

ЧАСТОТНЫЙ СЛОВАРЬ РУССКОГО ЯЗЫКА (под редакцией Л. Н. Засориной)

- Выходные данные
- От составителей
- ВВЕДЕНИЕ
 1. Назначение словаря
 2. Функциональные сферы речи и источники словаря
 3. Формирование словника
 - Единицы текста и элементы словаря
 - Аналитическая грамматика словаря
 - Правила анализа существительных
 - Нарисательные имена
 - Собственные имена
 - Правила анализа прилагательных
 - Правила анализа числительных
 - Правила анализа местоимений
 - Правила анализа глаголов
 - Правила анализа наречий
 - Правила анализа предлогов
 - Правила анализа союзов
 - Правила анализа частиц
 - Правила анализа междометий
 4. Техника обработки материала
 5. Источники словаря
 - Публицистические и научные тексты
 - Художественная проза
 - Драматургия
 - Журнальные и газетные тексты

Выходные данные

Словарь составили: В. А. АГРАЕВ В. В. БОРОДИН Л. Н. ЗАСОРИНА В. М. МУРАТОВА Э. В. ТИСЕНКО

Специальный научный редактор М. И. Привалова

Словарь представляет собой свод статистических данных о лексическом составе современного русского языка. Словарь составлен на основании обработки средствами вычислительной техники одного миллиона словоупотреблений, что дало около 40 тыс. единиц словаря.

Разнообразные статистические характеристики единиц словаря создают основу для изучения статистической структуры словарного состава русского языка, для определения границ основной лексики и других проблем общей и учебной лексикографии.

Словарь рассчитан на специалистов-филологов, он может быть использован и при составлении различных словарей грамматик и учебников русского языка.

Редакция русского языка

- Зав. редакцией В. В. Пчелкина
- Ведущий редактор Б. А. Поворотик
- Художественный редактор Е. В. Ноское
- Переплет художника А. Г. Кузькина
- Технический редактор Е. А. Сиротинская
- Корректоры: В. Н. Ермакова, Т. О. Тарасова

© Издательство «Русский язык», 1977

Памяти Бориса Александровича Ларина

От составителей

Настоящий словарь составлен коллективом сотрудников филологического факультета Ленинградского государственного университета им. А. А. Жданова и Лаборатории семиотики Научно-исследовательского института прикладной математики и кибернетики (НИИ ПМК) при Горьковском государственном университете им. Н. И. Лобачевского. Принципы и проект словаря получили одобрение со стороны Словарного сектора ЛО ИРЯЗ АН СССР и чл.-кор. АН СССР Ф. П. Филина.

Инициатором и руководителем работы является Л. Н. Засорина, ею же разработаны теоретические основы и практическая инструкция обработки лексического материала.

Проект словаря, аналитическая грамматика и «Введение» написаны Л. Н. Засориной. Обработка материалов словаря на ЭВМ проведена в Лаборатории семиотики НИИ ПМК под руководством В. А. Аграева и В. В. Бородина. Программы для подсчета словника на ЭВМ составлены В. В. Бородиным и В. М. Муратовой.

Перфорация текстов и отдельных словников осуществлена сотрудниками НИИ ПМК: А. Н. Елинкиной, С. В. Киселевой, В. Н. Некоркиной, Н. А. Рахманиной и Л. В. Харьковской. Проверка перфорированного текста проводилась Н. А. Гордеевой, С. М. Козокиной и В. В. Коноваловой. Вспомогательные программы ввода, распечатки и корректировки данных были составлены В. В. Бородиным, В. М. Муратовой и А. Н. Елинкиной. Контроль за вводом и выводом данных в ЭВМ осуществлен В. М. Муратовой.

Раздел «Статистические параметры словаря» написан Э. В. Тисенко.

Для подготовительной работы над источниками словаря были привлечены студенты, аспиранты и сотрудники кафедр математической лингвистики и русского языка филологического факультета ЛГУ (всего около 100 человек). Наиболее активно участвовали в расписывании текстов и перфорации студенты: В. Чернышев, О. Воробьева, Л. Парфенова, А. Тенегин, В. Тихомирова, Л. Мелешко, Л. Пименова, Л. Шишкина, В. Захаров, Л. Кулеш, Н. Михайлова, Т. Попова, В. Халябина.

Большую работу по подготовке текста для машинной обработки провели сотрудники и аспиранты филологического факультета ЛГУ: Л. М. Акуненко, Л. Н. Беляева, Е. А. Белоусова, А. С. Герд, А. Н. Еренкова, Л. Н. Иванова, Н. А. Коваль, Л. Ф. Клеопатрова, А. М. Лятина, М. И. Привалова, Л. Н. Смирнова, Г. Е. Спринчак, Е. Е. Талицкая, Э. В. Тисенко.

Особо должна быть отмечена работа на счетно-аналитических машинах лаборанта Е. С. Андреевой и инженеров Г. Г. Григорьева и А. С. Рахманова. Большую помощь в организации работ на машинах оказали директор Вычислительного центра ЛГУ Г. П. Самосюк и директор НИИ ПМК ГТУ Я. К. Любимцев.

В подготовке рукописи к печати участвовали сотрудники филологического факультета ЛГУ: С. А. Аверина, Л. М. Акуленко, Е. С. Андреева, С. И. Василькова, Е. М. Зайцева, А. В. Лемов, Н. Н. Майорова, В. В. Образцова, И. Д. Самарина, Л. В. Шашенкова, которым составители словаря выражают свою благодарность.

Редакторы словаря выражают искреннюю признательность рецензентам: д-ру филол. наук А. М. Бабкину, канд. филол. наук, доц. А. Я. Шайкевичу, канд. филол. наук В. М. Андрищенко и канд. филол. наук Л. К. Граудиной.

Составители словаря отдают себе отчет в том, что в первой работе такого объема им не удалось удовлетворительным образом решить все возникшие задачи. Поэтому они с признательностью примут все замечания и пожелания. Просьба направлять их по адресу: 103009, Москва, К-9, Пушкинская ул., 23, издательство «Русский язык», или: 199164, Ленинград, Университетская наб., 11, Филологический факультет ЛГУ.

ВВЕДЕНИЕ

1. НАЗНАЧЕНИЕ СЛОВАРЯ

В последние годы не ослабевает интерес к частотным словарям, актуальность которых для разных прикладных и исследовательских задач достаточно широко известна [*1]. На базе частотных словарей формируется особая ветвь статистической лингвистики — лексикостатистика, предмет изучения которой составляет лингвистическая и статистическая интерпретация распределений слов в языке и тексте. Однако до настоящего времени материалы лексикостатистики остаются скудными. Так, до сих пор нет ни одного большого частотного словаря русского языка, объем выборки которого превышал бы 400—500 тыс. словоупотреблений [*2]. Значительное расширение базы лексикостатистики возможно только за счет автоматизации обработки текстов. Вместе с тем большой диапазон приложений частотных словарей требует специализации каждого из них.

Частотные словари русского языка создавались на стыке лексикографии, статистики и автоматизации. Они различаются в зависимости от преобладания в них одной из этих сторон.

В настоящем словаре статистика и автоматизация стоят на службе интересов русской лексикографии. Это обстоятельство определяет и отбор источников словаря, и выбор единицы счета, и методы обработки текста.

В настоящее время остается открытой одна из чрезвычайно важных задач лексикологии — выяснение границ живой лексической системы современного русского языка. Мы не имеем отчетливых представлений об общей системе лексического состава языка, хотя предположение о наличии в ней общеупотребительного и периферического слоя само по себе кажется очевидным [*3].

