о том, что ранговое среднее возрастает с увеличением объема выборки, однако характеризуется некоторой нестабильностью.

Рисунок 7. График зависимости индекса концентрации от объема выборки.

Рисунок 7а. График зависимости индекса концентрации от объема выборки при сглаживании ряда.

 Индекс концентрации, как показывают Рисунки 7 и 7а, резко возрастает на начальных этапах, но при накоплении определенного объема выборки быстро затухает.

Рисунок 8. График зависимости индекса Хирша от объема выборки.

Рисунок 8а. График зависимости индекса Хирша от объема выборки при сглаживании ряда.

На Рисунках8 и 8а видно, что индекс Хирша увеличивается по мере прироста объема выборки, однако не сохраняет при этом постоянства.

Рисунок 9. График зависимости золотого сечения от объема выборки.

Рисунок 9а. График зависимости золотого сечения от объема выборки при сглаживании ряда.

Золотое сечение представляет собой еще один из постоянных показателей Частотника ОРД. Согласно Рисункам 9 и 9а, он варьируется в пределах четырех единиц, что в масштабах всей выборки можно считать минимальным изменением. При этом колебания не зависят от объема выборки.

Таким образом, мы получили три показателя, обладающих устойчивостью и не чувствительных к изменениям объема выборки: частота самого частотного слова, медианный ранг и золотое сечение.

**Выводы к Главе II**

В ходе подготовки и анализа Частотника ОРД, мы пришли к следующим выводам:

1. При построении частотника повседневной речи важно учитывать специфику лексического материала и исходную форму его представления в корпусе.

2. Для описания частотника могут быть использованы параметры: 1. Номинальные (Объем словаря (*n*) ; Частота самого частотного слова (*F*max)); количественные(Средняя частота (*Fср*); Коэффициент разнообразия (*K*)) и порядковые (Медианный ранг (*Meᵣ*); Ранговое среднее (*Rср*); Индекс концентрации (*γ*); Индекс Хирша (*h*); Золотое сечение (*Gᵣ*)). Оправданность их использования подтверждается результатами статистического анализа.

3. Анализ динамики статистических рядов показывает, что для частотного списка повседневной речи существует три параметра (Частота самого частотного слова (*F*max), Медианный ранг (*Meᵣ*) и Золотое сечение (*Gᵣ*)) , которые не зависят от объема выборки.

**Заключение**

В первой главе мы проанализировали связь статистики с современной лексикографией. Отметили ее возрастающую роль в лингвистических исследованиях. Рассмотрели типы частотных словарей и оценили преимущества современных частотных словарей на примере крупнейших корпусных.

Говоря о повседневной речи, мы отметили растущий интерес к ее изучению. Проанализировали особенности, потенциально влияющие на состав частотных списков. Также отметили отсутствие четкого определения для этого лингвистического явления и неоднозначность подходов.

Рассмотрев возможности и преимущества Корпуса ОРД, мы выделили такие его особенности, определяющие выбор его в качестве материала: методика 24-часовой записи информантов; сбалансированность Корпуса по нескольким параметрам; развитая система аннотирования расшифровок звукозаписей.

Во второй главе мы описали процесс составления Частотника ОРД, особенности методики и вошедших в него единиц. Также отобрали девять статистических характеристик, по которым можно описать его динамику. Произвели расчеты, построение статистических рядов, как эмирических, так и сглаженных, и соответствующих им графиков. Мы отметили, что все параметры, кроме индекса разнообразия, имеют тенденцию к росту по мере увеличения объема выборки. Наибольшую нестабильность проявили коэффициент разнообразия, ранговое среднее и индекс Хирша. Также мы выяснили, что повседневная речь отличается эгоцентричностью и несколько беднее художественной.

Анализ выявил три устойчивых параметра, не зависящих от объемы выборки: частота самого частотного слова, медиана и золотое сечение. На основании этих параметров может быть проведено сравнение различных частотных словарей и прогноз дальнейшего изменения словаря.

Проделанная работа в дальнейшем может лечь в основу построения профилей частотных словарей на порядковых статистиках.

