

Данный текст был найден в архивах Машинного фонда русского языка. Он первоначально создавался как отчет отдела Машинного фонда русского языка, отражавший состояние Машинного фонда на конец 2000 года. Позднее некоторые его разделы перерабатывались для публикации на сайте Машинного фонда. Так, написанный Л. И. Колодяжной раздел, посвященный лексикографической программной системе UNILEX-D, отражает состояние работ на 2004 год. Не удалось точно установить, какому времени соответствуют описания текстообработывающих систем (автор Ж. Г. Аношкина) и текстообработывающей и словарной системы UNILEX-A (В. М. Андрющенко).
Имеющийся в нашем распоряжении файл датирован 08.05.2004.

Е. Н. Морозова
Ноябрь 2018

Машинный фонд русского языка в Internet

От авторов

Настоящий Отчет подготовлен В.М. Андрущенко, Ж.Г. Аношкиной, Л.И. Колодяжной и Е.Н. Морозовой – сотрудниками отдела Машинного фонда русского языка РАН, продолжающими в 2000 г. дело, которое они начали 20 лет назад в составе гораздо большего коллектива и в другой социально-экономической среде. Мы выражаем свою искреннюю благодарность всем, кто сотрудничал с нами здесь и за рубежом, помогал нам и поддерживал нас в этом нелегком предприятии. Особую благодарность мы выражаем **Российскому гуманитарному научному фонду**, поддерживавшему нас последние пять лет. Мы пользуемся случаем еще раз подчеркнуть, что доступ к Машинному фонду русского языка для всех филологов открыт и бесплатен.

I. Введение

Идея создания Машинного фонда русского языка впервые была выдвинута советским математиком академиком А. П. Ершовым в докладе “Методология построения диалоговых систем: феномен деловой прозы” (1978 г.) Этот доклад был неоднократно опубликован, в частности, под названием “Методологические предпосылки продуктивного диалога с ЭВМ на естественном языке” в журнале “Вопросы философии”, 1981, № 8.)

Примечание [A1]:

В связи с проблемой взаимодействия человека и ЭВМ на естественном языке, прежде всего в производственных и иных регламентирующих отношениях, ученый поставил задачу: научить машину полностью воспринимать и понимать деловую прозу как языковой носитель модели, производственных отношений. Социально-экономическая посылка необходимости решения этой задачи сформулирована автором следующим образом: “Мне кажется нереальным насытить производственные отношения вычислительными машинами, расставляя на всех стыках бесчисленные интерфейсы в виде формализованных предписаний, стандартных бланков и других средств предварительной подготовки информации для машины. Время на обработку или синтез документа будет почти всегда соизмеримо со временем исполнения, подрывая тем самым все выгоды автоматизации. Мы не можем внедрить машины в повседневную жизнь, выделяя касту жрецов-посредников, но в таком случае мы должны быть готовы к тому, что число человеко-машинных интерфейсов в диалоговых системах возрастет в течение предстоящих 50 лет на несколько порядков. Не хватит никаких сил на то, чтобы снабдить эти миллионы интерфейсов специализированными языками и процессорами”.

Рассмотрев накопленные знания и технологические приемы системного программирования в области построения языковых процессоров, А. Ершов делает вывод, что любой прогресс в области построения моделей и алгоритмов останется лишь академическим упражнением, если не будет решена наиважнейшая задача — создание Машинного фонда русского языка. По мнению А. Ершова, это фундаментальная проблема, решение которой будет иметь большую научную, общекультурную и прикладную ценность. Желательно, подчеркнул он, чтобы создание такого фонда квалифицированными лингвистами опережало бы создание производственных лингвистических систем. Это не только позволило бы избежать дублирования больших усилий, но и защитило бы здоровую ткань русского языка от самоуправства и некачественного подхода.

Осознание необходимости использования ЭВМ в лингвистических исследованиях и в ходе реализации получаемых результатов как лингвистами, так и теоретиками и практиками автоматизированной переработки информации на естественных языках также вело к идее создания Машинного фонда русского языка.

Идея создания Машинного фонда русского языка как технологической основы для разработки систем общения с ЭВМ и обработки данных на естественном языке нашла отклик в среде системных программистов и разработчиков ЭВМ, послужила предметом дискуссий на конференциях и семинарах 1983–1986 гг., была поддержана лингвистической общественностью. Языко-

ведческая мысль связывала с этой идеей возможности подлинной и комплексной автоматизации лингвистических исследований и разработок.

Обсуждению вопросов о том, каким должен быть Машинный фонд русского языка по составу и структуре, решение каких задач он должен обеспечить в первую очередь, каков минимальный и максимальный круг его лингвистических источников, какими могли бы быть организационные и технические формы его существования, была посвящена состоявшаяся в феврале 1983 г. “Всесоюзная конференция по проблемам создания Машинного фонда данных для автоматизированной системы лексикографических исследований”. Конференция проводилась Научным советом по лексикологии и лексикографии АН СССР под руководством члена-корреспондента АН СССР **Ю. Н. Караулова**.

Участники конференции поддержали предложение Научного совета по лексикологии и лексикографии и Секции лингвистических проблем обработки информации Научного совета по комплексной проблеме “Кибернетика” АН СССР разработать проект Машинного фонда русского языка как системы комплексной автоматизации лингвистических исследований и разработок, ставящей следующие задачи:

1. Перевести академические словари и грамматики в машинную форму, т. е. создать академические словарно-грамматические базы данных. Решение этой задачи позволит:
 - а) передавать академические словари и грамматики в формализованном виде для использования в прикладных разработках;
 - б) регулярно и оперативно вести пополнение этих источников и их уточнение;
 - в) ускорить публикации исправленных и дополненных переизданий словарей и грамматик.
2. Создать автоматически пополняемые словоуказатели и словари на базе текстов деловых и разговорных стилей, а также текстов научно-технической литературы и документации. Решение этой задачи позволит:
 - а) ввести в научный обиход до сих пор мало использованный языковой материал;
 - б) уточнить данные академических словарей и грамматик в той части, которая в наибольшей степени связана с использованием русского языка в автоматизации процессов переработки информации во всех сферах народного хозяйства и общественной жизни;
 - в) ввести в научный обиход на широком материале принципиально новые типы лингвистических данных о статистико-комбинаторных свойствах лингвистических явлений.
3. Соединить в единый языковой фонд данные об общеупотребительном русском языке и данные терминологических фондов. Решение этой задачи позволит:
 - а) улучшить состав, структуру и лексикографическую информацию терминологических словарей;
 - б) расширить представление о современном русском литературном языке;
 - в) уточнить концепцию подязыка как проекцию общеупотребительного русского литературного языка на определенную предметную область или класс ситуаций общения.
4. Создать фонд лингвистических алгоритмов и программ, включая процессоры русского языка. Решение этой задачи позволит:
 - а) обеспечить типовыми программными средствами автоматизацию лингвистических исследований в академических и отраслевых институтах и вузах;
 - б) быстро настраивать общезыковые процессоры на конкретные применения;
 - в) применять общелингвистическое программное обеспечение для решения специальных задач, в частности задач автоматического анализа и синтеза звучащей речи.

По материалам этой конференции был издан сборник “*Машинный фонд русского языка: идеи и суждения*” (М.: “Наука”, 1986) и по ее рекомендации, поддержанной двумя научными советами АН СССР — по лексикологии и лексикографии и по комплексной проблеме “Кибернетика”, — а также Отделением литературы и языка АН СССР, задание на разработку Машинного фонда русского языка (1-я очередь 1986–1990 гг.) было включено в научно-техническую программу развития Государственной автоматизированной системы научно-технической информации на XII пятилетку. Был создан Координационный совет Машинного фонда русского языка, определена головная организация, ответственная за его создание, — Институт русского языка (ИРЯз) АН СССР.

Практическую координацию исследований, к которым были в 1985 г. привлечены свыше 40 научных коллективов, осуществлял организованный в ИРЯз отдел Машинного фонда русского языка (заведующий директор Института русского языка АН СССР член-корреспондент АН СССР Ю. Н. Караулов). На него была также возложена подготовка сводных проектных документов Фонда.

В слове “комплексная” заключены и смысл, и цель, и характер автоматизации лингвистических исследований. Уже в 1985 г. было понятно, что автоматизация только ведения текстовых источников (по существу, создание новых словарных картотек) не даст нужного результата, если не будут параллельно решены вопросы создания словарных баз данных, согласования словарной информации с грамматической, если выпадут из поля зрения вопросы систематического сбора языковой статистики, построения системы информационного обеспечения разработок.

Автоматизация затрагивала прежде всего словарное дело. Это естественно: именно здесь были накоплены и опыт, и образцы массового применения вычислительной техники и комплексности ее использования — от сбора источников до получения печатного продукта.

В понятие комплексности автоматизации входит также автоматизация исторических и диалектологических, социо- и психолингвистических исследований, изучения разговорной речи.