«Частотный словарь русского языка» составлялся в первую очередь для определения границ активного словарного состава. Одной из важных целей словаря является представление живого словоупотребления образованного человека наших дней [*4].

Количественный объем лексики современного русского языка может оцениваться только с известным приближением, так в Словаре современного русского литературного языка в 17-ти томах — 150000 лексем. Очевидно, что словарный состав языка в принципе является открытой системой, в связи с чем возникает задача более полной его инвентаризации и систематизации.

Вопросы систематизации словарного состава связаны с разграничением в нем активного и пассивного запаса, что в свою очередь требует дифференцированного подхода к устной и письменной речи, возрастным, профессиональным и образовательным категориям носителей языка.

Основное назначение словаря — дать достаточно полные сведения о современной лексике с учетом жанровой ее дифференциации. Исходя из наличия основных функциональных сфер речи, для данного словаря были определены четыре группы текстов-источников: художественная проза, драматургия, научные и публицистические тексты, газетные и журнальные тексты (см. подробнее в разд. 2). Полученные данные составят базу для последующих извлечений и интерпретаций как собственно лингвистических, так и лексикостатистических.

По первичным статистическим характеристикам можно определить с заданной относительной погрешностью ту часть словника, в которую входят слова с высокой частотой появления независимо от типа текста. Возможно также, введя ступенчатое упорядочение в словарь, получить серию словников, охватывающих первые 100, 1000, 5000 и т. д. частых слов. Статистические характеристики словаря вызывают интерес в связи со смысловым анализом лексики. Изучение предметно-идеологических групп и семантических полей показывает, что лексические объединения поддерживаются семантическими связями, которые концентрируются вокруг лексем с наиболее общим значением. Описание значений в пределах лексико-семантического поля может проводиться посредством идентификации слов с наиболее абстрактными по смыслу лексемами. По-видимому, «пустые» (с точки зрения номинативных потенций) единицы словаря, составляют статистически однородный пласт, выявление границ этого слоя может оказаться полезным при выборе метаязыковых формулировок для толкования значений.

Не меньшую ценность имеют и словники по отдельным жанрам. Изучение меры их сходства и характера статистических распределений даст интересные сведения о качественном расслоении лексики в зависимости от сферы речеупотребления.

Задачи систематизации лексики, а также определения ее базы и периферии до сих пор решались на основе частотных словарей малого и среднего объема. Возможности собственно статистического подхода к этой проблеме пока неясны. В частности это связано с исключительной трудоемкостью статистической обработки материала.

Составление большого частотного словаря («миллионного») потребовало обращения к вычислительной технике. Одна часть выборки (газетно-журнальные и научно-публицистические тексты) обшчитывалась на счетно-аналитических машинах, вторая часть и общий итоговый анализ — на ЭВМ. Введение частичной механизации и автоматизации в процесс работы над словарем представляет интерес как эксперимент машинной обработки словников к разным текстам. Такой словарь требует более строгой системы обработки и накопления словарного материала. Реализация системы на ЭВМ предъявляет требования к единообразному фиксации исходных объектов, средствам их записи и к программе операций по переработке информации. В миниатюре это информационно-поисковая система, которая способна выдавать сведения о различных сторонах текста и словаря. Некоторые основные запросы к этой системе планируются с самого начала: это — общее количество инвентаризованных слов, статистические характеристики отдельного слова и целых словников, упорядочение частых и редких зон словника и т. п. Машинная картотека позволяет автоматически строить обратные словари по отдельным жанрам

и источникам. Множество других полезных статистических сведений о языке будет извлечено из накопленного массива информации. Словарь с элементами автоматизации создает экспериментальную базу для перехода к более обширной автоматизации словарных работ.

Статистические данные «Частотного словаря русского языка» могут быть широко использованы и при решении других лингвистических задач. Так, обратный статистический словарь даст материалы для анализа словообразования и определения активных средств словообразования современного русского языка.

Укажем еще несколько возможных применений Частотного словаря. Вопросы усовершенствования графики и орфографии связаны с учетом статистических сведений о словарном составе. При этом важно учитывать вероятностные характеристики комбинаций графем, реализованные в словах типы буквосочетаний и т. п.

Не менее актуальны данные Частотного словаря и для практической транскрипции и транслитерации. Статистические параметры словаря будут полезны и при решении вопросов автоматизации печатного дела, распознавания и автоматического чтения буквенного текста.

2. ФУНКЦИОНАЛЬНЫЕ СФЕРЫ РЕЧИ И ИСТОЧНИКИ СЛОВАРЯ

Работа над частотным словарем выдвигает особые требования к отбору источников. Объем выборки, как бы он ни был велик, оказывается ограниченным. В связи с этим должны решаться вопросы о жанровых, хронологических и авторских границах текстов-источников. Составители опубликованных частотных словарей руководствовались различными системами жанров. Частотные словари русского языка созданы преимущественно по текстам художественной литературы; смотри соотношение текстов в словарях Г. Г. Йоссельсона и Э. А. Штейнфельдта (табл. 1), а также словарь Н. П. Вакара, в выборку которого включены только пьесы [5].

Таблица 1

Словари	Тексты	в %% ко всей выборке
Йоссельсон (выборка 1 млн.)	• художественная проза	59
	• драматургия	7
	• журналистика	20
	• литературная критика	14
Штейнфельдт (выборка 400 тыс.)	• детская литература	25
	• художественная проза	12,5
	• драматургия	12,5
	• радиопередачи для молодежи	25
	• публицистические статьи (из газет и журналов)	25

Данный частотный словарь впервые широко охватывает научно-публицистическую и деловую речь, бытующую в сфере массовой коммуникации. Схема функционально-речевых сфер для словаря разработана известным лексикологом, академиком Б. А. Лариным. По его мнению, современные нормы русского языка складываются под влиянием по крайней мере четырех самостоятельных сфер употребления лексики: художественной литературы, разговорной речи, научной и публицистической, а также деловой речи. Исходя из этого установлены четыре равновесные группы текстов: художественная проза, драматургия (как отображение разговорной речи), научно-публицистические и газетно-журнальные тексты. Каждой группе отведена $\frac{1}{4}$ часть выборки, около 250 тыс. словоупотреблений. Итоговое соотношение между текстами показано в табл. 2.

Таблица 2

Функционально-речевые сферы	в %% ко всей выборке
• художественная проза	25,4
• драматургия	27,2
• научные тексты и публицистические тексты	23,6
• газетные тексты и журнальные тексты	23,8

Современные толковые словари и грамматики русского языка в основном построены на базе литературно-художественных текстов. В последние годы все более отчетливо осознается противопоставление разговорной и книжной речи. Особенно остро обсуждается этот вопрос среди методистов, которые требуют поворота обучения в сторону разговорного языка. Однако специфика разговорной речи до сих пор остается необъясненной [6].

Не располагая записями разговоров в достаточном количестве, мы использовали в качестве источников разговорной речи современную реалистическую драматургию. Элементы разговорного словоупотребления безусловно проникают также в публицистические тексты, особенно массового назначения, газетные, а также литературно-художественные произведения, чаще—повести и рассказы.