**Список использованной литературы**

**Научная литература**

1. *Алексеев П. М.* Статистическая лексикография (типология, составление и применение частотных словарей) [Текст]: Учеб. пособие. Л.: ЛГПИ, 1975. 120 с.
2. *Алексеев П. М.* Частотные словари и приемы их составления // Статистика речи. Л.: Наука, 1968. С. 61-63.
3. *Алексеев П. М.* Частотные словари: Учебное пособие. СПб.: Изд-во С.-Петерб. ун-та, 2001. 156 с.
4. Анализ временных рядов // StatSoft, Inc. (2012). Электронный учебник по статистике. Москва, StatSoft [Электронный ресурс]. URL: http://www.statsoft.ru/home/textbook/default.htm (дата обращения: 16.06.2016).
5. *Асиновский А. С., Богданова Н. В., Русакова М. В. и др*. Звуковой корпус русского языка повседневного общения «Один речевой день»: концепция и состояние формирования // Компьютерная лингвистика и интеллектуальные технологии. Вып. 7 (14): По матер. межд. конф. «Диалог-2008». М., 2008. С. 488-494.
6. *Баева Е. М.* О способах социолингвистической балансировки устного корпуса (на примере «Одного речевого дня») // Вестник Пермского университета. Российская и зарубежная филология, Вып. 4 (28), 2014. С. 48-57.
7. *Байкулова А. Н.* Неофициальное общение и его разновидности: критерии выделения и реальное функционирование. Саратов: Изд. центр «Наука», 2012. 196 с.
8. *Байкулова А. Н.* Устное неофициальное общение и его разновидности: повседневная речь горожан. Саратов: Изд. центр «Наука», 2014. 216 с.
9. *Богданова Н. В.* Живые фонетические процессы русской речи: Учеб.-метод. пособ. по современному русскому литературному языку. СПб.: Филологический факультет СПбГУ, 2001. 186 с.
10. *Богданова Н. В., Степанова С. Б., Шерстинова Т. Ю.* Звуковой корпус русского языка: новый подход к исследованию речи // Корпусная лингвистика - 2011: Тр. Междунар. конф. СПб., 2011. С. 98-103.
11. *Богданова-Бегларян Н. В., Шерстинова Т. Ю., Блинова О. В. и др.* Звуковой корпус русского языка как база для социолингвистических исследований // Компьютерная лингвистика и интеллектуальные технологии. Материалы конференции, 2015 [Электронный ресурс]. URL: http://www.dialog-21.ru/digests/dialog2015/materials/pdf/Bogdanova-BeglarianNVetal.pdf (дата доступа: 26.04.2016).
12. *Васютинский Н.А.* Золотая гармония. М.: Молодая гвардия, 1990. 238 с.
13. Введение // Частотный словарь словоформ русского языка [Электронный ресурс]. URL: http://www.narusco.ru/STAT004/ (дата обращения: 11.04.2016).
14. *Герчук Я. П.* Графики в математико-статистическом анализе. М.: Статистика, 1972. 78 с.
15. *Герчук Я. П.* Графические методы в статистике. М.: Статистика, 1968. 212 с.
16. *Грешилов А. А., Стакун В. А., Стакун А. А.* Математические методы построения прогнозов. М.: Радио и связь, 1997. 112 с.
17. *Гринбаум О.Н.* Гармония строфического ритма в эстетико-формальном измерении (на материале «Онегинской строфы» и русского сонета). СПб.: Изд-во СПбГУ, 2000.160 с.
18. *Ежов А. И.* Выравнивание и вычисление рядов распределений. М.: Госстатиздат, 1961. 336 с.
19. *Ермолова О. Б.* «Один речевой день» говорящего с точки зрения прагматики // Вестник Пермского университета. Российская и зарубежная филология, Вып. 3 (27), 2014. С. 21-30.
20. Звуковой корпус как материал для анализа русской речи. Часть 2. Теоретические и практические аспекты анализа (Том 2. Звуковой корпус как материал для новых лексикографических проектов). Коллективная монография / Отв. ред. Н. В. Богданова-Бегларян. СПб.: Филологический факультет СПбГУ, 2015. 396 с.
21. *Земская Е. А., Китайгородская М. В., Ширяев Е. Н.* Русская разговорная речь. Общие вопросы. Словообразование. Синтаксис. М.: Наука, 1981. 276 с.
22. *Зобнина Е. А.* Перспективы использования звукового корпуса «один речевой день» в преподавании русского языка как иностранного [Текст] / Е. А. Зобнина // Мир русского слова. 2009, № 4. С. 99-109.
23. *Инфантова Г. Г.* Очерки по синтаксису современной русской разговорной речи. Ростов н/Д: Изд-во Рост. гос. пед. ин-та, 1973. 135 с.
24. *Козырев В. А., Черняк В. Д.* Лексикография русского языка: век нынешний и век минувший: монография. 2-е изд., испр. и доп. СПб.: Изд-во РГПУ им. А. И. Герцена, 2015. 631 с.
25. Лингвистический энциклопедический словарь [Электронный ресурс] / Гл. ред. В. Н. Ярцева. М.: Большая рос. энцикл., 2002. 707, [2] с. URL: http://tapemark.narod.ru/les/index.html (дата обращения: 11.05.2016 .)
26. *Мартыненко Г. Я.*  Золотое сечение формулы изобретения [Текст]: [Автоматизация] // Научно-техническая информация. Сер. 2. Информационные процессы и системы. 2002. N10. С. 22-25.
27. *Мартыненко Г. Я.* Математика гармонии и статистика [Электронный ресурс]. URL: http://www.trinitas.ru/rus/doc/0232/009a/02321149.htm (дата доступа: 30.05.2016).
28. *Мартыненко Г. Я.* Основы стилеметрии. Л.: Изд-во Ленингр. ун-та, 1988. 176 с.
29. *Мартыненко Г. Я., Мартинович Г. А.* Многопараметрический статистический анализ результатов ассоциативного эксперимента. СПб.: Изд-во С.-Петербургского ун-та, 2003. 28 с.
30. *Мартыненко Г. Я., Фомин С. В.* Ранговые моменты // Научно-техническая информация. Сер. 2 1989. № 5. С. 23-29.
31. Национальный корпус русского языка [сайт]. URL: http://www.ruscorpora.ru/.
32. *Овсиенко Ю. Г.* Частотно – семантический словарь русской разговорной речи [Электронный ресурс] . URL: http://www.euralex.org/elx\_proceedings/Euralex1992\_1/017\_J.%2
33. *Панов М. В.* О развитии русского языка в советском обществе // Вопр. языкознания. 1962, № 3.
34. Работа с программой ELAN [Электронный ресурс]. URL: http://www.mpi.nl/tools/elan/tp/how-to/ELAN\_handout\_Russian.pdf (дата доступа: 26.04.2016).
35. Русская авторская лексикография XIX-XX веков. Антология [Элекстронный ресурс]. URL: http://www.slovari.ru/default.aspx?p=5309 (дата доступа: 5.04.2016).
36. *Сиротинина О.Б.*Современная разговорная речь и ее особенности. М.: Знание, 1974. 260 с.
37. *Скребнев Ю. М.* Введение в коллоквиалистику / под ред.Сиротининой О. Б. Саратов: Изд-во Сарат.ун-та, 1985. 210 с.
38. Социально-демографический портрет России: По итогам Всероссийской переписи населения 2010 года / Федер. служба гос. статистики. М.: ИИЦ «Статистика России», 2012. 183 с.
39. Список CItot ≥ 1000 [сайт]. URL: http://www.expertcorps.ru/science/whoiswho/ci86?sortby=h.
40. Статистика речи. Сб. / Отв. ред. Р. Г. Пиотровский. Л.: Наука, 1968. 260 с.
41. Статистика: Курс лекций / Харченко Л. П., Долженкова В. Г., Ионин В. Г. и др.; под ред. канд. экон. наук В. Г. Ионина. Новосибирск: Изд-во НГАЭиУ; М.: ИНФРА-М, 2000. 310 с.
42. *Харченко В. К.* Современная повседневная речь. Изд-е 3-е. М.: Книжный дом «ЛИБРОКОМ», 2012.184 с.
43. *Четвериков Н. С.* Статистические и стохастические исследования. М.: Госстатиздат, 1963. 300 с.
44. *Шайкевич А. Я.* Дифференциальные частотные словари и изучение языка Достоевского (на примере романа «Идиот» // Слово Достоевского. Сб. ст. М.: ИРЯ РАН, 1996. С. 195-253.
45. *Шерстинова Т. Ю.* Коммуникативные макроэпизоды в корпусе повседневной русской речи «Один речевой день»: принципы аннотирования и результаты статистической обработки // Корпусная лингвистика - 2013: Тр. Междунар. конф. СПб., 2013. С. 449-456.
46. *Шерстинова Т. Ю.* Наиболее употребительные слова повседневной русской речи (в гендерном аспекте и в зависимости от условий коммуникации) // Компьютерная лингвистика и интеллектуальные технологии: По матер. ежегодн. межд. конф. «Диалог» (Москва, 1-4 июня 2016). Вып. 15 (22). М.: Изд-во РГГУ, 2016. С. 616-632.
47. *Шерстинова Т. Ю., Степанова С. Б., Рыко А. И.* Система аннотирования в звуковом корпусе русского языка «Один речевой день» // Мат-лы XXXVIII международной конференции. Секция: «Формальные методы анализа русской речи». Март 2009. СПбГУ: СПб. С. 66-75.
48. *Bogdanova-Beglarian N., Sherstinova T., Martynenko G.* The “One Day of Speech” Corpus: Phonetic and Syntactic Studies of Everyday Spoken Russian, Proc. 18th Int. Conf “Speech and Computer” (SPECOM-2015), LNAI, vol. 9319, Springer, Switzerland, 2015. pp. 429-437.
49. *Burnard L.* (ed.). Reference guide for the British National Corpus (XML edition). Published for the British National Corpus Consortium by Oxford University Computing Services. URL: http://www.natcorp.ox.ac.uk/docs/URG/ (дата обращения 15.06.2016).
50. *Campbell N.* Speech & Expression; the Value of a Longitudinal Corpus, LREC-04. Lisbon, 2004. pp. 183-186.
51. *Eysenck H. J.* An Experimental Study of the Good Gestalt. In: Psychological Rewiew. 1942. № 49. Pp. 344-364.
52. *Hircsh J. E*. An index to quantify an individual's scientific research output [Элекстронный ресурс]. URL: ttp://www.pnas.org/content/102/46/16569.full (дата доступа: 15.06.2016).