В проекте Фонда сконцентрировалось, таким образом, большое количество различных целей и задач, совместное и параллельное решение которых обеспечило бы комплексную автоматизацию лингвистических исследований.

Однако судьба решила иначе: только первые пять лет (1985–1989 гг.) исследования и разработки шли, как было задумано. Отдел Машинного фонда русского языка динамично развивался и оснащался вычислительной техникой, в Отдел стекались различные материалы, проводились всесоюзные конференции по созданию Фонда, издавались сборники и бюллетени. Была опубликована книга В. М. Андрищенко “*Концепция и архитектура Машинного фонда русского языка*”, М.: “Наука”, 1989. Был выработан проект создания Генерального словаря, Текстового, Словарного, Терминологического, Грамматического, Исторического, Фонетического и Диалектологического фондов русского языка.

Обстановка начала резко меняться в 1990 г., как раз тогда, когда идеи создания Машинного фонда русского языка начали приобретать свою законченность и осязаемость. Процесс создания Машинного фонда русского языка резко замедлился — в эти годы прекратилось финансирование по государственным научно-техническим программам. Новая научно-техническая политика отказала в поддержке проекту Машинного фонда русского языка, численность отдела сократилась вдвое, финансирование работ соисполнителей прекратилось, произошла полная смена и относительное сокращение общей мощности вычислительной техники. Коллектив, ранее работавший по научно-технической программе ГКНТ СССР, вынужден был заняться частными проектами, получавшими финансовую поддержку научных фондов.

Развитие Машинного фонда русского языка остановилось на стадии формирования Генерального словаря, Словарного, Текстового и Терминологического фондов. Прекратились работы по формированию Грамматического, Исторического, Диалектологического и Фонетического фондов. Развитие Словарного фонда остановилось на стадии формирования баз данных Словаря С. И. Ожегова, Орфографического и Грамматического словарей. Развитие Текстового фонда остановилось на стадии формирования 1-го Московского корпуса текстов, Терминологического фонда — на стадии формирования словаря. Все результаты (кроме исходных данных) по созданию Автоматического варианта Диалектологического атласа русского языка и формированию Диалектологического фонда оказались для Машинного фонда русского языка потерянными. Не осуществилась, и теперь уже, видимо, не осуществится идея коллективного создания Машинного фонда русского языка путем помещения в него результатов, полученных другими отделами Института с использованием программно-источниковых пакетов Фонда. Таким образом, был закрыт путь к саморазвитию Машинного фонда русского языка.

Не осуществились мероприятия по созданию профессиональной службы Архива Машинного фонда русского языка. Остановились работы по проверке, кодификации и маркированию источников, формированию текстовых корпусов.

Потеря специального финансирования, сокращение численности отдела Машинного фонда русского языка, начавшийся переход на персональные компьютеры, уменьшение возможностей для публикаций и научного обмена — все это не только не способствовало планомерному и массовому накоплению материалов и планомерной подготовке программных продуктов, но и препятствовало планомерной работе, отвлекая силы и средства на решение частных задач ради выживания. Эта ситуация продолжает иметь место и сейчас и, видимо, сохранится в будущем.

С 1996 г. развитие Машинного фонда русского языка определялось прямым или косвенным финансированием по грантам Российского гуманитарного научного фонда, Российского фонда фундаментальных исследований и Института “Открытое общество” (Фонд Сороса). Финансирование по грантам РГНФ 96-04-12033в “Включение Машинного фонда русского языка в Интернет” (руководитель — В. М. Андрищенко), РФФИ 97-06-80100 “Дистрибутивно-статистическое описание русской газеты” (руководитель — А. Я. Шайкевич), РФФИ 96-06-80044 “Разработка методики и системы для построения словарей по текстам” (руководитель — Н. А. Ребецкая), РФФИ 96-06-80143 “Дистрибутивно-статистическое описание русской прозы” (руководитель — А. Я. Шайкевич), РГНФ 96-04-06268 “Словарь Достоевского: лексикографическое отражение художественного текста” (руководитель — Ю. Н. Караулов), РГНФ 96-04-06264 “Дистрибутивно-статистический атлас русского классического романа”

(руководитель — А. Я. Шайкевич) позволили переоснастить Отдел новой вычислительной техникой и накопить значительное количество источников (см. раздел VI). Финансирование по грантам РГНФ 95-06-17344 “Словарь языка русской поэзии XX в. “Самовитое слово” (руководитель — В. П. Григорьев) и РГНФ 96-04-06103 “Создание лексико-грамматического словаря северного диалекта селькупского языка по материалам фольклорных текстов 20х–70х гг. XX века” (руководитель — О. А. Казакевич), в реализации которых принимали участие сотрудники Отдела Ж. Г. Аношкина и Л. И. Колодяжная, значительно способствовало развитию систем UNILEX-M и UNILEX-D соответственно. Развитие системы UNILEX-A было поддержано грантами CEU/RSS 589/95 (руководитель М. Г. Пшеничнова) и РФФИ 96-06-80044 или РФФИ 96-06-80143 (руководитель Н. А. Ребечкая).

Подводя итог пятнадцатилетней работы, постараемся точно сформулировать, что осталось жить, а что умерло или не родилось из романтических идей построения системы комплексной автоматизации лингвистических исследований 80-х годов в ИРЯз РАН.

1. Совершенно не проводились работы по информатизации в области *истории русского языка*, так как такие работы могли проводиться только силами самих историков. (Тем не менее нами были разработаны две системы электронной нотации древнерусских текстов в системах Chiwriter и TeX.)

2. Не проводились практические работы по накоплению научно-технической терминологии, также из-за недостатка сил и отсутствия конкретной терминологической тематики в Институте, хотя была проведена теоретическая работа и выпущена брошюра “*Лингвистическая концепция терминологического банка данных Машинного фонда русского языка (Проект)*” под ред. А. С. Герда (Институт русского языка АН СССР, МПНИИЯ им. М. Тореза, М.: 1989).

3. На машине ЕС-1022 был подготовлен и функционировал “*Автоматический вариант Диалектологического атласа русского языка*”, который погиб в 1992 г. в связи с экстренным демонтажем машины, но который оставил заметный след в исследованиях Н. Н. Пшеничновой (см. ее монографию “*Типология русских говоров*”, М.: “Наука”, 1996 и публикации в списке литературы).

В дальнейшем В. В. Пресновым за несколько лет (!) была воссоздана экспериментальная модель Диалектологического атласа русского языка (ДАРЯ), подготовленная в среде Paradox (на платформе Windows), представлявшая собой базу данных по русским народным говорам, собранным на территории центральных областей России. В базе представлено около 4000 населенных пунктов, в которых обнаружены диалектные различия по разделам: фонетика, морфология, синтаксис и лексика. Всего в базе данных зарегистрировано более 4000 признаков. Возможности разработанной системы управления позволяют вести поиск населенных пунктов в базе по заданным признакам методом итераций, сначала в пределах одного раздела, затем второго и далее, с использованием для дальнейших запросов уже выявленных пунктов. Аналогичным образом можно получать признаки, составляя запросы по конкретным населенным пунктам. Так как структура базы данных эквивалентна опубликованному Атласу, можно получить список карт и ее частей, в которых встречаются те или иные признаки. Для одной карты можно получить все населенные пункты, в которых встретился определенный признак. Запросы можно формировать и в пределах одной лингвогеографической карты (или ее части) по какому-то одному признаку.

После ухода В. В. Преснова из Института этот материал сохраняется в архиве.

Словарная компонента Диалектологического фонда, несмотря на все трудности, продолжала развиваться в Сыктывкарском университете. В **Приложении** приводится соответствующая справка.

4. В области создания Грамматического фонда были предприняты попытки построения информационных систем по “*Краткой русской грамматике*” (М.: 1989), по книге Г. А. Золотовой “*Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса*” (М.: 1988) и по книге Л. Л. Касаткина, Е. В. Клобукова и П. А. Леканта “*Краткий справочник по современному русскому языку*”; системы были построены, но кроме “*Синтаксического словаря*” не нашли применения и погибли, в то время как “*Синтаксический словарь*” вошел в состав системы UNILEX-D (см. ниже). В отделе выполнялись работы по созданию и подготовке к включению в Машинный фонд русского языка словаря-тезауруса “CrossLexica” И. А. Большакова. “CrossLexica” является многосвязным словарем-тезаурусом, характеризующим примерно 120 тыс. слов и выражений русского языка и отражающим семантические и синтагматические связи между ними. К семантическим связям относятся синонимы, антонимы, гипонимы, гиперонимы, меронимы, холонимы и семантические дериваты. К синтагматическим связям (словосочетаниям) относятся управляющие существительные и глаголы, сказуемые при данном существительном и модели управления. Каждая связь может рассматриваться с двух сторон, для синтагм — со стороны управляющего и управляемого слов. В итоге число всех односторонних связей семантического характера составляет примерно 1,5 млн., синтагматического характера — 1 млн. Родовые и синонимические связи конструктивно используются для порождения гипотетических словосочетаний, прямо в словаре не содержащихся (пить). Имеются

подпрограммы входной лемматизации и перевода на английский. И эта перспективная работа была потеряна в результате сокращения автора.