Наряду с языком художественной литературы в современной речи все более растет влияние научного словаря. В картотеку источников включены работы видных советских ученых — С. И. Вавилова, Н. Н. Качалова, Е. В. Тарле, К. А. Тимирязева, А. Е. Ферсмана. Эти труды представляют ведущие отрасли современной науки — физику, химию, биологию, минералогия и историю. Научные тексты составляют половину источников в научно-публицистическом жанре. В состав публицистической подгруппы текстов входят статьи и выступления В. И. Ленина, которые он сам предназначал для широких масс. Несомненно, что работы Ленина оказали и продолжают оказывать влияние на судьбы современной лексики. В состав этой группы источников словаря включены речи и выступления М. И. Калинина о коммунистическом воспитании, а также материалы съездов КПСС, постоянно находящие отражение в нашей прессе. Помимо собственно политических текстов представлены статьи и выступления известных советских писателей-публицистов периода Великой Отечественной войны — А. Н. Толстого, Л. М. Леонова и М. А. Шолохова. Из послевоенной публицистики выбраны статьи Н. М. Грибачева.

О газетно-журнальном жанре следует сказать особо. При ограниченном объеме выборки (на газетные тексты отводится 125 000 словоупотреблений) едва ли целесообразно просматривать газеты случайным образом на протяжении нескольких десятилетий. Перечень газет ограничивается несколькими изданиями общесоюзного значения, наиболее авторитетными и широкими по охвату тем: «Правда», «Известия», «Советская Россия» и др. Из этих изданий отобраны номера за 5 января 1968 года. Каждый номер газеты расписывался полностью, включая объявления, рекламу и т. д. Газетные тексты отличаются разно- и многотемностью. Следуя основной цели словаря — выявить общие лексические нормы русского языка — мы сочли возможным не разграничивать внутригазетные жанры. Журнальные тексты составляют вторую половину этой группы текстов. Длина выборки из одного журнала примерно равна длине 2-х номеров газет, всего использовано 5 наименований. Отобраны наиболее популярные журналы: «Вокруг света», «Наука и жизнь», «Новое время», «Новый мир», «Юность» (по 1—2 номера за период 1958—1962 гг.). При выборе материала исключались беллетристические очерки во избежание дублирования текстов художественной литературы. Журнальные тексты представляют репортажи на разные темы (научные, политические, экономические).

За пределами словаря остаются: а) стихотворные тексты (словарь лирики специфичен и часто дает отклонения по сравнению с обычной нормой); б) сатирические произведения, в которых исключительную роль играет переосмысление слова, что не может быть учтено в данном частотном словаре; в) тексты точных наук, словарь которых обусловлен спецификой предмета, насыщен научными терминами, элементами символических языков, представляющих самостоятельные знаковые системы в естественном языке.

Принимая изложенные принципы отбора источников, естественно одновременно с этим решать и вопросы хронологии. Частотный словарь представляет фонд лексики русского языка XX века. Таким образом, хронологические рамки словаря обнимают эпоху от произведений Ленина и Горького до 60-х годов.

При составлении частотного словаря необходимо соблюдать количественную однородность выборок в пределах четырех функционально-речевых сфер. Отбирая авторов, названия произведений, мы стремились оценить их с точки зрения наибольшей чистоты их языка. Так, Горький, Толстой, Леонов представлены в драматургии и прозе. Остановившись на автобиографической трилогии Горького, мы имели в виду, что это произведение является связующим звеном между русской классической литературой и советской литературой. В наборе источников не представлены произведения, хотя бы и крупнейших писателей, в которых значительны элементы диалектной лексики, например «Тихий Дон» Шолохова.

Сохранение равнопорционности выборок из конкретных источников оказалось трудно достижимым: специфика произведений, особенно в драматургии и публицистике, во многих случаях определяет размер текста. Примерное количество разных авторов в каждой из 4 функционально-речевых сфер колеблется от 11 до 15: в прозе—11, драматургии—14, научно-публицистических текстах—12 (научные тексты—5, публицистические—7), газетно-журнальных—15 разных изданий.

Каждый текст подвергался сплошному расписыванию: короткие тексты расписывались полностью, для длинных задавалось определенное количество страниц из расчета примерной длины одного источника выборки.

В целом решение вопроса об источниках словаря опиралось преимущественно не на статистические критерии. В числе главных были функционально-речевые признаки, авторское мастерство, значимость произведения для развития литературной нормы русского языка. Словарь в первую очередь отражает устойчивую часть лексики, общеупотребительную и нейтральную относительно темы, жанра, автора, составляющую общую основу для всех жанров и разновидностей современной речи. Во вторую очередь он представляет интерес для изучения дифференциации лексики по функционально-речевым сферам.

3. ФОРМИРОВАНИЕ СЛОВНИКА

Единицы текста и элементы словаря

Определение тождеств и различий речевых единиц, иначе вопрос об отдельности слова, — кардинальная проблема лексикографии и лингвостатистики.

Текстовые материалы даны составителю в виде множества графических слов. Любой словарь соотносится с текстом по правилам грамматики данного языка. Элементами словаря являются лексемы, слово-типы, обладающие свойством «порождать» текстовые слова (слово-знаки). В лексеме, имплицитно или эксплицитно, заключен грамматический шифр, позволяющий развернуть весь ряд форм слова в тексте. Тот же грамматический механизм,

хранящийся в памяти носителя языка, позволяет сводить любые словоупотребления в тексте к исходным формам слова.

В традиционной лексикографии обычно не фиксируются или указываются неполностью правила переработки текстовых слов в словарные. В предисловиях к словарям уделяется внимание омонимии и заглавной форме представления лексем, системе грамматических помет.

Частотные словари составляются в процессе сплошного подсчета текстовых единиц. Они формируются не выборочно, а в соответствии с появлением в тексте новых слов. Поэтому отождествление графических слов проводится при расписывании текста.

Частотные словари составляются с опорой на различные единицы счета: словоформы, лексемы (с различием или неразличием разных типов омонимов) или словосочетания. Обычно противопоставляют друг другу две методики составления частотных словарей: за единицу словника принимается либо словоформа, либо лексема. Первый способ имеет кажущиеся преимущества перед вторым, он кажется более последовательным. Выбирая в качестве единицы счета словоформу, составитель словаря опирается только на графическую эквивалентность, почти никакого анализа текста не производится. Считается, что переход от словоформ к лексемам можно осуществить легко по общеизвестным правилам. В действительности дело обстоит сложнее. Правила идентификации словоформ далеко не просты, эксплицитно не формулируются (в настоящее время не существует надежных и простых алгоритмов преобразования текстовых слов в лексемы), из-за чего невозможен машинный анализ текста. Кроме того, при большой выборке словаря объем статистических перерасчетов при такой методике был бы так велик, что приблизился бы по затратам времени и труда к обработке материала заново.

В «Частотном словаре русского языка» была выбрана вторая методика получения словника. За элемент словника принимается лексема. Это потребовало кодификации правил сведения графических (текстовых) слов к исходным формам слова. В результате была составлена аналитическая грамматика русского языка, которая может послужить эскизом анализирующей модели переработки сегментов текста в элементы словаря.

Материалом для отбора элементов словника служили тексты, расчлененные на отрезки, букворяды и другие графические символы, разделенные пробелами. В число графических знаков, образующих текстовые слова, входят буквы русского алфавита, дефис и точка, не являющаяся концом предложения (ср.: т. п., т. е.). Все остальные графические знаки не учитывались при обработке текста: знаки препинания, эмфатические знаки, кавычки, скобки, под- и надстрочные знаки, цифровые, выделительные и другие знаки, буквы других алфавитов, а также иностранные слова, написанные некирилловским алфавитом, напр., *summa*. Транслитерированные слова, заимствованные из других языков, отбирались в словник, напр., *нкосаан*, *падаунги* (из зулу). Исключались формулы, графики, чертежи, а также слова бессмысленные, ср. у С. П. Антонова: *татути*, *напотутоте*.