**Словари и справочники**

1. *Венецкий И. Г., Венецкая В. И.* Основные математико-статистические понятия и формулы в экономическом анализе. Справочник. 2-е изд., перераб. и доп. М.: Статистика, 1979. 447 с.
2. *Ефремова Т. Ф.* Толковый словарь служебных частей речи русского языка [Текст] : ок. 15000 слов. статей. Ок. 22000 семантических единиц / Т. Ф. Ефремова. 2-е изд., испр.. М. : АСТ : Астрель, 2004. 815 с.
3. *Караулов Ю.Н.* Частотный словарь семантических множителей русского языка. М.: Наука, 1980. 207 с.
4. Корпусной словарь неоднословных лексических единиц (оборотов) [сайт]. URL: http://ruscorpora.ru/obgrams.html.
5. *Кривенко Б. В.* Частотный словарь языка массовой коммуникации. Воронеж : Изд-во Воронеж. ун-та, 1992. 218 с.
6. *Леннгрен Л*. Частотный словарь современного русского языка (Lönngren, Lennart. The Frequency Dictionary of Modern Russian). Acta Univ. Ups., Studia Slavica Upsaliensia Uppsala 32. Uppsala, 1993.
7. *Ляшевская О. Н., Шаров С. А.* Новый частотный словарь русской [Электронный ресурс]. URL: http://dict.ruslang.ru/freq.php (дата доступа: 20.05.2016).
8. *Мартинович Г.А.* Частотный тематический словарь «Городской общественный транспорт». СПб, Филологический факультет СПбГУ, 2005. 521
9. *Рогожникова Р. П.* Толковый словарь сочетаний, эквивалентных слову: Ок. 1500 устойчивых сочетаний рус. яз. / Р. П. Рогожникова. М.: ООО «Издательство Астрель»: ООО «Издательство АСТ», 2003. 416 с.
10. Словарь русского языка: В 4-х т. / РАН, Ин-т лингвистич. исследований; Под ред. А. П. Евгеньевой; 4-е изд., стер. М.: Рус. яз.; Полиграфресурсы, 1999.
11. Частотный грамматико-семантический словарь языка художественных произведений А. П. Чехова [Электронный ресурс]. URL: http://www.philol.msu.ru/~lex/chehov.html (дата доступа: 20.05.2016).
12. Частотный словарь рассказов А. И. Куприна / Автор-сост. А.О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: Изд-во С.-Петерб. ун-та, 2012. 552 с.
13. Частотный словарь рассказов А. П. Чехова / Автор-сост. А.О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: Изд-во С.-Петерб. ун-та, 1999. 172 с.
14. Частотный словарь рассказов И. А. Бунина / Автор-сост. А.О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: Изд-во С.-Петерб. ун-та, 2012. 296 с.
15. Частотный словарь рассказов Л. Н. Андреева / Автор-сост. А.О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: Изд-во С.-Петерб. ун-та, 2003. 398 с.
16. Частотный словарь русского языка [Электронный ресурс] / под ред. Л. Н. Засориной. М.: Изд-во «Русский язык», 1977. URL: http://project.phil.spbu.ru/lib/data/slovari/zasorina/zasorina.html (дата доступа: 20.05.2016).
17. Частотный словарь словоформ русского языка [Электронный ресурс]. URL: http://www.narusco.ru/STAT004/ (дата обращения: 11.04.2016).
18. *Шайкевич А. Я., Андрющенко В. М., Ребецкая Н. А.* Статистический словарь языка русской газеты (1990-е годы). Т. 1. М.: Языки славянской культуры, 2008. 580 с.
19. *Штейнфельдт Э. А.* Частотный словарь современного русского литературного языка: справочник для преподавателей русского языка / под ред. В. А. Ицковича. М.: Прогресс, 1973. 228 с. [1-е изд. в 1963 г.].