5. Проверку временем выдержали лишь текстообработывающие и словарные системы со своими корпусами и базами данных, на описании которых мы далее и сосредоточимся.

Всего нами были разработаны 5 вариантов систем, предназначенных для обработки текстов и ведения автоматических словарей:

1. **UNILEX-D** — автор Л. И. Колодяжная; система предназначена для ведения автоматических словарей на платформе MS DOS и выполнена на языке C1rper. Результаты этих работ многократно публиковались (см. описание системы и список литературы в разделе IV). Система используется в Институте русского языка РАН для подготовки разнообразных словарей.
2. **UNILEX-T** — автор Ж. Г. Аношкина; система предназначена для обработки текстов (подготовки частотных словарей и автоматических конкордансов, эксплуатации автоматических конкордансов как автоматических словарей) — на платформе MS DOS и выполнена на языке C++. Результаты работ многократно публиковались, наиболее добротное описание системы дано в брошюре **Ж. Аношкина** “Подготовка частотных словарей и конкордансов на компьютере (пособие для филологов)”, М.: Институт русского языка РАН, 1995. Система внедрена в Институте русского языка РАН, а также во многих организациях России и за рубежом.
3. **UNILEX-M** — автор Ж. Г. Аношкина; система предназначена для обработки текстов (создания текстовых корпусов, подготовки частотных словарей и автоматических конкордансов, эксплуатации автоматических конкордансов как автоматических словарей) на платформе Windows (95/98/NT). Описание помещено в разделе III.
4. **Рабочее место лексикографа (UNI)** — автор В. М. Андриященко; система предназначена для обработки больших массивов текстов (подготовки частотных словарей и автоматических конкордансов, включая подготовку их к печати и пересылку электронной почтой) на платформе MS DOS. Система разработана на языке Паскаль и эксплуатируется А. Я. Шайкевичем в теме “Дистрибутивно-статистический анализ языка”.
5. **UNILEX-A** — автор В. М. Андриященко; система предназначена для обработки текстов (создания текстовых корпусов, подготовки частотных словарей и автоматических конкордансов, эксплуатации автоматических конкордансов как автоматических словарей ведения автоматических словарей) на платформе Windows (95/98/NT). Система разработана на языке Ява, ее описание помещено в разделе V.

Кроме того, нами разработаны две системы для подготовки филологических изданий на платформе MS DOS (CfTeX и более ранняя — MfChi). Система MfChi не эксплуатировалась, так как развитие технологии издательского дела в этот период было особенно бурным. Довольно скоро выяснилось, что система TeX имеет значительные преимущества, и была начата разработка системы CfTeX и связанных с ней правил разметки текстов (см. В. М. Андриященко “Об организации архива источников Машинного фонда русского языка, их разметке и комментировании”, в Бюллетене Машинного фонда русского языка, вып. 1, Институт русского языка РАН, М.: 1992). Система CfTeX эксплуатировалась при подготовке к изданию журнала “Русистика сегодня” и нескольких монографий, в том числе по истории русского языка. (См. В. Б. Крысько “Развитие категории одушевленности в истории русского языка”, М.: Lyceum, 1994.)

В условиях новой научно-технической политики, в новых экономических и технологических условиях необходимо разрабатывать новую концепцию Машинного фонда русского языка. Исходя из реальных возможностей и реальных потребностей Машинный фонд русского языка нам видится теперь как совокупность файл-серверов, содержащих источники (тексты, словари, грамматики, диалектологические, психо- и социолингвистические данные), результаты их обработки в виде частотных словарей и конкордансов к текстовым корпусам и словарным базам данных, а также программные средства обработки этих и других источников.

Конечно, жаль расставаться со старой концепцией Фонда, рассматриваемого как множество фондов-составляющих, каждому из которых соответствовал свой лингвистический программно-источниковый пакет. Такая концепция могла быть реализована только в условиях долгосрочного планирования и устойчивого финансирования. Ее преимущества перед новой концепцией очевидны: она соответствовала естественному членению науки о языке и была направлена на комплексную информатизацию филологии. Комплексная информатизация филологии в свою очередь содействовала бы развитию прикладных направлений языкознания и расширению участия филологов в разработке современных средств автоматической обработки данных на естественных языках. Именно этот пункт был исходной точкой идеи А. П. Ершова о создании Машинного фонда русского языка.

Новая концепция, спасая материальные результаты предыдущей эпохи, отступает перед стихийностью информатизации филологии, давая простор случайности конкретных проявлений текущих потребностей и индивидуальных способностей. Это означает, что прорываясь на отдельном участке, в отделе Машинного фонда русского языка, на мировой уровень формирования лингвистических ресурсов, мы отказываемся от вовлечения в этот процесс других отделов, других направлений русистики. Этим мы не только содержательно обедняем концепцию Машинного фонда русского языка, но и вновь упускаем исторический шанс обогатить русистику прикладными направлениями.

В 1994–1998 гг. локальная сеть отдела Машинного фонда русского языка была подключена к Интернет (благодаря помощи и через сервер соседствовавшего тогда с нами подразделения Института физики Земли РАН) по адресам **cfrl0.cfrl.synapse.ru**, **cfrl1.cfrl.synapse.ru** и **cfrl2.cfrl.synapse.ru** (эти адреса больше не существуют). Функционировали http- и ftp-серверы Машинного фонда русского языка, на которых были размещены произведения 11 русских писателей XIX в. (Зеркало http-сервера сохранилось по адресу **solaris3.ids-mannheim.de**).

Осенью 1998 г. Отдел был и остается до сих пор отключенным от Интернет, но сохраняет и развивает далее свои серверы в автономном режиме, в надежде на подключение в будущем. Для этого есть существенные резоны:

1. Сохранение источников и программ их обработки на серверах Интернет является более подходящей формой организации Машинного фонда русского языка, чем просто хранилища носителей или локальные базы данных, как она нам представлялась в начале 80-х годов.
2. Программы обработки источников, выполненные как апплеты http-страниц, позволяют обрабатывать источники методом удаленного доступа (UNILEX-A — первая попытка создания такой универсальной программы).
3. Выполнение двух предыдущих условий обеспечивает сохранение и включение в Фонд результатов обработки источников, что позволяет нам вновь вернуться к идее саморасширения Фонда.

В настоящее время нами созданы четыре варианта хранилища данных Машинного фонда русского языка:

1. **http://www.cfrl6.irras0.iph.ras.ru** и **ftp://ftp.cfrl6.irras0.iph.ras.ru** — на платформе **Windows NT** в локальной сети irras0 на машине cfrl0 под Windows 98 (WinNT/2000-сервер на машине cfrl6);
2. **http://www.cfrl3.irras0.iph.ras.ru** и **ftp://ftp.cfrl3.irras0.iph.ras.ru** — на платформе **Windows 2000** в локальной сети irras1 на машине cfrl10 под Windows NT/2000 (WinNT/2000-сервер на машине cfrl3);
3. **http://www.cfrl3.irras1.iph.ras.ru** и **ftp://ftp.cfrl3.irras1.iph.ras.ru** — на платформе **Linux** в локальной сети irras1 на машине cfrl2 под Linux (LINUX-сервер на машине cfrl3);
4. **http://www.cfrl3.irras1.iph.ras.ru** и **ftp://ftp.cfrl3.irras1.iph.ras.ru** — на платформе **Linux** в локальной сети irras1 на машине cfrl10 под Linux (LINUX-сервер на машине cfrl3).

На этих серверах расположены произведения 27 писателей XIX и XX вв. и один корпус из русских газет за 1997 г. Доступ к этим источникам осуществляется из локальной сети Института русского языка РАН.

Весь архив источников переведен с внешних машинных носителей на жесткие диски этой же локальной сети и доступ к ним осуществляется локальными средствами.

В следующем разделе приведены полные списки источников Машинного фонда русского языка 2000.

II. Серверы и источники Машинного фонда русского языка

На серверах cfr13 и cfr16 хранятся и доступны по http и по ftp следующие источники и описание Архива источников Машинного фонда русского языка:

Русская проза XIX–XX вв.