Выделение графических слов в тексте (сегментация его на графические слова) задается правилами орфографии, которые, в известной степени, являются независимыми от фонологических, морфологических, грамматических, а также семантических характеристик системы языка.

Правила отождествления графических слов разрабатывались с опорой на естественное, живое формообразование современного русского языка. При факторизации текста ведущим принципом считалась относительная графическая эквивалентность словоформ, подчиненным — принцип лексико-синтаксической эквивалентности.

Лексическое единство слова в плане выражения обеспечивается разными видами эквивалентностей: фонетической, графической, морфемной (слово- и формообразовательной) и синтаксической (синтагматической). Первые три более формальны и потому наглядно прослеживаются в отдельном текстовом слове. Обычно их учитывают в лексикографической практике. При работе над Частотным словарем, мы ограничились учетом фонетической, графической, морфемной и синтаксической эквивалентности слов, аспект семантического варьирования и эквивалентности не мог быть принят во внимание.

АНАЛИТИЧЕСКАЯ ГРАММАТИКА СЛОВАРЯ

В словнике Частотного словаря лексемы представлены в исходных формах, которые не сопровождаются дополнительными грамматическими пометами. Однако это не означает, что словарь строился без учета грамматики русского языка. Переход от текста к словнику регулировался правилами грамматики, которые зафиксированы отдельно для каждой части речи [*7]. Они применялись при домашней обработке текста.

Правила анализа существительных

Нарицательные имена

1. Исходной формой считается им. пад. ед. ч.: *дерева*→*дерево*, *книгой*→*книга*, *отцу*→*отец*.
2. Звательные формы существительных сводятся к им. пад.: *господи*→*господь*, *друзе*→*друг*, *старче*→*старик*. Употребление этих форм в функции междометия не ведет к обособлению их в отдельную лексему: *боже (мой!)*→*бог*.
3. Супплетивные формы с одной основой сводятся к опорной форме ед. ч.: *армянину*→*армянин*, *армяне*→*армянин*; *хозяева*→*хозяин*; *листья*→*лист*, *листья*→*лист*.

Супплетивные формы от разных основ считаются разными лексемами: *людей*→*люди*, *человека*→*человек*; *дети*→*дети*, *ребенком*→*ребенок*; *лет*→*лета*, *годы*→*год*.

4. Существительные, употребляющиеся только во мн. ч., возводятся к им. пад. мн. ч.: *недрах*→*недра*, *щипцов*→*щипцы*.

5. Существительные от одной основы, имеющие разные формы рода. даются как разные слова: *казака*→*казак*, *казачке*→*казачка*.

Просторечные, диалектные и т. п. варианты существительных, имеющие отклонения в формах рода и падежа, фиксируются отдельно: *проток* и *протока*, *туфель* и *туфля*.

6. Уменьшительные и увеличительные имена считаются самостоятельными лексемами: *ручья*→*ручей*, *ручейка*→*ручеек*; *топора*→*топор*, *топоришки*→*топоришко*.

7. Фонетические варианты слов даются разными единицами при условии имеющих стилистических различий между ними: *волнение* и *волненье*, *поколение* и *поколенье*, *камешек* и *камушек*. Фонетически удлинённые формы с повторами букв считаются текстовыми вариантами исходных форм: *гла-а-зыньки*→*глазыньки*.

8. Сокращённые формы являются вариантами полной формы лексемы, в том числе и буквенные сокращения: *г-жа*→*госпожа*, *г.*→*город*, *им.*→*имя*. Буквенные сокращения — собственные имена (*СССР*, *ЦК КПСС* и т. п.) выпускаются из текста. Также не учитываются буквенные сокращения названий организаций, марок машин и т. п. (*ГЭС*, *МТС*).

Эпизодические, индивидуальные сокращения, значение которых непонятно без текстовых пояснений, также не выписываются, напр., у Антонова: «*Разрешбор*» («*Разрешаю*». *Борисов*).

Общепринятые сокращения (*т. д.*, *т. е.*, *т. п.*) даются в текстовой форме, поскольку они соответствуют сочетанию слов.

9. Сложные существительные с дефисом считаются отдельным словом: *бой-девка*, *мама-слониха*, *тракторист-скоростник* [*8].

10. Существительные с частицами, присоединёнными дефисами, даются как два отдельных слова: *господину-то*→*господин*, *то*, *часы-таки*→*часы*, *таки*.

11. Устойчивые сочетания с существительными расчленяются. Если они содержат формы, не встречающиеся в свободном сочетании, то заглавные формы выводятся искусственно: *во сто крат*→*в*, *сто*, *крат*; *на всем скаку*→*на*, *весь*, *скок*.

12. Омонимия существительных со словами других грамматических классов большей частью не затрагивает их словарных форм. Так, случаи субстантивации прилагательных, причастий, местоимений и числительных регистрируются соответствующими формами ср. и ж. рода ед. ч. или им. пад. мн. ч.: *заливное*, *ванная*, *передовая*, *молодые*, *данные*, *свое*, *первое*, *второе* и т. п. Неразличенной остается омонимия в случаях: а) субстантивации прилагательных только муж. р. ед. ч. (*военный*, *раненый*); б) совпадения именных форм с инфинитивом (*печь*, *сечь*, *течь*); в) совпадения имени и частицы (*мол*, *уж*, *чай*) (см. Приложение 2 — Перечень словарных грамматических омонимов).

Собственные имена

1. Собственные имена (Иванова, Галилей, М. А. Шолохов) не выписываются из текста. Собственные имена лиц, употребляющиеся как нарицательные, написанные со строчной буквы, фиксируются: *кондратий* (паралич), *георгий* (орден).

Если при собственных именах употреблены титулы или названия должностей, то они выписываются: *принц Ольденбургский*→*принц*; *царь Иван*→*царь*; *председатель колхоза Петров*→*председатель*, *колхоз*. Иностранные частицы — титулы (де, фон) при именах (Бодуэн де Куртене) выпускаются вместе с собственными именами. Если названия чинов и должностей написаны с заглавной буквы, то они также выписываются: *Полномочный Посол Великобритании*→*полномочный*, *посол*. Прозвища лиц, совпадающие с соответствующими нарицательными именами, выписываются: *Иван Грозный*→*грозный*. В составных прозвищах, первый компонент которых собственное имя, выписываются только нарицательные слова: *Аника-воин*→*воин*.

2. Не фиксируются клички животных, не имеющие параллелей в нарицательных именах, напр., *Фафик* (но *Шарик*→*шарик*).

Клички-названия мифических существ, в составе которых имеются компоненты, употребляющиеся и как нарицательные имена, выписываются: *Василиса Прекрасная*→*прекрасный*, *Илья-пророк*→*пророк*.

3. Географические названия (названия государств, стран, городов, рек, морей, озер, заливов и т. д.), названия планет, которые не известны в функции нарицательных имен, не фиксируются: напр., *Африка*, *Байкал*, *Марс*, *Москва*. Если в их составе имеются компоненты, которые совпадают с нарицательными именами, то они возводятся к соответствующей заглавной форме: *Великие Луки*→*великий*, *Булонский лес*→*лес*, *Черное море*→*черный*, *море*; *Орел*→*орел*.

Сложносокращённые собственные имена, первый компонент которых собственное имя (*Днепрогэс*, *Кузбассталь*, *Ленсовет*), не фиксируются.