**Приложение 1**

**Частотный список неоднословных единиц**

|  |  |
| --- | --- |
| как бы Part | 476 |
| потому что | 403 |
| то есть Conj | 360 |
| в общем | 157 |
| может быть | 146 |
| что ли | 140 |
| вот тАк | 135 |
| а вот | 123 |
| в принципе | 115 |
| ну ладно | 111 |
| на самом деле | 101 |
| на хуй | 100 |
| то есть Part | 98 |
| и всё | 96 |
| ну и | 90 |
| да нет | 82 |
| вот так вот | 77 |
| всё равно Adv | 77 |
| тАк вот | 75 |
| всё время | 72 |
| если … то | 72 |
| всё равно Part | 61 |
| так и | 59 |
| как раз | 56 |
| так что | 55 |
| и так далее | 50 |
| а то Conj | 47 |
| мне кажется | 43 |
| хотя бы Part | 41 |
| что за | 41 |
| и вот | 40 |
| ну что | 40 |
| ну вот | 38 |
| в смысле | 36 |
| а что | 34 |
| ну всё | 34 |
| то ли | 34 |
| если бы Conj | 33 |
| один раз | 33 |
| до свидания | 32 |
| ещё раз | 32 |
| да и | 30 |
| ну и что | 28 |
| ну как | 28 |
| в течение | 25 |
| просто так | 25 |
| да ладно | 24 |
| ну как там | 24 |
| ну конечно | 24 |
| первый раз | 23 |
| по вечерам | 22 |
| вот и | 21 |
| до сих пор | 21 |
| если б Conj | 21 |
| слава богу | 21 |
| то же самое | 21 |
| да ну | 20 |
| в основном | 19 |
| скорее всего | 19 |
| если что | 18 |
| и так | 18 |
| к сожалению | 18 |
| как будто Conj | 18 |
| по поводу | 17 |
| только что Adv | 17 |
| тут же | 17 |
| а ещё | 16 |
| в итоге | 16 |
| не так | 16 |
| так сказать | 16 |
| вместе с | 15 |
| как и | 15 |
| по крайней мере Part | 15 |
| шутка ли | 15 |
| вот и всё | 14 |
| вряд ли | 14 |
| ничего себе Intj | 14 |
| при этом | 14 |
| так же | 14 |
| тем более Conj | 14 |
| в общем-то | 13 |
| не то что | 13 |
| ну а | 13 |
| рядом с | 13 |
| сразу же | 13 |
| надо же | 12 |
| при чём здесь | 12 |
| в конце концов | 11 |
| на месте | 11 |
| по одному | 11 |
| во время | 10 |
| для того чтобы | 10 |
| до свиданья | 10 |
| между прочим | 10 |
| не просто | 10 |
| по идее | 10 |
| по сути дела | 10 |
| с другой стороны | 10 |
| в гости | 9 |
| в порядке Adv | 9 |
| вот именно | 9 |
| вроде как Part | 9 |
| друг друга | 9 |
| каждый раз | 9 |
| как говорится | 9 |
| не только | 9 |
| одну секундочку | 9 |
| а вот и | 8 |
| во всяком случае Part | 8 |
| да уж | 8 |
| для начала | 8 |
| доброе утро | 8 |
| как сказать | 8 |
| мало ли Part | 8 |
| на всякий случай | 8 |
| не раз | 8 |
| от природы | 8 |
| так как | 8 |
| что угодно | 8 |
| а если | 7 |
| в конце Pr | 7 |
| в районе | 7 |
| да и всё | 7 |
| и то Conj | 7 |
| к примеру | 7 |
| короче говоря | 7 |
| не больше Part | 7 |
| от того что | 7 |
| так же как | 7 |
| тем не менее | 7 |
| чем … тем | 7 |
| чуть ли не | 7 |
| в виде | 6 |
| в долг | 6 |
| в плане | 6 |
| в результате Adv | 6 |
| в чём дело | 6 |
| вроде бы Part | 6 |
| к себе | 6 |
| к чему | 6 |
| как только | 6 |
| мало ли Adv | 6 |
| не больше Adv | 6 |
| не то | 6 |
| не то чтобы | 6 |
| ни разу | 6 |
| ну что же | 6 |
| пока ещё | 6 |
| ради бога | 6 |
| с удовольствием | 6 |
| со стороны Pr | 6 |
| судя по | 6 |
| у себя | 6 |
| а то Part | 5 |
| а тут | 5 |
| а что ж | 5 |
| больше всего | 5 |
| в курсе | 5 |
| в первую очередь | 5 |
| всего лишь | 5 |
| ещё бы | 5 |
| и прочее | 5 |
| и тут | 5 |
| как … так и | 5 |
| как всегда | 5 |
| ну тебя | 5 |
| оттого что | 5 |
| пока не | 5 |
| пока что | 5 |
| после того как | 5 |
| сам по себе | 5 |
| ух ты | 5 |
| честно говоря | 5 |
| что ль | 5 |
| что называется | 5 |
| в качестве | 4 |
| в крайнем случае | 4 |
| в начале Pr | 4 |
| за счёт | 4 |
| как его | 4 |
| как правило | 4 |
| кроме того что | 4 |
| мало того что | 4 |
| можно сказать | 4 |
| на фиг | 4 |
| на хер | 4 |
| ни при чём | 4 |
| ни фига | 4 |
| ни хуя | 4 |
| ну что ж | 4 |
| по имени | 4 |
| по крайней мере Adv | 4 |
| по ночам | 4 |
| по очереди | 4 |
| сколько раз | 4 |
| так чтобы | 4 |
| хоть бы | 4 |
| чёрт-те что | 4 |
| что ж | 4 |
| чуть не | 4 |
| а там | 3 |
| будь здоров | 3 |
| в гостях | 3 |
| в дальнейшем | 3 |
| в пределах | 3 |
| в силу | 3 |
| в среднем | 3 |
| в том числе | 3 |
| вок тАк | 3 |
| всё же | 3 |
| да вОт | 3 |
| да ещё | 3 |
| дай бог | 3 |
| если только | 3 |
| ещё и | 3 |
| и то Part | 3 |
| к тому же | 3 |
| как бы Conj | 3 |
| как есть | 3 |
| как обычно | 3 |
| как положено | 3 |
| каким образом | 3 |
| как-то раз | 3 |
| на виду | 3 |
| на грани | 3 |
| на мой взгляд | 3 |
| на ночь | 3 |
| на прощанье | 3 |
| на ходу | 3 |
| на хуя | 3 |
| начиная с | 3 |
| не до | 3 |
| не за что | 3 |
| не так уж | 3 |
| ни в коем случае | 3 |
| ни фига себе | 3 |
| ну а если | 3 |
| ну так | 3 |
| один и тот же | 3 |
| одно и то же | 3 |
| под ногами | 3 |
| прежде всего | 3 |
| прежде чем | 3 |
| с детства | 3 |
| с помощью | 3 |
| сам понимаешь | 3 |
| само собой | 3 |
| своим ходом | 3 |
| так же как и | 3 |
| так называемый | 3 |
| тем более Part | 3 |
| только бы Part | 3 |
| что вы | 3 |
| что же | 3 |
| а вон | 2 |
| а как же | 2 |
| а ну-ка | 2 |
| более чем | 2 |
| в зависимости от | 2 |
| в количестве | 2 |
| в любом случае | 2 |