1. Николай Васильевич Гоголь
 - Вечера на хуторе близ Диканьки
 - Старосветские помещики
 - Тарас Бульба
 - Вий
 - Повесть о том, как поссорились Иван Иванович с Иваном Никифоровичем
 - Невский проспект
 - Нос
 - Портрет
 - Шинель
 - Коляска
 - Записки сумасшедшего
 - Рим
 - Мертвые души
 - а также наброски, ранние редакции
2. Иван Александрович Гончаров
 - Обломов
 - Обрыв
3. Михаил Юрьевич Лермонтов
 - Герой нашего времени
 - Разная проза (Вадим, Княгиня Лиговская, Ашик-Кериб)
4. Иван Сергеевич Тургенев
 - Три встречи
 - Фауст
 - Ася
 - Первая любовь
 - Бригадир
 - Степной король Лир
 - Клара Милич
 - Рудин
 - Дворянское гнездо
 - Накануне
 - Отцы и дети
 - Дым
 - Вешние воды
 - Новь
5. Федор Михайлович Достоевский
 - Бедные люди
 - Двойник
 - Господин Прохарчин
 - Хозяйка
 - Роман в девяти письмах
 - Ползунков
 - Слабое сердце
 - Чужая жена и муж под кроватью
 - Честный вор
 - Елка и свадьба
 - Белье ночи
 - Неточка Незванова
 - Маленький герой
 - Дядюшкин сон
 - Село Степанчиково и его обитатели

- Униженные и оскорбленные
 - Записки из Мертвого дома
 - Скверный анекдот
 - Зимние заметки о летних впечатлениях
 - Записки из подполья
 - Крокодил
 - Преступление и наказание
 - Игрок
 - Идиот
 - Вечный муж
 - Бесы
 - Бобок
 - Подросток
 - Кроткая
 - Мальчик у Христа на елке
 - Мужик Марей
 - Сон смешного человека
 - Братья Карамазовы
6. Алексей Феофилактович Писемский
- Взбаламученное море
7. Михаил Евграфович Салтыков-Щедрин
- Господа Головлевы
8. Николай Гаврилович Чернышевский
- Что делать?
9. Лев Николаевич Толстой
- Детство
 - Отрочество
 - Юность
 - Война и мир
 - Анна Каренина
10. Николай Семенович Лесков
- Некуда
 - На ножах
 - Соборяне
11. Иван Алексеевич Бунин
- Антоновские яблоки (*Антоновские яблоки, Деревня*)
 - Чаша жизни (*Чаша жизни, Бернар, Книга, Преображение, Зойка и Валерия, Таня, Чистый понедельник, Кавказ, Смарагд, Ворон, Солнечный удар, Волки, Визитные карточки, Антигона, Танька, Поздней ночью, Туман, Тишина, Заря всю ночь*)
 - У истока дней (*У истока дней, Море богов, Снежный бык, Клаша, Косцы, Мистраль, Холодная осень, Лапти, Поздний час, Баллада, Ида, Хорошая жизнь, Личарда, Митина любовь, Последнее свидание, Казимир Станиславович*)
 - Суходол (*Суходол, Сверчок, Захар Воробьев, Лирник Родион, Братья, Грамматика любви, Господин из Сан-Франциско*)
12. Михаил Афанасьевич Булгаков
- Белая гвардия
 - Театральный роман
 - Мастер и Маргарита

Русская поэзия XIX–XX вв.

1. Константин Николаевич Батюшков
 - Полное собрание стихотворений. Л.: Советский писатель. 1964
2. Александр Сергеевич Пушкин
 - Полное собрание сочинений. Тт. 1–7. Изд-во Академии наук. 1937–1949
3. Евгений Абрамович Баратынский
 - Полное собрание стихотворений. Л.: Советский писатель. 1989
4. Афанасий Афанасьевич Фет
 - Полное собрание стихотворений. Л.: Советский писатель. 1959

5. Николай Алексеевич Некрасов
 - Полное собрание стихотворений в трех томах. Тт. 1–3. Л.: Советский писатель. 1967
6. Иннокентий Федорович Анненский
 - Стихотворения и трагедии. Л.: Советский писатель. 1990
7. Михаил Алексеевич Кузмин
 - Избранные произведения. Л.: Художественная литература. 1990
8. Александр Александрович Блок
 - Собрание сочинений в восьми томах. Тт. 1–4. М.–Л.: Художественная литература. 1960
9. Велимир Хлебников
 - Творения. М.: Советский писатель. 1986
10. Анна Андреевна Ахматова
 - Сочинения. Т. 1. М.: Художественная литература. 1986
 - В то время я гостила на земле... СПб.: Лениздат. 1995
11. Борис Леонидович Пастернак
 - Сочинения в четырех томах. Т. 3. М.: Художественная литература. 1989
12. Осип Эмильевич Мандельштам
 - Сочинения в двух томах. Т. 1. М.: Художественная литература. 1990
13. Марина Ивановна Цветаева
 - Собрание сочинений в семи томах. Тт. 1–3. М.: Эллис Лак. 1994
14. Владимир Владимирович Маяковский
 - Стихотворения. Поэмы. Пьесы. М.: Художественная литература. 1969
15. Сергей Александрович Есенин
 - Собрание сочинений в пяти томах. Тт. 1–3. М.–Л.: Художественная литература. 1961

Русский газетный корпус

1. Известия (июнь–декабрь 1997)
2. Литературная газета (январь, апрель, июль–декабрь 1997)
3. Московский комсомолец (июль–декабрь 1997)
4. Независимая газета (сентябрь–декабрь 1997)
5. Новая газета (июнь–декабрь 1997)
6. Правда–5 (сентябрь–декабрь 1997)
7. Российские вести (октябрь–декабрь 1997)
8. Сегодня (май–ноябрь 1997)
9. Санкт-Петербургские ведомости (июль–декабрь 1997)

Для каждого писателя на соответствующей [http-странице](#) приведены его портрет и краткая биографическая справка. Произведения доступны как в кодировке CP 1251, так и в альтернативной кодировке или в архивированном виде.

Архив источников Машинного фонда русского языка

Введение

Ниже приведена распечатка базы данных Архива источников Машинного фонда русского языка. Карта описания источника состоит из 23-х полей, которые группируются в 5 блоков. Отдельные характеристики образуют ПОЛЯ карты. Ниже описаны блоки и их поля.

Поля 1-5.

Блок информации об архивных характеристиках источника:

- номер источника,
- тип источника,
- способ хранения,
- код жанра в типологии Машинного фонда русского языка,
- жанр источника.

Номер источника — это его порядковый номер в описании Архива (1, 2, 3, ...).

Тип источника — это поле, в котором указывается цифра от 1 до 8, обозначающая одну из характеристик:

- 1 — текст прозаический,
- 2 — текст поэтический,
- 3 — массив текстов, составителем которого как источника является не автор произведения,
- 4 — текст словаря,
- 5 — база данных,
- 8 — текст разговорной речи.

Способ хранения — поле, в котором указывается одна из двух характеристик:

- база данных,
- текст.

Код жанра — поле, в котором указывается шифр, значение которого раскрывается в следующем поле.

Жанр — поле, в котором расшифровывается код жанра, напр.:

- Д — Драматургия,
- П — Поэзия,
- ПД — Поэзия, драматургия,
- ПР — Проза,
- ПЦ — Публицистика,
- Р — Разговорная речь,
- СБ — Словарная база данных,
- СЛ — Словарь (в виде текста),
- УЧ — Учебник.

Поля 6-9.

Блок библиографической информации:

- век создания произведения,
- автор источника,
- название источника,
- библиографическая справка.

Век создания произведения — поле, в котором записывается 18 в., или 18-19 вв., или 19 в., или 19-20 вв., или 20 в.

Автор источника — поле, в котором указывается:

- автор книги (рукописи),
- создатель машинного корпуса текстов источника, разработчик базы данных источника,

Название источника — поле, в котором указывается название источника в Архиве Машинного фонда русского языка. Название источника может:

- совпадать с названием произведения, напр.,
Словарь синонимов,

- служить названием корпуса и не совпадать с названием произведения (произведений), напр.,
Русский фольклор: Духовные стихи,
Тексты русской разговорной речи.

Библиографическая справка — поле, содержащее:

- библиографическую справку о печатном произведении (рукописи), напр.,
Русская баллада. М.: Сов. писатель, 1936. 502 с. (Б-ка поэта). С. 3-404.
- отсылку к другой единице описания Архива, напр.,
См. N 8 данного каталога.

Поля 10 — 16.

Блок информации, описывающий характеристики источника, связанные с его расположением на магнитном носителе:

- код носителя источника,
- носитель источника,
- количество носителей,
- архивный номер,
- число файлов,
- тип файла,
- код файла.

Код носителя источника — поле, содержащее символ:

D — копия источника находится на дискетах.

Носитель источника — поле, в котором записывается тип дискеты:

5"DS/DD — носителем являются пятидюймовые дискеты для персонального компьютера.

Количество носителей — поле, в котором записывается цифра, равная количеству дискет, занимаемых источником.

Архивный номер — поле, содержащее архивные номера дискет, на которых находится источник:

- # 1, 2, 8 — источник находится на дискетах с архивными номерами 1, 2, 8,
- # 105a — источник находится на дискете с номером 105a.

Число файлов — поле, содержащее цифру, которая указывает число файлов с информацией источника.

Тип файла — поле, содержащее одну из следующих характеристик файла:

арх. — архивированный файл (для файлов на дискетах),
арх., послед. — архивированный последовательный файл,
арх. и неарх., послед. — информация источника распределена по архивированным и неархивированным последовательным файлам,
неарх., послед. — неархивированный последовательный файл.