4. Названия статей, книг, изданий и организаций обрабатываются по тем же правилам—фиксируются лишь те компоненты, которые могут встретиться в качестве нарицательного имени, напр.: «*Поднятая целина*»→*поднятый*, *целина*; «*Комсомольская правда*»→ *комсомольский*, *правда*; «*Большевичка*» (фабри-

ка)→большевичка; «Красный треугольник» (завод)→красный, треугольник. Транслитерированные собственные имена (Ассошиейтед пресс) выпускаются из текста.

Правила анализа прилагательных

1. Заглавной формой прилагательных является им. пад. ед. ч. муж. р.: красного→красный, волчья→волчий.

2. Краткие формы прилагательных возводятся к полной: велики→великий, сине→синий (исключение составляют: рад, должен).

3. Притяжательные прилагательные, образованные от собственных имен, с помощью суффиксов -ин, -ов, -ев, даются со строчной буквы: Ньютонов {закон}→ньютонов.

4. Формы степеней сравнения сводятся к форме положительной степени: красивее →красивый, наисильнейшие→сильный, тончайший→тонкий. Супплетивные формы прилагательных считаются разными лексемами: плохой, худший; хороший, лучший. Повторы типа старый-старый возводятся к исходной форме старый.

5. Уменьшительные и увеличительные образования считаются отдельной лексемой: большой, большущий; тихий, тихонький.

6. Прилагательные, парадигма которых состоит из единичных форм, даются в соответствующей текстовой форме. Такие прилагательные с неразвитой парадигмой часто встречаются в устойчивых и номенклатурных наименованиях: глауберова (соль), незапамятные {времена}, подозрная (труба), стельная (корова). Прилагательные многие, немногие представляют образования *pluralia tantum* подобно разряду существительных *pluralia tantum*. Субстантивированные формы многое, немногое представляют самостоятельные лексемы.

7. Фонетические варианты прилагательных записываются как разные слова при наличии смысловой и стилистической дифференциации между ними: развитой и развитый, родной и родный.

8. Сокращенные формы считаются вариантами полных форм: др.→другой, проч.→прочий.

9. Сложные имена прилагательных с дефисом даются как отдельное слово: рабоче-крестьянский, русско-болгарский, серебристо-черный [*9].

10. Прилагательные с частицами, присоединенными дефисом, разделяются на две лексемы: белым-таки→белый, таки; иных-то→иной, то.

Правила анализа числительных

1. Количественные и собирательные числительные записываются в форме им. пад.: пятью→пять, ста→сто, двоих→двое, троим→трое. Числительные, имеющие формы рода, сводятся к муж. р.: одно, одна→один; обеих, обоих→оба.

2. Порядковые числительные, как и прилагательные, сводятся к им. пад. ед. ч. муж. р.: вторые→второй, двухсотого→двухсотый.

3. Неопределенно-количественные числительные (местоименные числительные) записываются в именительном падеже: нескольких→несколько, скольким→сколько, столькими→столько.

4. Числительные дробные представляются как два слова: три пятых→три, пятый, также и составные числительные: двадцать пять→двадцать, пять.

5. Составные числительные с дефисом записываются как одна лексема, при этом возводятся к заглавной форме: двух-трех→два-три, пяти-шести→пять-шесть.

6. Цифро-буквенные варианты числительных исключаются из текста: 10-ый, 17-ая. Также не фиксируются сложные образования типа 50-летие.

7. Числительные с частицами записываются как два отдельных слова: третьего-то→третий, то.

8. Субстантивированные числительные, чаще всего в им. пят. ср. р., представляют отдельные лексемы: (на) первое (суп)→первое.

Правила анализа местоимений

1. Для местоимений, различающих формы рода и числа, заглавной формой является им. пад. ед. ч. муж. р.: наших→наш, твоего→твой. Исключение составляют личные местоимения. По характеру синтаксических связей они подобны существительным и так же, как существительные, фиксируются в соответствующих формах рода и числа: я→я, мы→мы и т. д.

Притяжательные местоимения отличаются заглавной формой от личных: его (книга)→его, но (вижу) его→он.

2. Супплетивные формы возводятся к основе им. пад. ед. ч.: меня→я, ему→он, те→тот, чего→что.

3. Краткие формы таков, каков представляют отдельные лексемы, отличные от таковой, каковой.

4. Вопросительные и относительные местоимения с неопределенными частицами кое-, либо-, -нибудь, -то даются как одно слово и при этом возводятся к заглавной форме: кое-кому→кое-кто, чем-либо→что-либо, чьим-то→чей-то. Местоимения с частицами -де, -таки сводятся к двум лексемам: какого-де→какой, де; что-таки→ что, таки.

5. Для местоимений с усеченной парадигмой заглавной формой считается род. пад.: некого, нечего, себя.

6. Архаические формы личных местоимений сводятся к современной заглавной форме: (помилуй)мя, (грешного)→я. Диалектные и просторечные формы, отклоняющиеся от нормативных одним-двумя знаками не возводятся к общепринятым при наличии у них стилевых оттенков: ихний→ихний, ктой-то→ктой-то.

7. Грамматикализованные словосочетания, сложные союзы, образованные от местоименных форм, расчленяются на отдельные слова: потому что→потому, что; в связи с тем, что→в, связь, с, то, что; чем..., тем→чем. тем.

8. В случае перехода местоимений в разряд других частей речи происходит утрата типичных для них грамматических парадигм. Многочисленны случаи субстантивации: наши, свое, такое. Не следует смешивать местоимения этот и это, тот и то, весь и все; первые изменяются по родам и числам (этот, эта, -о, -и), а вторые только по падежам (это, -ого, -им). Ср.: это издание быстро разошлось (это→этот); этого мне только не хватало (это-го→это).

Наблюдается использование местоименных форм в функции частиц. Ср.: все (блее), все (чаще)→все; всего {дольше}→всего.

Правила анализа глаголов

1. К исходной форме инфинитива сводятся все личные, деепричастные, безличные и повелительные формы: думая→думать, значит→значить, может быть→мочь, быть; пиши→писать. Супплетивные формы также сводятся к инфинитиву: будет (играть)→быть; шел, иду→идти.

Причастия возводятся к инфинитиву только в случаях, когда имеют при себе зависимые слова: окрашенная (в голубой цвет вода)→окрасить, но окрашенная (вода)→окрашенный. Исключения представляют устаревшие формы вроде: (власть) имущие→имущий. Страдательные причастия, утратившие связь с производящей глагольной основой, сводятся к форме им. пад. ед. ч. муж. р.: преданный (чему)→преданный, но преданный (кем)→предать. Краткие формы от этих причастий возводятся к полной форме, а в остальных случаях—к инфинитиву: влюблен→влюбленный, но рекомендован→рекомендовать.

2. Разными словами считаются видовые формы глагола (делать и сделать, красить и покрасить, разбросать и разбрасывать), супплетивные видовые пары (говорить и сказать, класть и положить), видовые формы глаголов, сходных по звуковому составу (удостоить и удостаивать) и разные глаголы одного вида (гнать и гонять).

3. Формы с возвратными частицами даются отдельными лексемами: купать и купаться, мыть и мыться.

4. В составном глагольном сказуемом обе формы возводятся к инфинитиву: начал вздрагивать→начать, вздрагивать; пошел ругаться→пойти, ругаться.

5. Усеченные глагольные формы даются в текстовом виде: скок, толк.

6. Глагольные формы с частицами, написанными через дефис, сводятся к двум лексемам; скажи-таки→сказать, таки; писать-то→писать, то.