| в одиночку | 2 |
| в результате Pr | 2 |
| в самом деле | 2 |
| в самый раз | 2 |
| в связи с | 2 |
| в связи с тем что | 2 |
| в середине Pr | 2 |
| в сторону Adv | 2 |
| в том числе и | 2 |
| в честь | 2 |
| во всяком случае Adv | 2 |
| вот это да | 2 |
| всего хорошего | 2 |
| всякого рода | 2 |
| где угодно | 2 |
| дело не в | 2 |
| до конца | 2 |
| до тех пор пока | 2 |
| если б … то | 2 |
| если бы … то | 2 |
| если не | 2 |
| за бесплатно | 2 |
| за границей | 2 |
| за границу | 2 |
| здравия желаю | 2 |
| из-за того что | 2 |
| исходя из | 2 |
| как будто бы Conj | 2 |
| как можно | 2 |
| кстати говоря | 2 |
| ладно бы | 2 |
| меньше всего | 2 |
| на глаз | 2 |
| на днях | 2 |
| на память | 2 |
| на самом деле  | 2 |
| на свете | 2 |
| на хрен | 2 |
| насчёт того что | 2 |
| не без | 2 |
| не более Part | 2 |
| не дай бог | 2 |
| не считая | 2 |
| не то чтобы … но | 2 |
| не только … но и | 2 |
| ни в какую | 2 |
| ни с того ни с сего | 2 |
| ничего подобного | 2 |
| но и | 2 |
| одним словом | 2 |
| ох ты | 2 |
| перед тем как | 2 |
| по блату | 2 |
| по отдельности | 2 |
| по порядку | 2 |
| по сути | 2 |
| при чём тут | 2 |
| с радостью | 2 |
| с трудом | 2 |
| со временем | 2 |
| стало быть | 2 |
| судя по всему | 2 |
| так и быть | 2 |
| так и так | 2 |
| тем более что | 2 |
| тоже мне | 2 |
| тому назад | 2 |
| фу ты | 2 |
| хоть и | 2 |
| чаще всего | 2 |
| честное слово | 2 |
| что касается | 2 |
| а всё-таки | 1 |
| а именно | 1 |
| а ну | 1 |
| а также | 1 |
| благо что | 1 |
| более того Adv | 2 |
| бог знает что | 1 |
| бога ради | 1 |
| более или менее | 1 |
| будто бы Conj | 1 |
| в будущем | 1 |
| в долгу | 1 |
| в клеточку | 1 |
| в надежде | 1 |
| в надежде на то что | 1 |
| в ногу | 1 |
| в обход Adv | 1 |
| в один прекрасный день | 1 |
| в один присест | 1 |
| в особенности Part | 1 |
| в порядке Pr | 1 |
| в процессе | 1 |
| в результате | 1 |
| в своё время | 1 |
| в силах | 1 |
| в случае | 1 |
| в случае чего | 1 |
| в состоянии | 1 |
| в счёт | 1 |
| в то же время | 1 |
| в том числе Conj | 1 |
| в упор | 1 |
| в целом | 1 |
| видишь ли | 1 |
| вместо того чтобы | 1 |
| во главе Adv | 1 |
| во как Adv | 1 |
| вон как | 1 |
| вон какая | 1 |
| вот что | 1 |
| вплоть до того что | 1 |
| всё ещё | 1 |
| всё ж | 1 |
| где бы ни | 1 |
| грубо говоря | 1 |
| да вон | 1 |
| да ещё и | 1 |
| да мало ли | 1 |
| дело в том что | 1 |
| для того чтоб | 1 |
| до свиданьячки | 1 |
| дО смерти | 1 |
| до того как | 1 |
| до хера | 1 |
| до хуя | 1 |
| ёшкин кот | 1 |
| за один присест | 1 |
| за счёт того что | 1 |
| затем чтобы | 1 |
| знаешь ли | 1 |
| и правда | 1 |
| к лицу | 1 |
| казалось бы | 1 |
| как будто б Part | 1 |
| как же | 1 |
| как минимум | 1 |
| как нарочно | 1 |
| как ни | 1 |
| как таковой | 1 |
| как хочешь | 1 |
| каким-то образом | 1 |
| конечно же | 1 |
| куда угодно | 1 |
| лишь бы | 1 |
| мало кто | 1 |
| между тем | 1 |
| менее всего | 1 |
| мне казалось | 1 |
| на вид | 1 |
| на время | 1 |
| на выбор | 1 |
| на здоровье | 1 |
| на основании | 1 |
| на основе | 1 |
| на почве | 1 |
| на равных | 1 |
| на радостях | 1 |
| на слух | 1 |
| на удивление | 1 |
| на фига | 1 |
| навряд ли | 1 |
| надо ж | 1 |
| надо сказать | 1 |
| насчёт того | 1 |
| не в состоянии | 1 |
| не говоря о | 1 |
| не грех | 1 |
| не правда ли | 1 |
| не скажи | 1 |
| не скажите | 1 |
| не смотря на то что | 1 |
| не спеша | 1 |
| не так уж и | 1 |
| не то что … но | 1 |
| не то чтоб | 1 |
| не то чтоб … но | 1 |
| не то чтобы … а | 1 |
| не только … но | 1 |
| независимо от того | 1 |
| несмотря на то что | 1 |
| нет чтоб | 1 |
| нечто вроде | 1 |
| ни к чему | 1 |
| ни хера | 1 |
| ни хуя себе | 1 |
| ничего себе Adv | 1 |
| ну вот и всё | 1 |
| ну и всё | 1 |
| ну и как | 1 |
| ну и ну | 1 |
| ну и чего | 1 |
| ну и что ж | 1 |
| ну как же | 1 |
| ну что ты | 1 |
| одну минуточку | 1 |
| от начала до конца | 1 |
| ох как | 1 |
| ох уж этот | 1 |
| первым делом | 1 |
| по виду | 1 |
| по дешёвке | 1 |
| по мере того как | 1 |
| по отношению к | 1 |
| по памяти | 1 |
| по поводу того что | 1 |
| по полной | 1 |
| по причине | 1 |
| по пути Adv | 1 |
| по слогам | 1 |
| по сравнению с | 1 |
| по старой памяти | 1 |
| по существу | 1 |
| под утро | 1 |
| понимаешь ли | 1 |
| после чего | 1 |
| постольку поскольку | 1 |
| при себе | 1 |
| при случае | 1 |
| при том что | 1 |
| про себя | 1 |
| ради того чтоб | 1 |
| раз и навсегда | 1 |
| рано или поздно | 1 |
| с благодарностью | 1 |
| с интересом | 1 |
| с лишним | 1 |
| с небольшим | 1 |
| с нетерпением | 1 |
| с тем чтобы | 1 |
| с тех пор | 1 |
| с точки зрение | 1 |
| с целью | 1 |
| сколько угодно | 1 |
| скорей всего | 1 |
| со вкусом | 1 |
| со стороны Adv | 1 |
| собственно говоря | 1 |
| спокойной ночи | 1 |
| так ведь | 1 |
| так и есть | 1 |
| так нет | 1 |
| так себе | 1 |
| так уж и | 1 |
| таким образом | 1 |
| то бишь | 1 |
| только что Part | 1 |
| хотя б Part | 1 |
| хотя и | 1 |
| честно сказать | 1 |
| что ты | 1 |