Код файла — поле, содержащее характеристики файла, являющиеся указанием на текстовой или базовый файл, или на тип двоичной кодировки символов текста:

ASCII — текстовый файл со стандартной кодировкой для компьютеров типа IBM PC XT/AT,

ASCII (КОИ-7)<EBCDIC (ДКОИ) — текстовый файл на дискетах, созданный путем копирования и перекодировки информации с магнитной ленты с кодировкой EBCDIC (ДКОИ),

база ".DBF", программа EXE — файл базовый типа ".DBF" с обрабатывающей программой типа "EXE".

Поля 17 — 18.

Блок статистической информации:

- число слов,
- число стихов.

Поля 19 — 23.

Блок дополнительной информации:

- источник поступления,
- коррекция,
- разметка,
- адресат копии,
- пользователи.

Источник поступления — поле, содержащее название организации или фамилию филолога, поставившего источник в Машинный фонд русского языка, имя лица или название организации, например, УПЦ

Адресат копии — поле, содержащее фамилию филолога или название организации, в которой находится копия источника,

Коррекция — поле, содержащее информацию о филологе, корректировавшем источник на компьютере и дату окончания коррекции, напр.:

Салий А. Д. 1990,
Орф. группа 1990.

Разметка — поле, содержащее указание о разметке источника, соответствующей некоторой программе обработки источника на компьютере, напр.:

ETC — разметка текста для программы ETC, составляющей частотные словари, словоуказатели и конкордансы;

OSP — разметка текста для Оксфордской программы OSP, составляющей частотные словари, словоуказатели, и конкордансы;

WORD- разметка текста для текстового процессора WORD.

Пользователи — поле, содержащее фамилию филолога, использующего источник Архива Машинного фонда в работе.

ДРАМАТУРГИЯ

2 Текст ПД Поэзия, драматургия 20 в. Анненский И. Стихотворения и трагедии. Л.: Сов. писатель, 19 59. 669 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 65-228, 305-578 Д 5"DS/DD 1 # 117а 1 арх., послед. ASCII (КОИ – 7 альтернативный) USA Кукушкина Е.Ю. (1990). Коробова М.М. (1991) OSP USA 1991 D.Worth, R.Vroon.

2 Текст ПД Поэзия, драматургия 20 в. Берггольц О.Ф. Стихотворения. Поэма. Трагедия. Берггольц О.Ф. Избранные произведения. Л.: Сов. писатель, 1983. 607 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 63-530 Д 5"DS/DD 1 # 112а 1 арх., послед. ASCII (КОИ – 7 альтернативный) УПЦ 1989-1990.

1 Текст Д Драматургия 19 в. Гоголь Н.В. Ревизор, Женитьба. Собрание сочинений в 7 томах. Т. 4. С. 5-153. М.: Худож. лит., 1978 г. Д 3"DS/DD 2 # 152, 153. 4 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Шайкевич М., 1994 г.) Шайкевич М., 1994 г.

2 Текст ПД. Поэзия, драматургия 20 в. Гумилев Николай. Стихотворения и поэмы. Драматические произведения. Стихотворения и поэмы. М.: Сов. писатель, 1988 г. Драматические произведения. Л.: Искусство, 1990г. Д 5"DS/DD 1 # 173. 2 арх., послед. ASCII (КОИ-7 альтернативный) с перфолент (Ловля Е.Н., Шумилевская В.Н., 1992 г.).

2 Текст ПД Поэзия, драматургия 19 в. Жемчужников А.М. Стихотворения. Поэмы. Сцены в стихах. Жемчужников А.М. Избранные произведения. М.-Л.: Сов. писатель 1963. 415 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 69-336 Д 5"DS/DD 1 # 43 42 неарх., послед. ASCII (КОИ – 7 альтернативный) УПЦ 1989-1990.

2 Текст ПД Поэзия, драматургия 19 в. Катенин П.А. Стихотворения. Поэмы. Драм. произведения Катенин П.А. Избранные произведения. М.-Л.: Сов. писатель, 1965. 740 с. (Большая сер., 2-е изд.). С. 61-656 Д 5"DS/DD 1 # 53 63 неарх., послед. ASCII (КОИ – 7 альтернативный) УПЦ 1989-1990.

2 Текст ПД Поэзия, драматургия 19 в. Лермонтов М.Ю. Драммы и трагедии. Лермонтов М.Ю. Полное собрание сочинений.

Т. 3. М.-Л.: Гослитиздат, 1947. 486 с. С. 23-470 Д 5"DS/DD 1 # 58 44 неарх., послед. ASCII (КОИ-7 альтернативный) УПЦ 1988-1989 гг.

2 Текст ПД Поэзия, драматургия 19 в. Майков А.Н. Избранные произведения. Л.: Сов. писатель, 1977. 910 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 47-784 Д 5"DS/DD 2 # 87, 88 2 арх., послед. ASCII (КОИ - 7 альтернативный) УПЦ 1991.

2 Текст ПД Поэзия, драматургия 19 в. Мей Л.А. Стихотворения и драмы. М.: Сов. писатель, 1947. 540 с. (Б-ка поэта). С. 12-507 Д 5"DS/DD 2 # 69, 71 неарх., послед. ASCII (КОИ - 7 альтернативный) УПЦ 1989-1990 гг.

2 Текст Д Драматургия 19 в. Пушкин А.С. Борис Годунов, Маленькие трагедии. Д 5"DS/DD 2 # 218 1 арх., послед. В формате WORD. Амстердамский университет.

2 Текст ПД Поэзия, драматургия 18 в. Сумароков А.П. Стихотворения. Трагедии. Сумароков А.П. Избранные произведения. Л.: Сов. писатель, 1957. 607 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 56-510. Д 5"DS/DD 2 # 67, 68 71 неарх., послед. ASCII (КОИ - 7 альтернативный) УПЦ, 1989-1990 гг.

2 Текст ПД Поэзия, драматургия 19 в. Хомяков А.С. Стихотворения и драмы. Л.: Сов. писатель, 1962. 549 с. (Б-ка поэта. Большая сер. 2-е изд.). С. 59-540 Д 5"DS/DD 2 # 5, 6 93 неарх., послед. ASCII (КОИ - 7 альтернативный) УПЦ 1989-1990.

2 Текст ПД Поэзия, драматургия 20 в. Цветаева М.И. Стихотворения. Поэмы. Драм. произведения Цветаева М. Избранные произведения. М.-Л.: Сов. писатель, 1965. 810 с. (Б-ка поэта. Большая серия. 2-е изд.). С. 57-558 Д 5"DS/DD 1 # 80 1 арх., послед. ASCII (КОИ - 7 альтернативный) УПЦ 1989-1990.

1 Текст Д Драматургия 19 в. Чехов А.П. Чайка. С. 3-60. Дядя Ваня. С. 61-116. Три сестры. С. 117-188. Вишневый сад. С. 195-254. Чехов А.П. Полное собрание сочинений и писем. В 30 т. Т. 13. М.: Наука, 1986. Д 5"DS/DD 1 # 139 1 арх., послед. ASCII (КОИ-7) <EBCDIC (ДКОИ) ЛО ИЯЗ. Амстердамский университет.

2 Текст ПД Поэзия, драматургия 19 в. Шаховской А.А. Комедии. Стихотворения. Шаховской А.А. Комедии. Стихотворения. Л.: Сов. писатель, 1961. 828 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 1-275 Д 5"DS/DD 1 # 98 1 арх., послед. ASCII (КОИ - 7 альтернативный) УПЦ 1989-1990.

ПУБЛИЦИСТИКА. РАЗГОВОРНАЯ РЕЧЬ. УЧЕБНИКИ

3 Текст ПЦ Публицистика 20 в. Андрющенко В.М., Климова О.Г. (сост. корп.) Общественно-политические тексты (Статист. выборка) Журналы "Коммунист", 1977-1986 гг. (Статистич. выборка, 1 млн. словоупотреблений) Д 5"DS/DD 20 # 180 - 200 10 арх., послед. ASCII (КОИ-7 альтернативный) 1000000 УПЦ 1987-1988 Копыл М., Коробова М., Пономарева Н.

1 Текст ПР Проза 20 в. Бахтин М. Тексты по литературоведению. Д 5"DS/DD 1 # 199 1 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ребецкая Н.А., 1995)

1 Текст ПЦ Публицистика 19 в. Белинский В.Г. <Письмо к Н.В.Гоголю 15 июля н.с. 1847 г.> Белинский В.Г. Полное собрание сочинений. Т. 10. М.: Изд. Академии наук СССР, 1956. С. 212-220. Д 5"DS/DD 1 # 276 арх., послед. ASCII (КОИ-7 альтернативный) 3283 ЛО ИЯЗ.

1 Текст ПЦ Публицистика 20 в. Бродский И.А. Нобелевская речь. Д 5"DS/DD 2 # 76, 78 2 арх., послед. ASCII (КОИ – 7 альтернативный) Климова О.Г. 1990.