7. Устойчивые сочетания расчленяются на отдельные лексемы: не мудрствуя лукаво→не, мудрствовать, лукаво.

8. Лексикализованные глагольные формы, а также глагольные частицы даются как отдельные лексемы: благодаря (опозданию)→благодаря; (расстались) молча→молча; (принимался) бывало (писать)→бывало; (он) поди (умылся)→поди.

9. Фонетические варианты типа весть, цвествь, возводятся к основной форме вести, цвести. Исключение составляют варианты, стилистически отифференцированные от исходных форм: подойти и подойтить, нагибаться и нагинаться.

Правила анализа наречий

1. Наречия, имеющие степени сравнения, сводятся к положительной степени: быстрее→быстро, суше→сухо.

Поскольку при параллельных формах на -ее и -ше часто имеет место лексикализация, такие формы даются как разные слова: далее и дальше, ранее и раньше.

Слова более и менее в роли показателей степени являются самостоятельными лексемами. В текстовой форме записываются также слова наиболее и наименее.

Формы сравнительной степени с приставкой и суффиксом записываются как самостоятельные лексемы: поскорее, потеплее.

Супплетивные формы учитываются как отдельные лексемы: много, больше, побольше; плохо, хуже.

2. Словообразовательные варианты наречий записываются как отдельные слова: давно и давненько, немного и немножко.

3. Фонетические варианты наречий возводятся к нормативному виду при отсутствии явных стилистических различий между ними: зимою→зимой, поскорей→поскорее, уж→уже, но даются как самостоятельные формы: даве, маненько, покеда, теперича.

4. Наречия с дефисом записываются как одно слово: по-волчьи, во-вторых, давным-давно, как-никак [*10]. Если дефис соединяет повторы, то эти формы считаются вариантом основной: тихо-тихо→тихо. Исключение из этого правила делается для трех форм: еле-еле, чуть-чуть, только-только, которые употребляются и

как союзы или частицы. При эмфазе дефис также, как и повторяющиеся буквы, опускается: о-очень, очень-очень→очень.

5. Устойчивые наречные выражения расчленяются: в насмешку→в, насмешка.

Наречные выражения, образованные от числительных, типа по трое, по четверо расчленяются как и устойчивые сочетания.

6. Онареченные формы существительных, деепричастий и др. разрядов отличаются от соответствующих омонимичных употреблений слов: порядком (надоело)→порядком, но в порядке→порядок.

Правила анализа предлогов

1. Фонетические варианты предлогов сводятся к основным формам: безо→без, изо→из; об, обо→о. Исключение составляют варианты со стилевой дифференциацией: меж и между, посреде и посреди.

2. Сложные предлоги с дефисом считаются отдельными словами: из-за, из-под.

3. Предложные составные сочетания расчленяются на отдельные лексемы: по случаю→по, случай; в силу→в, сила; в зависимости от→в, зависимость, от.

4. Предлоги отграничиваются от омонимичных употреблений существительных, деепричастий и т. д.: (мир достигнут) путем (переговоров)→путем; (он шел тернистым) путем→путь.

Правила анализа союзов

1. Сложные союзы расчленяются на отдельные лексемы: потому что→потому, что; с тем чтобы→с, то, чтобы.

Союзы чем, тем в сравнительных оборотах (более... чем; чем..., тем) записываются в текстовой форме.

Двойные союзы представляются как два употребления слова: ни... ни→ни, ни; то... то→то, то.

2. Фонетические варианты союзов хоть и хотя считаются разными словами. Союзы чтобы и чтоб сводятся к чтобы, иль и или — к или.

Правила анализа частиц

1. Фонетические варианты частиц сводятся к основной (полной) форме при отсутствии стилистических различий: ж→же, ль→ли. Пары али — аль, коли — коль стилистически дифференцированы, они фиксируются как разные единицы.

2. Частицы, присоединяемые к знаменательному слову дефисом, записываются как отдельные лексемы: де, таки, то.

Частица -то, а также -либо, -нибудь, сцепленные с вопросительными и относительными местоимениями, считаются морфемными элементами (см. выше: Правила анализа местоимений).

3. Частицы, омонимичные другим частям речи, представлены единичными формами. Словарные формы частиц и омонимов глаголов дифференцированы: (он) поди (умылся)→поди-, поди (за угол)→пойти. Неразличимыми от соответствующих омонимов-существительных остаются лишь слова: мол, раз, уж, чай.

Правила анализа междометий

Междометия с дефисом, функционально отличающиеся от соответствующих бездефисных форм, записываются как отдельное слово: ей-ей, ой-ой, те-те, то-то.

Для междометий-повторов заглавной формой считается двойное повторение: а-а-а→а-а.

Фонетически удлиненные междометия, как и прочие удлинения, даются отдельной лексемой, если включают нетождественные компоненты: н-да, тра-та-та, эх-ма.

4. ТЕХНИКА ОБРАБОТКИ МАТЕРИАЛА

Первоначальный проект предусматривал составление частотного словаря с помощью счетно-аналитических машин (САМ). Процесс обработки первой порции текстов в 120.000 словоупотреблений описан в книге Л. Н. Засориной «Автоматизация и статистика в лексикографии». По той же методике были обработаны тексты научно-публицистические и газетно-журнальные. Технические и организационные трудности эксплуатации САМ в дальнейшем привели к необходимости завершить работу над словарем на ЭВМ. При обработке большого массива информации обращение к ЭВМ имеет бесспорные преимущества по сравнению с машинами малой механизации.

Теперь уже известно, что САМ с цифровым вводом крайне неэкономичны при решении лингвистических задач. Процесс обработки требует значительного объема операций, осуществляемых вручную. В связи с этим малая механизация оправдывает себя только при работе с небольшими массивами информации. Так, сортировки в пределах 100000 карт уже неэффективны по затратам времени. Самой трудоемкой операцией на САМ является алфавитная сортировка. Изучались возможности поручить выполнение этой операции ЭВМ, рассматривались условия объединения САМ и ЭВМ в одну систему [*11].

Обработка первичных материалов словаря на САМ проводилась в ВЦ Ленинградского университета. Окончательная обработка материалов—в Горьковском государственном университете на ЭВМ БЭСМ—ЗМ. Ввод данных осуществлялся: а) с итоговых словников по газетно-журнальному и научно-публицистическому жанру; б) с первичных заготовок по отдельным источникам; в) с итоговых словников по текстам прозы и драматургии, выведенных на табулограммы.

Накопление и последующая обработка материалов проводилась по программам, составленным В.В.Бородиным [*12]. Контрольная печать сводного алфавитно-частотного словаря подвергалась многократному редактированию. Ошибки в записи лексем устранялись специальной программой корректировки данных.

На итоговую печать были последовательно выведены алфавитно-частотный словарь и частотный словарь. Первый из них полностью вошел в словарь, второй — в сокращенном виде, включая слова с частотой 10 и выше. Таблицы статистических распределений были составлены по окончательно отредактированному словарю.

Обобщая опытные данные о применении вычислительных машин в словарной работе, следует подчеркнуть необходимость резкого изменения организационных принципов в этом деле. Ожидаемый экономический эффект от ЭВМ может быть получен только при переходе к комплексной системе инвентаризации лексикографических данных по типу крупных информационных банков [*13].

5. ИСТОЧНИКИ СЛОВАРЯ

Публицистические и научные тексты

Ленин В.И. Военная программа пролетарской революции. — Сочинения. Изд. 4-е. Т. 23. Госполитиздат; 1 Всероссийский съезд по внешкольному образованию 6-19 мая 1919 г. Там же, т. 29; Задачи союзов молодежи. Там же, т. 31; Карл Маркс. Там же, т. 21; Лучше меньше, да лучше. Там же, т. 33; Марксизм и восстание. Там же, т. 26; О государстве. Там же, т. 29.