**Приложение 2**

**Верхушка Частотника ОРД**

|  |  |  |
| --- | --- | --- |
| Ранг | Лемма | Частота |
| 1 | я | 7975 |
| 2 | не | 4863 |
| 3 | вот | 4344 |
| 4 | да Part | 4259 |
| 5 | ну Part | 4256 |
| 6 | и Conj | 3343 |
| 7 | а Conj | 3275 |
| 8 | быть | 2991 |
| 9 | в | 2968 |
| 10 | это Spro | 2879 |
| 11 | ты | 2763 |
| 12 | у Pr | 2724 |
| 13 | он | 2484 |
| 14 | на Pr | 2271 |
| 15 | там Advpro | 2096 |
| 16 | мы | 1930 |
| 17 | она | 1825 |
| 18 | с | 1599 |
| 19 | они | 1586 |
| 20 | этот | 1546 |
| 21 | такой | 1533 |
| 22 | что Conj | 1482 |
| 23 | вы | 1408 |
| 24 | что Spro | 1370 |
| 25 | нет Part | 1321 |
| 26 | знать V | 1313 |
| 27 | говорить | 1284 |
| 28 | угу | 1275 |
| 29 | всё Spro | 1153 |
| 30 | так Part | 1115 |
| 31 | там Part | 1079 |
| 32 | сейчас | 1036 |
| 33 | э-э | 975 |
| 34 | это Part | 905 |
| 35 | но Conj | 871 |
| 36 | надо Adv | 867 |
| 37 | ещё Adv | 835 |
| 38 | так Advpro | 817 |
| 39 | уже Adv | 799 |
| 40 | просто Part | 766 |
| 41 | же Part | 762 |
| 42 | вообще | 704 |
| 43 | как Advpro | 689 |
| 44 | по | 677 |
| 45 |  -то | 662 |
| 46 | тоже Adv | 594 |
| 47 | сказать | 589 |
| 48 | а Intj | 565 |
| 49 | какой | 563 |
| 50 | весь Apro | 546 |
| 51 | здесь | 532 |
| 52 | мочь V | 523 |
| 53 | два | 520 |
| 54 | потом | 514 |
| 55 | хотеть | 501 |
| 56 | когда | 480 |
| 57-58 | если | 476 |
| 57-58 | как бы Part | 476 |
| 59 | к | 470 |
| 60 | очень | 465 |
| 61 | значит Adv | 463 |
| 62 | как Conj | 459 |
| 63 | тут Advpro | 454 |
| 64 | э Intj | 433 |
| 65 | блядь Intj | 432 |
| 66 | ой | 417 |
| 67 | давай | 410 |
| 68-69 | или Conj | 409 |
| 68-69 | пойти | 409 |
| 70 | делать | 408 |
| 71 | потому что | 403 |
| 72 | кто | 401 |
| 73 | где | 400 |
| 74 | который | 399 |
| 75-77 | бы | 394 |
| 75-77 | за Pr | 394 |
| 75-77 | можно | 394 |
| 78 | что-то Spro | 378 |
| 79 | самый | 368 |
| 80-81 | ага Part | 367 |
| 80-81 | какой-то | 367 |
| 82 | смотреть | 363 |
| 83 | то есть Conj | 360 |
| 84 | человек | 359 |
| 85-86 | думать | 352 |
| 85-86 | мой | 352 |
| 87-88 | понимать | 351 |
| 87-88 | то Spro | 351 |
| 89-90 | год | 342 |
| 89-90 | идти | 342 |
| 91 | сделать | 337 |
| 92 | работать | 335 |
| 93 | видеть | 317 |
| 94 | только Part | 315 |
| 95 | один Anum | 302 |
| 96 | чтоб Conj | 301 |
| 97-98 | три | 297 |
| 97-98 | чтобы Conj | 297 |
| 99 | дать | 286 |
| 100 | даже Part | 285 |