8 Текст Р Разговорная речь 20 в. Тексты русской разговорной речи. I Русская разговорная речь. Тексты. М.: Наука, 1979. Д 5"DS/DD 3 # 119 арх., послед. ASCII (КОИ-7) <ЕВCDIC НИВЦ МГУ 1986 Марянян Т.В. (МГУ), Столярова Э.А. (Саратов). Авт. конкорданс Ж.Г. Мошкович, Саратов ГУ, Столярова Э.А. (Саратов), Красильникова Е.В.

8 Текст Р Разговорная речь 20 в. Тексты русской разговорной речи. II Русская разговорная речь. Тексты / Инфантова Г.Г. Экономия сегментных средств в синтаксисе соврем. рус. разг. речи. Дис. на соиск. уч. степени д-ра филол. наук. (Приложение) Л., 1975. (ЛПИ им. А.И.Герцена) Д 5"DS/DD 3 # 119 арх., послед. ASCII (КОИ-7 альтернативный) НИВЦ МГУ 1986 Марянян Т.В. МГУ, Ломковская М.В. 1991. Авт. конкорданс Ж.Г.Мошкович, Саратов ГУ, Столярова Э.А., Красильникова Е.В.

8 Текст Р Разговорная речь 20 в. Тексты русской разговорной речи. III Материалы Саратовского университета. Рукопись. Д 5"DS/DD 3 # 119 арх., послед. ASCII (КОИ-7 альтернативный) НИВЦ МГУ 1986.

8 Текст Р Разговорная речь 20 в. Тексты русской разговорной речи. IV Записи разговорной речи из фонотеки ИРЯ АН СССР. Рукопись. Д 5"DS/DD # 119 арх., послед. ASCII (КОИ-7 альтернативный).

8 Текст Р Разговорная речь 20 в. Тексты русской разговорной речи. IV Телефонная справочная служба "09" (Рукопись). Д 5"DS/DD 1 # 119 арх., послед. ASCII (КОИ-7 альтернативный) МГУ Лаб. А.Е.Кибрика, Морозова Е.Н., Ломковская М.В. 1991.

2 Текст ПР Проза 20 в. И.Сурат. Пушкинист Владислав Ходасевич. И.Сурат. Д 5"DS/DD 1 # 219 1 арх., послед. ASCII (КОИ-7 альтернативный). Сдала в архив Богуславская О.Ю.

3 Текст УЧ Учебник 20 в. Матюшенко А.Г. (составитель). Хрестоматия по марксистско-ленинской философии. Учебники по марксистско-ленинской философии. Д 5"DS/DD 1 # 114 арх., послед. ASCII (КОИ-7) <ЕВCDIC Матюшенко А.Г. 1987.

СЛОВАРИ

4 Текст СЛ Словарь 20 в. Грамматический словарь русского языка А.А.Зализняка Зализняк А.А. Грамматический словарь русского языка: Словоизменение. Ок. 100 000 слов. М.: Рус. яз., 1977. Д 5"DS/DD # 213 арх., послед. ASCII (КОИ – 7 альтернативный) НИВЦ МГУ.

4. Текст СЛ. Словарь 20 в. 1994 г. Аношкина Ж.Г. Краткий словарь омонимичных словоформ русского языка. Зализняк А.А.

Грамматический словарь русского языка. Д 5"DS/DD 1 # 164. 1 арх., послед. ASCII (КОИ-7 альтернативный) Аношкина Ж.Г.

4 Текст СЛ Словарь 20 в. Орфографический словарь русского языка Орфографический словарь русского языка: 106 000 слов/АН СССР. Ин-т русского языка; Под ред. С.Г.Бархударова и др. изд., стер. М.: Рус. яз., 1986. Д 5"DS/DD 10 # 101 арх., послед. ASCII (КОИ – 7 альтернативный). 106000. Орфографическая группа ИРЯ. 1990. Орф. группа. 1990. Word.

4 Текст СЛ Словарь 20 в. Золотова Г.А. Синтаксический словарь: Репертуар элементарных единиц русского синтаксиса. М.: Наука, 1988. Д 5"DS/DD арх., послед. ASCII (КОИ – 7 альтернативный). ИРЯ РАН.

4 Текст СЛ Словарь 20 в. Букчина Б.З., Калакуцкая Л.П. Слитно-раздельно?: (Опыт словаря-справочника). Ок. 82 000 слов. 6-е изд., стереотип. М.: Рус. яз., 1987. 876 с. Д 5"DS/DD 3 # 128 1 арх., послед. ASCII (КОИ – 7) <EBCDIC. 82000. НИВЦ МГУ.

4 Текст СЛ Словарь 20 в. Ожегов С.И. Словарь русского языка: 70 000 слов /Под ред. Н. Ю.Шведовой. 22-е изд., стер. М.: Рус. яз., 1990. Д 5"DS/DD 13 # 167 – 172 арх., послед. ASCII (КОИ – 7 альтернативный). ГВЦ. 1990.

4 Текст СЛ Словарь 20 в. Словарь синонимов. Под ред. А.П.Евгеньевой. М.: Наука, 1975. Д 5"DS/DD # 177 арх., послед. ASCII (КОИ – 7 альтернативный) ИРЯ РАН, МГУ.

4. Текст СЛ. Словарь 20 в. 1994 г. Аношкина Ж.Г. Словарь словоформ русского языка. Зализняк А.А. Грамматический словарь русского языка. Д 5"DS/DD 2 # 165,166. 20 арх., послед. ASCII (КОИ-7 альтернативный) 2000000. Аношкина Ж.Г.

4 Текст СЛ Словарь 20 в. Тихонов А.Н. Словообразовательный словарь русского языка: В 2 т. Ок. 145 000 слов. М.: Рус. яз., 1985. Д 5"DS/DD 5 # 148,149 1 арх., послед. ASCII. 145000. НИВЦ МГУ. Салий А.Д. 1990.

4 Текст СЛ Словарь 20 в. Розенталь Д.Э. Управление в русском языке: Словарь-справочник 2-е изд., М.: Книга, 1986. Д 5"DS/DD # 179 арх., послед. ASCII (КОИ – 7 альтернативный)

СЛОВАРНЫЕ БАЗЫ ДАННЫХ

5 База данных СБ Словарная база данных 20 в. Колодяжная Л.И. (разработчик). Синтаксический словарь русского языка Г.А.Золотовой. Компьютерная версия словаря работает под управлением лексикографического процессора UNILEX-D. Кол-во статей 1374. Д 5"DS/DD 4 арх. ASCII (.dbf).

5 База данных СБ Словарная база данных 20 в. Колодяжная Л.И. (разработчик). Словарь управления в русском языке Д.Э.Розенталя. Компьютерная версия словаря работает под управлением лексикографического процессора UNILEX-D. Кол-во статей 2500. Д 5"DS/DD арх. ASCII (.dbf).

5 База данных СБ Словарная база данных 20 в. Колодяжная Л.И. (разработчик). Словарь синонимов русского языка А.П.Евгеньевой. Компьютерная версия словаря работает под управлением лексикографического процессора UNILEX-D. Кол-во

статей 7000. Д 5"DS/DD арх. ASCII (.dbf). Поликарпов А.А. (МГУ).

5 БД СБ Словарная база данных 20 в. Колодяжная Л.И. (разработчик). Ожегов С.И. Словарь русского языка: / Под ред. Н.Ю.Шведовой. 22-е изд., стер. М.: Рус. яз., 1990. Компьютерная версия словаря С.И.Ожегова реализована как система словарных баз, каждая из которых содержит массив статей на одну букву (А-Я), работающих под управлением лексикографического процессора UNILEX-D. Кол-во статей 70.000. Д 5"DS/DD 4 # 167 – 171 28 арх. ASCII (.dbf). Колодяжная Л.И., Пшеничнова М.Г. В системе STAR С.Старостиным, а также различными филологами как в ИРЯ, так и в других институтах.

5 БД СБ Словарная база данных 20 в. Колодяжная Л.И. (разработчик). Букчина Б.З., Калакуцкая Л.П. Слитно-раздельно?: (Опыт словаря-справочника). 6-е изд., М.: Рус. яз., 1987 г. Компьютерная версия словаря "Слитно или раздельно?" реализована как система словарных баз, каждая из которых содержит массив статей на одну букву (А – Я), работающих под управлением лексикографического процессора UNILEX-D. Количество статей 80000. Д 5"DS/DD 4 # 128 28 арх. ASCII (.dbf). Словарь используется в Институте русского языка.

ПРОЗА

1 Текст ПР Проза 20 в. Белов В.И. Повести, рассказы, очерки. М.: Худож. лит., 1989г. Д 5"DS/DD 1 # 144 1 арх., послед. ASCII (КОИ-7 альтернативный). С перфолент (Ловля Е.Н., Шумилевская В.А., 1992 г.).