Материалы XXII съезда КПСС. М., Госполитиздат, 1962, с. 334—373.

Материалы XXIII съезда КПСС. М., Госполитиздат, 1966, с. 3—63.

Калинин М. И. Боевые задачи комсомольцев в колхозах. — В кн.: О коммунистическом воспитании. Избр. речи и статьи. [М.], «Молодая гвардия», 1956; О коммунистическом воспитании, (Докл. на собрании партийного актива г. Москвы 2 октября 1940 г.). Там же; О моральном облике нашего народа. Там же; Речь на собрании учащихся восьмых, девярых, десятых классов средних школ Ленинского района города Москвы 17 апреля 1941 года. Там же; О корреспондентах и корреспонденциях. — В кн.: Об искусстве и литературе. [Статьи, речи, беседы]. М., Гослитиздат, [Ленингр. отд-ние], 1957.

Вавилов С. И. Глаз и Солнце. — О «теплом» и «холодном» свете. М., Изд-во АН СССР, 1961, с. II—45, 113—157.

Грибачев Н. М. Оса из Лайфа: Оборотни; Туз из старой колоды; Лимонные корки; Штраус — новый троянский конь; Исландская сага с американской слезой; Шпейдель—тень над Европой; Карлос Ромуло—паяц с фирменной улыбкой; Голос нового века; Новая дистанция; Фаст—псаломщик ревизионизма; Человек и время; Время и цель.—Избр. произв. В 3-х т. Т. 3. М., Гослитиздат, 1960, с. 397—493.

Качалов Н. Н. Стекло. М., Изд-во АН СССР, 1959, с. 9—39, 284—295, 308—338.

Леонов Л. М. Горький сегодня.—Собр. соч. В 9-ти т. Т. 8. М., Гослитиздат, 1962; Наша Москва. Там же; О природе начистоту. Там же; Падение Зарядья. Там же; Слава России. Там же; Слово о Толстом. Там же.

Тарле Е. В. Наполеон. — Сочинения. В 12-ти т. Т. 7. М. Изд-во АН СССР, 1959, с. 44—133.

Тимирязев К. А. Жизнь растения. М., Изд-во АН СССР, 1962, с. 42-142.

Толстой А. Н. Что мы защищаем?—Полн. собр. соч. [В 15-ти т.]. Т. 14. М., Гослитиздат, 1950; Москве угрожает враг. Там же; Разгневанная Россия. Там же; Народ и армия. Там же; Вековая сила. Там же.

Ферсман А. Е. Занимательная минералогия. Изд. 2-е. М.—Л Детгиз, 1945, с. 24-130.

Шолохов М. А Слово о Родине. — Собр. соч. В 8-ми т. Т. 8. М., Гослитиздат, 1960.

Художественная проза

Антонов С. П. Весна.—В кн.: Весна. Рассказы. М., Гослитиздат, 1958; Поддубенские частушки.—В кн.: Деревенские повести. М., «Сов. Россия», 1957.

Гайдар А. П. Голубая чашка. — Сочинения. В 2-х т. Т. 2. М.—Л., Детгиз, 1949; Чук и Гек, Собр. соч. В 4-ч т. Т. 3. М., Детгиз, 1959.

Горький А. М. В людях. - Собр. соч. В 30-ч т. Т. 13. М., Гослитиздат, 1951, с. 205—344.

Леонов Л. М. Русский лес. Собр. соч. В 9-ч т. Т. 9. М., Гослитиздат, 1962, с. 121—242.

Овечкин В. В. В том же районе.—В кн.: Трудная весна. (Районные будни). Л., Лениздат, 1957.

Паустовский К. Г. Рождение моря. М., Воениздат, 1952, с. 14—129.

Пришвин М.М. Кашеева цепь. — Собр. соч. В 6-ти М., Гослитиздат, 1956, с. 19—120.

Соболев Л. С. Зеленый луч. — Избр. произв. В 3-х т, Т. 3 Гослитиздат, 1962, с. 7—141.

Толстой А. Н. Хмурое утро.—Собр. соч. В 10-ти т. Т. 6. М., Гослитиздат, 1959, с. 7—126.

Федин К. А. Первые радости. — Собр. соч. В 9-ти т. Т. 6. М., Гослитиздат, 1960, с. 123—242.

Шолохов М. А. Поднятая целина.—Собр. соч. В 8-ми т. Т. 6. М., Гослитиздат, 1960, с. 7—75.

Драматургия

Алешин С. И. Все остается людям.—В кн.: Пьесы. М., «Сов. писатель», 1962.

Арбузов А. Н. Годы странствий.—«Театр», 1954, №3; Тая.—В кн.: Театр. [Пьесы]. М., «Сов. писатель», 1961.

Афиногенов А. Н. Машенька.—В кн.: Пьесы. [М.], «Сов. писатель», 1956; Страх. Там же.

Булгаков М. А. Дни Турбиных. М., «Искусство», 1955.

Горький А. М. Враги.—Собр. соч. В 30-ти т. Т. 6. М., Гослитиздат, 1950; Достигаев и другие. Там же.

Зорин Л. Г. Друзья и годы,—«Театр», 1962, № 8.

Иванов В. В. Бронепоезд 14—69. — Собр. соч. В 8-ми т. Т. 1. М., Гослитиздат, 1958.

Лавренев Б. А. За тех, кто в море. — Избр. произв. В 2-х т. Т. 2. М., Гослитиздат, 1958; Разлом. Там же.

Леонов Л. М. Нашествие.—Собр. соч. В 9-ти т. Т. 7. М., Гослитиздат, 1961.

Погодин Н. Ф. Кремлевские куранты.— Собр. драм. произв. В 5-ти т. Т. 2. М., «Искусство», 1960; Человек с ружьем. Там же.

Розов В. С. В добрый час. М., «Искусство», 1959; В поисках радости. М., «Искусство», 1963.

Софронов А. В. Московский характер.—Избр. произв. В 2-х т. Т. 2. М., Гослитиздат, 1955; Стряпуха.—«Театр», 1959, № 8.

Симонов К. М. Русский вопрос. — Сочинения. В 3-х т. Т. 2. М., Гослитиздат, 1952.

Штейн А. П. Океан.—В кн.: Драммы. М., «Сов. писатель», 1966.

Журнальные и газетные тексты

«Вокруг света», 1958, № 5, с. 1—7, 17—18, 22—29, 33—49, 61—64; № 6, с. 5—41.

«Наука и жизнь», 1961, № 7, с. I—57.

«Новое время», 1960, № II, с. 1—25; № 13, с. 14—18.

«Новый мир», 1959, № 12, с. 194—253.

«Юность», 1962, № II, с. 66—76, 80, 82—112.

Газеты за 5 января 1968 г.: «Водный транспорт», «Известия», «Комсомольская правда», «Ленинградская правда», «Литературная Россия», «Правда», «Сельская жизнь», «Советская Россия», «Советский спорт», «Труд».

[*1] См. материалы 1, 2, 3 межвуз. конференций по частотным словарям (Межвузовская конференция по вопросам частотных словарей и автоматизации лингвостатистических работ. Тезисы докладов и сообщений. Л., 1966: Частотные словари и автоматическая переработка лингвистических текстов. Минск. 1968; Автоматическая переработка текста методами прикладной лингвистики. Кишинев, 1971), а также обзор Андриющенко В. М. «Новые работы в области статистической лексикографии» (В Я, 1968 № 5).