1. *П. М. Алексеев*. Частотные словари: Учебное пособие. СПб.: Изд-во С.-Петерб. ун-та, 2001. С. 22. [↑](#footnote-ref-2)
2. *Алексеев П. М.* Статистическая лексикография (типология, составление и применение частотных словарей) [Текст]: Учеб. пособие. Л.: ЛГПИ, 1975 С. 32. [↑](#footnote-ref-3)
3. *Штейнфельдт Э. А*. Частотный словарь современного русского литературного языка: справочник для преподавателей русского языка / под ред. В. А. Ицковича. М.: Прогресс, 1973. 228 с. [1-е изд. в 1963 г.]. [↑](#footnote-ref-4)
4. Частотный словарь русского языка [Электронный ресурс] / под ред. Л. Н. Засориной. М.: Изд-во «Русский язык», 1977. URL: http://project.phil.spbu.ru/lib/data/slovari/zasorina/zasorina.html (дата доступа: 20.05.2016). [↑](#footnote-ref-5)
5. *Леннгрен Л.* Частотный словарь современного русского языка (Lönngren, Lennart. The Frequency Dictionary of Modern Russian). Acta Univ. Ups., Studia Slavica Upsaliensia Uppsala 32. Uppsala, 1993. [↑](#footnote-ref-6)
6. Статистика речи. Сб. / Отв. ред. Р. Г. Пиотровский. Л.: Наука, 1968. 260 с. [↑](#footnote-ref-7)
7. *Алексеев П. М*. Частотные словари и приемы их составления // Статистика речи. Л.: Наука, 1968. С. 62-63. [↑](#footnote-ref-8)
8. *Ляшевская О. Н., Шаров С. А.* Новый частотный словарь русской [Электронный ресурс]. URL: http://dict.ruslang.ru/freq.php (дата доступа: 20.05.2016) [↑](#footnote-ref-9)
9. Национальный корпус русского языка [сайт]. URL: http://www.ruscorpora.ru/. [↑](#footnote-ref-10)
10. Частотный словарь словоформ русского языка [Электронный ресурс]. URL: http://www.narusco.ru/STAT004/ (дата обращения: 11.04.2016). [↑](#footnote-ref-11)
11. *Леннгрен Л*. Частотный словарь современного русского языка. [↑](#footnote-ref-12)
12. *Мартинович Г.А*. Частотный тематический словарь «Городской общественный транспорт». СПб, Филологический факультет СПбГУ, 2005. 521. [↑](#footnote-ref-13)
13. *Караулов Ю.Н.* Частотный словарь семантических множителей русского языка. М.: Наука, 1980. 207 с. [↑](#footnote-ref-14)
14. *Кривенко Б. В.* Частотный словарь языка массовой коммуникации. Воронеж : Изд-во Воронеж. ун-та, 1992. 218 с. [↑](#footnote-ref-15)
15. *Шайкевич А. Я., Андрющенко В. М., Ребецкая Н. А.* Статистический словарь языка русской газеты (1990-е годы). Т. 1. М.: Языки славянской культуры, 2008. 580 с. [↑](#footnote-ref-16)
16. Частотный словарь рассказов А.П.Чехова / Автор-сост. А.О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: Изд-во С.-Петерб. ун-та, 1999. 172 с. [↑](#footnote-ref-17)
17. Частотный словарь рассказов Л. Н. Андреева / Автор-сост. А.О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: Изд-во С.-Петерб. ун-та, 2003. 398 с. [↑](#footnote-ref-18)
18. Частотный словарь рассказов И. А. Бунина / Автор-сост. А.О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: Изд-во С.-Петерб. ун-та, 2012. 296 с. [↑](#footnote-ref-19)
19. Частотный словарь рассказов А. И. Куприна / Автор-сост. А.О. Гребенников; под ред. Г. Я. Мартыненко. СПб.: Изд-во С.-Петерб. ун-та, 2012. 552 с. [↑](#footnote-ref-20)
20. Частотный грамматико-семантический словарь языка художественных произведений А. П. Чехова [Электронный ресурс]. URL: http://www.philol.msu.ru/~lex/chehov.html (дата доступа: 20.05.2016). [↑](#footnote-ref-21)
21. Русская авторская лексикография XIX-XX веков. Антология [Элекстронный ресурс]. URL: http://www.slovari.ru/default.aspx?p=5309 (дата доступа: 5.04.2016). [↑](#footnote-ref-22)
22. *Шайкевич А. Я.* Дифференциальные частотные словари и изучение языка Достоевского (на примере романа «Идиот» // Слово Достоевского. Сб. ст. М.: ИРЯ РАН, 1996. С. 195-253. [↑](#footnote-ref-23)
23. *Овсиенко Ю. Г.* Частотно – семантический словарь русской разговорной речи [Электронный ресурс] . URL: http://www.euralex.org/elx\_proceedings/Euralex1992\_1/017\_J.%20G.%20Ovsienko%20-Castotno-semanticeskij%20slovar%20russkoj%20razgovornoj%20reci.pdf (дата доступа: 13.03.2016) [↑](#footnote-ref-24)
24. *Козырев В. А., Черняк В. Д.* Лексикография русского языка: век нынешний и век минувший: монография. 2-е изд., испр. и доп. СПб.: Изд-во РГПУ им. А. И. Герцена, 2015. С. 367-369. [↑](#footnote-ref-25)
25. *Лаптева О. А*. Разговорная речь [Электронный ресурс] // Лингвистический энциклопедический словарь / Гл. ред. В. Н. Ярцева. М.: Большая рос. энцикл., 2002. 707, [2] с. URL: http://tapemark.narod.ru/les/407c.html (дата обращения: 11.05.2016 .) [↑](#footnote-ref-26)
26. *Скребнев Ю. М.*Введение в коллоквиалистику / под ред.Сиротининой О. Б. Саратов: Изд-во Сарат.ун-та, 1985. 210 с. [↑](#footnote-ref-27)
27. *Земская Е. А., Китайгородская М. В., Ширяев Е. Н.* Русская разговорная речь. Общие вопросы. Словообразование. Синтаксис. М.: Наука, 1981 [↑](#footnote-ref-28)
28. *Инфантова Г. Г.* Очерки по синтаксису современной русской разговорной речи. Ростов н/Д: Изд-во Рост. гос. пед. ин-та, 1973. 135 с. [↑](#footnote-ref-29)
29. *Сиротинина О.Б.*Современная разговорная речь и ее особенности. М.: Знание, 1974. 260 с. [↑](#footnote-ref-30)
30. *Байкулова А. Н.* Неофициальное общение и его разновидности: критерии выделения и реальное функционирование. Саратов: Изд. центр «Наука», 2012. С. 26. [↑](#footnote-ref-31)
31. *Байкулова А. Н.* Устное неофициальное общение и его разновидности: повседневная речь горожан. Саратов: Изд. центр «Наука», 2014. С. 7. [↑](#footnote-ref-32)
32. *Харченко В. К.* Современная повседневная речь. Изд-е 3-е. М.: Книжный дом «ЛИБРОКОМ», 2012. С. 15-23. [↑](#footnote-ref-33)
33. *Харченко В. К.* Современная повседневная речь. С. 25. [↑](#footnote-ref-34)
34. *Земская Е. А., Китайгородская М. В., Ширяев Е. Н.* Русская разговорная речь. С. 12. [↑](#footnote-ref-35)
35. *Лаптева О. А.* Разговорная речь. [↑](#footnote-ref-36)
36. *Харченко В. К.* Современная повседневная речь. С. 56. [↑](#footnote-ref-37)
37. *Земская Е. А., Китайгородская М. В., Ширяев Е. Н.* Русская разговорная речь. С. 6. [↑](#footnote-ref-38)
38. Там же. [↑](#footnote-ref-39)
39. *Панов М. В.* О развитии русского языка в советском обществе // Вопр. языкознания. 1962, № 3. [↑](#footnote-ref-40)
40. *Асиновский А. С., Богданова Н. В., Русакова М. В. и др.* Звуковой корпус русского языка повседневного общения «Один речевой день»: концепция и состояние формирования // Компьютерная лингвистика и интеллектуальные технологии. Вып. 7 (14): По матер. межд. конф. «Диалог-2008». М., 2008. С. 489. [↑](#footnote-ref-41)
41. *Burnard L.* (ed.). Reference guide for the British National Corpus (XML edition). Published for the British National Corpus Consortium by Oxford University Computing Services. URL: http://www.natcorp.ox.ac.uk/docs/URG/ (дата обращения 15.06.2016). [↑](#footnote-ref-42)
42. *Campbell N.* Speech & Expression; the Value of a Longitudinal Corpus, LREC-04. Lisbon, 2004. pp. 183-186. [↑](#footnote-ref-43)
43. *Bogdanova-Beglarian N., Sherstinova T., Martynenko G*. The “One Day of Speech” Corpus: Phonetic and Syntactic Studies of Everyday Spoken Russian, Proc. 18th Int. Conf “Speech and Computer” (SPECOM-2015), LNAI, vol. 9319, Springer, Switzerland, 2015. pp. 429-437. [↑](#footnote-ref-44)
44. *Богданова Н. В., Степанова С. Б., Шерстинова Т. Ю.* Звуковой корпус русского языка: новый подход к исследованию речи // Корпусная лингвистика - 2011: Тр. Междунар. конф. СПб., 2011. С. 99. [↑](#footnote-ref-45)
45. Звуковой корпус как материал для анализа русской речи. Часть 2. Теоретические и практические аспекты анализа (Том 2. Звуковой корпус как материал для новых лексикографических проектов). Коллективная монография / Отв. ред. Н. В. Богданова-Бегларян. СПб.: Филологический факультет СПбГУ, 2015 С. 11. [↑](#footnote-ref-46)
46. *Шерстинова Т. Ю.* Коммуникативные макроэпизоды в корпусе повседневной русской речи «Один речевой день»: принципы аннотирования и результаты статистической обработки // Корпусная лингвистика - 2013: Тр. Междунар. конф. СПб., 2013. С. 449-456. [↑](#footnote-ref-47)