1 Текст ПР Проза 20 в. Белый А. Сочинения в 2-х томах. М.: Худож. лит., 1990 г. Д 5"DS/DD 1 # 144 1 арх., послед. ASCII (КОИ-7 альтернативный). С перфолент (Ловля Е.Н., Шумилевская В.А., 1992 г.).

1 Текст ПР Проза 20 в. Булгаков М.А. Белая гвардия; Театральный роман; Мастер и Маргарита: (Романы). М.: Худож. лит., 1988. Д 5"DS/DD 1 # 135 1 арх., послед. ASCII (КОИ – 7 альтернативный). Типография. Ломковская М.В. 1991 "OK".

2 Текст ПР Проза 20 в. Бунин И.А. Отдельные произведения. (Антоновские яблоки. Деревня. Суходол. Чаша жизни. У истока дней.) Бунин И.А. Избранное. М.: Молодая гвардия, 1987. Д 5"DS/DD 1 # 155, 158, 159. арх., послед. ASCII (КОИ-7) < EBCDIC (ДКОИ).

1 Текст ПР Проза 20 в. Бунин И.А. Собр. сочинений в 6 томах. Том 6. Собр. сочинений в 6 томах. Том 6. Худ. лит. 1989 г. Д 5"DS/DD 1 # 154 1 арх., послед. ASCII (КОИ-7 альтернативный). С перфолент.

1 Текст ПР Проза 19 в. Гоголь Н.В. Вечера на хуторе близ Диканьки. Миргород. Петербург. Мертвые души. Собрание сочинений в 7 томах. Т. 1-3, т. 5. М.: Худож. лит., 1978 г. Д 5"DS/DD 2 # 152, 153. 4 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Шайкевич М., 1994 г.) Шайкевич М. 1994г.

1 Текст ПР Проза 19 в. Гончаров И.А. Обломов. Роман в четырех частях. Школьная библиотека. Л.: Государственное

изд-во детской литературы министерства просвещения РСФСР. Стр. 5-540. Д 5"DS/DD 1 # 200 1 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ловля Е.Н., 1994 г.), коррекция – Ловля Е.Н. 1994 г.

1. Текст ПР Проза 19 в. Гончаров И.А. Обрыв. Собрание сочинений в 8-ми томах. Т. 5, 6. М.: Худож. лит., 1979-1980г., стр. 31-702. Д 5"DS/DD 3. # 272-274. 3 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Морозова Е.Н., Ребецкая Н.А., 1994), коррекция – Морозова Е. Н., Ребецкая Н.А., 1994 г.

1. Текст ПР Проза 19 в. Достоевский Ф.М. Бесы. Подросток. Полное собрание сочинений в тридцати томах. Т. 10, 11, 13. Д 5"DS/DD 1 # 229. 2 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ребецкая Н.А., Ловля Е.Н., 1994), коррекция сделана Ребецкой Н.А., Ловля Е.Н., 1994 г.

1. Текст ПР Проза 19 в. Достоевский Ф.М. Бедные люди, Двойник. Полное собрание сочинений в тридцати томах. Т. 1. Л.: Наука, 1972г. Д 5"DS/DD 1 # 224. 2 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ребецкая Н.А. Отдел экспериментальной лексикографии – Двойник.) 1994 г. Коррекция сделана Ребецкой Н.А., отдел экспериментальной лексикографии – Двойник.

1. Текст ПР Проза 19 в. Достоевский Ф.М. Униженные и оскорбленные. Вечный муж. Игрок. Скверный анекдот. Зимние заметки о летних впечатлениях. Записки из подполья. Крокодил. Полное собрание сочинений в тридцати томах. Т. 3, 5, 9. Л.: Наука. 1972 г. Д 5"DS/DD 1 # 223. 7 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ребецкая Н.А., Морозова Е.Н., 1994), коррекция сделана Ребецкой Н.А., Морозовой Е.Н., 1994 г.

1. Текст ПР Проза 19 в. Достоевский Ф.М. Село Степанчиково. Ползунков. Слабое сердце. Чужая жена и муж под кроватью. Честный вор. Елка и свадьба. Белые ночи. Неточка Незванова. Маленький герой. Полное собрание сочинений в тридцати томах. Т. 2, 3. Л.: Наука, 1972 г. Д 5"DS/DD 1 # 226. 9 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ребецкая Н.А., Морозова Е.Н., 1994), коррекция сделана Ребецкой Н.А., Морозовой Е.Н., 1994 г.

1 Текст ПР Проза 19 в. Достоевский Ф.М. Идиот. Полное собрание сочинений в тридцати томах. Т. 8. Л.: Наука, 1972 г. Д 5"DS/DD 1 # 225. 1 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ребецкая Н.А., Морозова Е.Н., 1994), коррекция сделана Ребецкой Н.А., Морозовой Е.Н., 1994 г.

2 Текст ПР Проза 19 в. Ермоленко Г.В. (составитель корпуса). Рассказы А.П.Чехова. Чехов А.П. Полное собрание сочинений и писем. В 30 т. М.: Наука, 1974. Т. 2, с. 195-198, 201-221, 250-251; Т. 3, с. 52-55; Т. 4, с. 121-125; Т. 7, с. 340-352; Т. 8, с. 7-31, 310-322; Т. 9, с. 161-173, 174-191; Т. 10, с. 55-68. Д 5"DS/DD 1 # 139 арх., послед. ASCII (КОИ-7) < EBCDIC (ДКОИ). Гродн. ГУ.

2 Текст П Проза 20 в. Ерофеев Венедикт. Москва-Петушки. Поэма. М.: СП "Интербук". 1990 г., 128 стр. Д 5"DS/DD 1 # 145 2 арх. послед. ASCII (КОИ-7 альтернативный). Введена И.Вольфсон в Саратовском университете в 1994 г.

1 Текст ПР Проза 19 в. Лермонтов М.Ю. Герой нашего времени. Лермонтов М.Ю. Собрание сочинений. В 4 т. Т. 4. М.: Худож. лит., 1958. С. 7-159 Д 5"DS/DD 1 # 276 арх., послед. ASCII (КОИ-7 альтернативный). 26222. Амстердамский университет.

1 Текст ПР Проза 19 в. Лермонтов М.Ю. Сочинения в двух томах. Т. 2. М.: Правда, 1990. С. 275-455. Д 5"DS/DD 1 # 276 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Галиневич, 1995), коррекция сделана Галиневич Н.

1 Текст ПР Проза 19 в. Лесков Н.С. Некуда. Роман. Собрание сочинений. Т. 1. М.: АО Экран, 1993. С. 15-695. Д 3"DS/DD 1 # 291 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Морозова Е.Н., 1995 г.), коррекция сделана Морозовой Е.Н., 1995 г.

1 Текст ПР Проза 20 в. Маканин В.С. Человек свиты. Избранное. М.: Сов. пис., 1987 г. С. 169-216. Д 5"DS/DD 1 # 73. 1 арх., послед. ASCII (КОИ-7 альтернативный) УПЦ 1990 г.

1 ПР Проза 20 в. Пастернак Б.Л. Доктор Живаго. / Публ., подготовка текста Е.Б. Пастернака, В.М. Борисова. Вильнюс: Вага, 1988. Кн. 1 - 222 с. Кн. 2 - 252 с. Д 5"DS/DD 2 # 136 арх., послед. ASCII (КОИ - 7 альтернативный). Климова О.Г. 1989.

1 Текст ПР Проза 20 в. Платонов А. Котлован. Новый мир. N 6. 1987 г. Стр. 5-120. Д 5"DS/DD 1 ASCII (КОИ - 7) < EBCDIC (ДКОИ).

1 Текст ПР Проза 20 в. Платонов А. Чевенгур. Роман. М.: Худож. лит., 1988 г. Д 5"DS/DD 1 # 143. 1 арх., послед. ASCII (КОИ-7 альтернативный). С перфолент (Ловля Е.Н., Шумилевская В.А., 1992 г.).

1 Текст ПР Проза 19 в. Пушкин А.С. Повести покойного Ивана Петровича Белкина. Текст по изданию: "Книга для чтения с комментарием на нидерландском языке и словарем". 2-е издание. М.: Русский язык, 1986 г. Д 3"DS/DD 1 # 218 1 арх., послед. ASCII (КОИ-7 альтернативный). Амстердамский университет.

1 Текст ПР Проза 19 в. Салтыков-Щедрин М.Е. Господа Головлевы. Салтыков-Щедрин М.Е. Господа Головлевы. Сказки. Отечественная проза. М.: Правда. 1988 г. С. 3-283. Д 5"DS/DD 1 # 104 1 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Морозова Е.Н., 1994 г.), коррекция сделана Морозовой Е.Н., 1994 г.

1 Текст ПР Проза 20 в. Солженицын А.И. Архипелаг ГУЛАГ. Д 5"DS/DD 4 # 126, 127 арх., послед. ASCII (КОИ - 7 альтернативный). Климова О.Г. 1989.