[*2] Объем выборки в словаре Г. Г. Йоссельсона равен 1 млн. словоупотреблений. Однако методика подсчета слов, разработанная по типу математико-статистического анализа, привела к такой форме представления данных, которая не дает полного перечня всех слов даже в зоне наиболее частых. Сведения о частоте даются в относительных пределах, в явном виде частоты не сопоставлены словам. Списки «первых» 2200 употребительных слов были откорректированы по нестатистическим критериям. См.: Josselson H. H. The Russian Word Count and Frequency Analysis of Grammatical Categories of Standard Literary Russian. Detroit, Wayne University Press, 1953. В журнале «Русский язык в национальной школе» (1960, № 6) перепечатаны первые 5 списков словаря Йоссельсона. Некоторые разделы предисловия к словарю в русском переводе включены в сб. «Автоматизация в лингвистике», М.—Л., «Наука», 1966.

[*3] Тот факт, что в словарях обычно не применяется какой-либо иной принцип упорядочения единиц, кроме алфавитного, свидетельствует о том, что для лексической системы до сих пор не найдено собственно лингвистической классификации.

[*4] Засорина Л. Н., Воробьева Э.В. К применению статистических методов при выделении базового словаря. — В сб.: Актуальные проблемы лексикологии. Тезисы докладов лингвистической конференции. Вып. II, ч. 1. Новосибирск, 1969; Засорина Л. Н. О деривационном словаре русского языка. — В сб.: Актуальные проблемы лексикологии. Тезисы докладов лингвистической конференции. Вып. II, ч. II. Новосибирск. 1969; Засорина Л. Н. Проект метаязыка для деривационного словаря. — В сб.: Структурно-математические методы моделирования языка. Тезисы докладов и сообщений всесоюзной научной конференции. Киев, 1970.

[*5] Josselson H. H. The Russian Word Count and Frequency Analysis of Grammatical Categories of Standard Literary Russian. Detroit, Wayne University Press, 1953; Штейнфельдт Э. А. Частотный словарь современного русского литературного языка. Таллин, 1963; Vakar N. P. A Word Count of Spoken Russian. The Soviet Usage. [Columbus], Ohio State University Press. 1966.

[*6] Земская Е. А. О понятии «разговорная речь». — В сб.: Русская разговорная речь. Саратов, Изд-во Саратовского ун-та, 1970, с. 4.

[*7] В первоначальном виде правила обработки текстовых слов были опубликованы в 1966 г. (Засорина Л. Н. Автоматизация и статистика в лексикографии. Л., 1966. Разд. II. Инструкция по подготовке текста к перфорации). Подробное рассмотрение этих правил положено в основу книги ВП Тимофеева «Исходная (словарная) форма слова в русском языке» (Свердловск, 1971).

[*8] В словник включены из текстов написания сложных существительных, не совпадающие с формами, данными Орфографическим словарем: динамонашна, еранито-гнейс, человекодень и др.

[*9] В дефисных написаниях существует известный разнобой, что отражает неустоявшиеся нормы в орфографии. В связи с этим формирование словника проводилось строго по тексту: заглавная форма отражает написание в тексте. Так, в алфавитно-частотном словнике имеются дефисные написания, отклоняющиеся от написаний, зарегистрированных Орфографическим словарем, напр.: англо-саксонский, густомалиновый, западно-сибирский, ново-кемеровский, нормально-деловой, общественно-полезный, сыр-дарьинский и др.

[*10] В источниках словаря зарегистрировано несколько случаев слитного и дефисного написания наречий, отличающихся от написаний Орфографического словаря, напр.: всердцах, дозарезу, полночь, на-днях, наруку, нестрашно, по-старинке и др.

[*11] См. подробнее Маслиева О. В. Опыт применения ЭВМ для алфавитной классификации слов. — В сб.: Межвузовская конференция по вопросам частотных словарей и автоматизации лингвостатистических работ. Тезисы докладов и сообщений. Л., Изд-во ЛГУ, 1960.

[*12] См. подробнее Бородин В. В. Автоматизация лексикографических работ. — В сб.: Прикладная математика и кибернетика. (Материалы к Всесоюз. межвузовскому симпозиуму по прикладной математике и кибернетике). Горький, 1967; Бородин В.В. Коновалова В.В. Обработка распределительных словников на ЭВМ. — «Труды ЦНИИПИ. Серия 3. Вопросы лингвостатистики и автоматизации лингвистических работ». 1972, вып. 5.

[*13] См. подробнее Засорина Л.Н., Сильвестров П.В. Информационный банк словарей. — В сб.: Материалы научного семинара «Семиотика средств массовой коммуникации». Ч. II, МГУ, М., 1973.

Распределение по частям речи

При контрольном обсчете данных словаря устанавливалась принадлежность каждой лексемы к той или иной грамматической категории. Рассматривались следующие грамматические категории: имя существительное, глагол, имя прилагательное, имя числительное, предлог, местоимение, союз, наречие, частица. Остальные разряды слов (междометия, модальные слова и др.) объединялись вместе, особо выделялись субстантивированные прилагательные (напр, *рядовой*), причастия, субстантивированные причастия (*ученый, учащийся*), омонимы типа: существительное/глагол (*печь, сечь, знать*) (см. Приложение 2 — Перечень словарных грамматических омонимов). В табл. 4 представлено частотное распределение частей речи в словаре. Здесь указывается относительная доля (в %) каждой категории в общем числе словоупотреблений по всему словнику в целом и по каждому функциональному стилю в отдельности.

Интересно отметить, что существительные занимают первое место по частоте употребления, следующее место занимают глаголы, причем в газетно-журнальных и научно-публицистических текстах количество глаголов заметно меньше за счет увеличения числа существительных и причастий. В текстах драматургии отмечено большее количество глаголов по сравнению с текстами других функциональных стилей. Распределение частот грамматических категорий знаменательных слов (существительное, глагол, прилагательное, наречие, местоимение) различается для каждого функционального стиля. Распределение частот служебных (нейтральных) слов (предлогов, союзов) оказывается равномерным для всех функциональных стилей.

Таблица 4

Грамматические категории	Частота				
	общая	по жанрам			
		газ.-журн.	драм.	н.-публ.	худ. пр.
1. Существительное	26,65	32,77	20,4	31,03	23,44
2. Глагол	17,12	14,5	20,88	13,50	18,96
3. Прилагательное	9,37	11,97	6,24	12,46	7,37
4. Наречие	8,096	6,95	9,01	7,26	8,98
5. Числительное	1,17	1,55	1,13	1,026	0,9969
6. Местоимение	13,29	10,01	16,18	11,55	14,94
7. Союз	7,39	6,57	6,81	7,61	8,56
8. Предлог	11,1	11,47	11,18	11,23	10,54
9. Частица	1,002	0,57	1,60	0,67	1,078
10. Причастие	0,979	1,05	0,523	1,36	1,05
11. Субстантив. причастие	0,053	0,08	0,06	0,032	0,044
12. Субстантив. прилагат.	0,4457	0,55	0,42	0,518	0,301
13. Омонимы (типа сущ./глагол.)	0,035	0,023	0,029	0,080	0,0489
14. Остальные	3,295	1,92	5,53	1,71	3,69
<i>Итого</i>	<i>99,996</i>	<i>99,983</i>	<i>99,992</i>	<i>100,036</i>	<i>99,999</i>