47. Там же. [↑](#footnote-ref-48)
48. Работа с программой ELAN [Электронный ресурс]. URL: http://www.mpi.nl/tools/elan/tp/how-to/ELAN\_handout\_Russian.pdf (дата доступа: 26.04.2016). [↑](#footnote-ref-49)
49. *Шерстинова Т. Ю., Степанова С. Б., Рыко А. И.* Система аннотирования в звуковом корпусе русского языка «Один речевой день» // Мат-лы XXXVIII международной конференции. Секция: «Формальные методы анализа русской речи». Март 2009. СПбГУ: СПб. С. 66-75. [↑](#footnote-ref-50)
50. Социально-демографический портрет России: По итогам Всероссийской переписи населения 2010 года / Федер. служба гос. статистики. М.: ИИЦ «Статистика России», 2012. 183 с. [↑](#footnote-ref-51)
51. *Баева Е. М*. О способах социолингвистической балансировки устного корпуса (на примере «Одного речевого дня») // Вестник Пермского университета. Российская и зарубежная филология, Вып. 4 (28), 2014. С. 50-51. [↑](#footnote-ref-52)
52. *Баева Е. М.* О способах социолингвистической балансировки устного корпуса (на примере «Одного речевого дня»). С. 51.  [↑](#footnote-ref-53)
53. Там же. С. 52. [↑](#footnote-ref-54)
54. *Богданова Н. В.* Живые фонетические процессы русской речи: Учеб.-метод. пособ. по современному русскому литературному языку. СПб.: Филологический факультет СПбГУ, 2001. С. 58. [↑](#footnote-ref-55)
55. Звуковой корпус как материал для анализа русской речи. Часть 2. С. 86. [↑](#footnote-ref-56)
56. Звуковой корпус как материал для анализа русской речи. Часть 2.  [↑](#footnote-ref-57)
57. *Ермолова О. Б.* «Один речевой день» говорящего с точки зрения прагматики // Вестник Пермского университета. Российская и зарубежная филология, Вып. 3 (27), 2014. С. 21-30. [↑](#footnote-ref-58)
58. *Шерстинова Т. Ю*. Наиболее употребительные слова повседневной русской речи (в гендерном аспекте и в зависимости от условий коммуникации) // Компьютерная лингвистика и интеллектуальные технологии: По матер. ежегодн. межд. конф. «Диалог» (Москва, 1-4 июня 2016). Вып. 15 (22). М.: Изд-во РГГУ, 2016. С. 616-632. [↑](#footnote-ref-59)
59. *Богданова-Бегларян Н. В., Шерстинова Т. Ю., Блинова О. В. и др.* Звуковой корпус русского языка как база для социолингвистических исследований // Компьютерная лингвистика и интеллектуальные технологии. Материалы конференции, 2015 [Электронный ресурс]. URL: http://www.dialog-21.ru/digests/dialog2015/materials/pdf/Bogdanova-BeglarianNVetal.pdf (дата доступа: 26.04.2016). [↑](#footnote-ref-60)
60. *Зобнина Е. А.* Перспективы использования звукового корпуса «один речевой день» в преподавании русского языка как иностранного [Текст] / Е. А. Зобнина // Мир русского слова. 2009, № 4. С. 99-109. [↑](#footnote-ref-61)
61. *Мартыненко Г. Я.* Основы стилеметрии. Л.: Изд-во Ленингр. ун-та, 1988. С. 19 [↑](#footnote-ref-62)
62. *Рогожникова Р. П.* Толковый словарь сочетаний, эквивалентных слову: Ок. 1500 устойчивых сочетаний рус. яз. / Р. П. Рогожникова. М.: Астрель: АСТ, 2003. 416 с. [↑](#footnote-ref-63)
63. *Ефремова Т. Ф.* Толковый словарь служебных частей речи русского языка [Текст] : ок. 15000 слов. статей. Ок. 22000 семантических единиц / Т. Ф. Ефремова. 2-е изд., испр.. М. : АСТ : Астрель, 2004. 815 с.  [↑](#footnote-ref-64)
64. Корпусной словарь неоднословных лексических единиц (оборотов) [сайт]. URL: http://ruscorpora.ru/obgrams.html. [↑](#footnote-ref-65)
65. Словарь русского языка: В 4-х т. / РАН, Ин-т лингвистич. исследований; Под ред. А. П. Евгеньевой; 4-е изд., стер. М.: Рус. яз.; Полиграфресурсы, 1999. [↑](#footnote-ref-66)
66. Частотный словарь словоформ русского языка [Электронный ресурс]. URL: http://www.narusco.ru/STAT004/ (дата обращения: 11.04.2016). [↑](#footnote-ref-67)
67. *Мартыненко Г. Я., Мартинович Г. А*. Многопараметрический статистический анализ результатов ассоциативного эксперимента. СПб.: Изд-во С.-Петербургского ун-та, 2003. С. 6. [↑](#footnote-ref-68)
68. *Мартыненко Г. Я., Мартинович Г. А.* Многопараметрический статистический анализ результатов ассоциативного эксперимента. С. 7. [↑](#footnote-ref-69)
69. *Алексеев П. М.* Частотные словари: Учебное пособие. СПб.: Изд-во С.-Петерб. ун-та, 2001. С. 10. [↑](#footnote-ref-70)
70. *Шерстинова Т. Ю.* Наиболее употребительные слова повседневной русской речи (в гендерном аспекте и в зависимости от условий коммуникации) [↑](#footnote-ref-71)
71. *Мартыненко Г. Я.* Математика гармонии и статистика [Электронный ресурс]. URL: http://www.trinitas.ru/rus/doc/0232/009a/02321149.htm (дата доступа: 30.05.2016). [↑](#footnote-ref-72)
72. *Eysenck H. J*. An Experimental Study of the Good Gestalt. In: Psychological Rewiew. 1942. № 49. Pp. 344-364. [↑](#footnote-ref-73)
73. Введение // Частотный словарь словоформ русского языка [Электронный ресурс]. URL: http://www.narusco.ru/STAT004/ (дата обращения: 11.04.2016). [↑](#footnote-ref-74)
74. *Мартыненко Г. Я., Мартинович Г. А*. Многопараметрический статистический анализ результатов ассоциативного эксперимента. [↑](#footnote-ref-75)
75. *Мартыненко Г. Я.* Основы стилеметрии. [↑](#footnote-ref-76)
76. *Мартыненко Г. Я., Фомин С. В.* Ранговые моменты // Научно-техническая информация. Сер. 2 1989. № 5. С. 23-29. [↑](#footnote-ref-77)
77. *Мартыненко Г. Я., Мартинович Г. А*. Многопараметрический статистический анализ результатов ассоциативного эксперимента. [↑](#footnote-ref-78)
78. *Hirsch J. E.* An index to quantify an individual's scientific research output [Элекстронный ресурс]. URL: ttp://www.pnas.org/content/102/46/16569.full (дата доступа: 15.06.2016). [↑](#footnote-ref-79)
79. Список CItot ≥ 1000 [сайт]. URL: http://www.expertcorps.ru/science/whoiswho/ci86?sortby=h. [↑](#footnote-ref-80)
80. *Мартыненко Г. Я.*Золотое сечение формулы изобретения [Текст]: [Автоматизация] // Научно-техническая информация. Сер. 2. Информационные процессы и системы. 2002. N10. С. 23. [↑](#footnote-ref-81)
81. *Гринбаум О.Н.* Гармония строфического ритма в эстетико-формальном измерении (на материале «Онегинской строфы» и русского сонета). СПб.: Изд-во СПбГУ, 2000.160 с. [↑](#footnote-ref-82)
82. *Васютинский Н.А*. Золотая гармония. М.: Молодая гвардия, 1990. 238 с. [↑](#footnote-ref-83)
83. *Мартыненко Г. Я.* Основы стилеметрии. С. 80. [↑](#footnote-ref-84)
84. *Венецкий И. Г., Венецкая В. И*. Основные математико-статистические понятия и формулы в экономическом анализе. Справочник. 2-е изд., перераб. и доп. М.: Статистика, 1979. С.77. [↑](#footnote-ref-85)
85. *Герчук Я. П.* Графические методы в статистике. М.: Статистика, 1968. С. 20 [↑](#footnote-ref-86)
86. *Ежов А. И.* Выравнивание и вычисление рядов распределений. М.: Госстатиздат, 1961. С. 15. [↑](#footnote-ref-87)
87. Анализ временных рядов // StatSoft, Inc. (2012). Электронный учебник по статистике. Москва, StatSoft [Электронный ресурс]. URL: http://www.statsoft.ru/home/textbook/default.htm (дата обращения: 16.06.2016). [↑](#footnote-ref-88)
88. *Герчук Я. П*. Графические методы в статистике. М.: Статистика, 1968. 140 [↑](#footnote-ref-89)
89. *Грешилов А. А., Стакун В. А., Стакун А. А*. Математические методы построения прогнозов. М.: Радио и связь, 1997. С. 74. [↑](#footnote-ref-90)
90. *Венецкий И. Г., Венецкая В. И.* Основные математико-статистические понятия и формулы в экономическом анализе. Справочник. 2-е изд., перераб. и доп. М.: Статистика, 1979. С.86. [↑](#footnote-ref-91)
91. *Венецкий И. Г., Венецкая В. И*. Основные математико-статистические понятия и формулы в экономическом анализе. Справочник. 2-е изд., перераб. и доп. М.: Статистика, 1979. С.86. [↑](#footnote-ref-92)
92. Статистика: Курс лекций / Харченко Л. П., Долженкова В. Г., Ионин В. Г. и др.; под ред. канд. экон. наук В. Г. Ионина. Новосибирск: Изд-во НГАЭиУ; М.: ИНФРА-М, 2000. С. 102. [↑](#footnote-ref-93)
93. *Четвериков Н. С.* Статистические и стохастические исследования. М.: Госстатиздат, 1963. С. 201. [↑](#footnote-ref-94)