1 Текст ПР Проза 20 в. Стругацкие А.Н. и В.Н. Далекая радуга. Д 5"DS/DD 1 # 157 1 арх., послед. ASCII (КОИ - 7 альтернативный). Богуславская О.Ю.

1 ПР Проза 20 в. Стругацкие А.Н. и В.Н. Пикник на обочине. Стругацкие А.Н. и В.Н. За миллиард лет до конца света: Повести. М.: Сов. писатель, 1984. Д 5"DS/DD 1 # 157 (157d) 4 неарх., послед. ASCII (КОИ - 7 альтернативный). Климова О.Г., Ломковская М.В. 1991.

1 Текст ПР Проза 20 в. Стругацкие А.Н. и Б.Н. Трудно быть богом. Стругацкие А.Н. и Б.Н. За миллиард лет до конца света: Повести. М.: Сов. писатель, 1984. 415 с. Д 5"DS/DD 1 # 157 1 арх., послед. ASCII (КОИ – 7 альтернативный). Богуславская О.Ю.

1 Текст ПР Проза 20 в. Стругацкие А.Н. и Б.Н. Улитка на склоне. Стругацкие А.Н. и Б.Н. Волны гасят ветер: Повести. Л.: Сов. писатель, 1989. Д 5"DS/DD 1 # 157 1 арх., послед. ASCII (КОИ – 7 альтернативный). Богуславская О.Ю.

1 Текст ПР Проза 19 в. Толстой Л.Н. Анна Каренина. Собрание сочинений. Том 8, 9. М.: Худож. лит., 1958 г. Д 5"DS/DD 2 # 146, 147 1 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ребецкая Н.А., Морозова Е.Н., 1994 г.), коррекция сделана Ребецкой Н.А., Морозовой Е.Н., 1994 г.

1 Текст ПР Проза 19 в. Толстой Л.Н. Война и мир. Собрание сочинений. Том 4-7. М.: Худож. лит., 1958 г. Д 5"DS/DD 2 # 209, 210 2 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ребецкая Н.А., 1994 г.), коррекция сделана Ребецкой Н.А., 1994 г.

1 Текст ПР Проза 19 в. Толстой Л.Н. Детство. Отрочество. Юность. Собрание сочинений. Том 1. М.: Худож. лит., 1958 г. Д 5"DS/DD 1 # 210 1 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ребецкая Н.А., 1994 г.), коррекция сделана Ребецкой Н.А., 1994 г.

1 Текст ПР Проза 19 в. Тургенев И.С. Дворянское гнездо. Тургенев И.С. Собрание сочинений и писем. В 12 т. Т. 2. М.: Худож. лит., 1954 г. Стр. 141-307. Д 5"DS/DD 1 # 163 арх., послед. ASCII (КОИ – 7 альтернативный).

1 Текст ПР Проза 19 в. Тургенев И.С. Отцы и дети. Тургенев И.С. Собрание сочинений и писем. В 12 т. Т. 3. Отцы и дети. М.: Худож. лит. 1954. С. 167-370. Д 5"DS/DD 1 # 163 арх., послед. ASCII (КОИ – 7 альтернативный).

1 Текст ПР Проза 19 в. Тургенев И.С. Рудин. Тургенев И.С. Собрание сочинений и писем. В 12 т. Т. 2. М.: Худож. лит. 1954 г. С. 7-137. Д 5"DS/DD 1 # 163 арх., послед. ASCII (КОИ-7 альтернативный).

2 Текст ПР Проза 19 в. Тургенев И.С. Стихотворения в прозе. Тургенев И.С. Собрание сочинений. Т. 10. М.: Правда, 1949. (Б-ка "Огонек"). С. 141-207. Д 5"DS/DD 1 # 90 1 арх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1990 г.

1 Текст ПР Проза 19 в. Тургенев И.С. Новь. М.: Худ. лит., 1970 г. Текст напечатан по изданию: И.С. Тургенев. Полное собр. сочинен. в 28 томах. Сочинения в 15 томах, Наука, т. 9. М.-Л. 1965, т. 12, М.-Л. 1966, стр. 177-447. Д 5"DS/DD 1 # 150 1 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Шумилевская В.Н., 1994 г.), коррекция сделана Шумилевской В.Н., 1994 г.

1 Текст ПР Проза 19 в. Тургенев И.С. Вешние воды, Дым. Сочинения в 2-х томах. Том 2. Повести и романы. М.: Худож. лит., 1980 г. Д 5"DS/DD 1 # 150 1 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Шумилевская В.Н., 1994 г.), коррекция сделана Шумилевской В.Н., 1994 г.

1 Текст ПР Проза 19 в. Тургенев И.С. Три встречи. Фауст. Ася. Первая любовь. Бригадир. Степной король Лир. Клара Милич. Полное собрание сочинений. Стр. 408-673. Д 5"DS/DD 1 # 151 9 арх., послед. ASCII (КОИ-7 альтернативный). С перфолент (Ловля Е.Н., Шумилевская В.Н., 1992 г.). Гинзбург Е.Л., 1994 г.

1 Текст ПР Проза 19 в. Тургенев И.С. Накануне. Тургенев И.С. Собрание сочинений в двенадцати томах. Т. 3. М.: Худож. лит., 1954 г. С. 7-164. Д 5"DS/DD 1 # 163 1 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование (Ребецкая Н.А., 1994 г.), коррекция сделана Ребецкой Н.А., 1994 г.

1 Текст ПР Проза 20 в. Ходасевич В.Д. Избранное. Л.: Сов. писатель, 1990 г. Д 5"DS/DD 1 # 176 1 арх., послед. ASCII (КОИ-7 альтернативный). С перфолент (Ловля Е.Н., Шумилевская В.Н., 1992 г.)

1 Текст ПР Проза 20 в. Цветаева М.И. Проза, письма. Сочинения в 2 томах. М.: Худож. лит., 1988 г. Д 5"DS/DD 1 # 175 1 арх., послед. ASCII (КОИ-7 альтернативный). С перфолент (Ловля Е.Н., Шумилевская В.Н., 1992 г.)

1 Текст ПР Проза 20 в. Шукшин В.М. Любавины: Роман. М.: Сов. писатель, 1990. 421 с. Д 5"DS/DD 2 # 155 (156) 6 арх., послед. ASCII (КОИ - 7 альтернативный). Типография. Ломковская М.В. 1991.

ПОЭЗИЯ

2 Текст П Поэзия 20 в. Антокольский П.Г. Стихотворения и поэмы. Антокольский П.Г. Избранные произведения. В 2 т. Т. 2. Стихотворения и поэмы, 1941-1976. М.: Худож. лит., 495 с. С. 3-460. Д 5"DS/DD 1 # 113а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ. 1991.

2 Текст П Поэзия 19 в. Апухтин А.Н. Стихотворения. Апухтин А.Н. Песни моей отчизны. Тула: Приок. кн. изд-во, 1985. С. 16-186. Д 5"DS/DD 1 # 38 31 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Асеев Н.Н. Стихотворения и поэмы. Л.: Сов. писатель, 1967. 734 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 65-656. Д 5"DS/DD 2 # 62, 67 73 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Ахмадулина Б.А. Стихотворения. Ахмадулина Б.А. Избранное. Стихи. М.: Сов. писатель, 1988. 477 с. С. 7-473. Д 5"DS/DD 1 # 92 1 арх., послед. ASCII (КОИ-7 альтернативный). УПЦ, 1991. Копыл М. 1991.

2 Текст П Поэзия 20 в. Ахмадулина Б.А. Стихотворения. М.: Худож. лит., 1988. 333 с. (БСП. Б-ка сов. поэзии). С. 9-328. Д 5"DS/DD 1 # 107а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 20 в. Багрицкий Э. Стихотворения и поэмы. М.-Л.: Сов. писатель, 1964. 569 с. (Б-ка поэта. Большая серия. Изд. 2-е). С. 18-476 Д 5"DS/DD 2 # 39, 40 60 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Бальмонт К.Д. Стихотворения. Л.: Сов. писатель, 1969. 610 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 77-604. Д 5"DS/DD 1 # 81 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Батюшков К.Н. Стихотворения. Батюшков К.Н. Полное собрание стихотворений. М.-Л.: Сов. писатель, 1964. 354 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 55-252. Д 5"DS/DD 1 # 42 34 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Бедный Д. Стихотворения и поэмы. М.-Л.: Сов. писатель, 1965. 591 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 49-520. Д 5"DS/DD 3 # 65, 66, 67 80 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Безыменский А.И. Поэмы. Безыменский А. Избранные произведения. В 2 т. Т. 2. Поэмы, пьесы. М.: Гослитиздат, 1958. 297 с. С. 7-247. Д 5"DS/DD 2 # 108а. 2 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Безыменский А.И. Стихотворения. Безыменский А. Избранные произведения. В 2-х т. М.: Гослитиздат, 1958. Т. 1. Стихотворения. 311 с. С. 5-297. Д 5"DS/DD 2 # 108а 2 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Белый А., Стихотворения и поэмы. М.-Л.: Сов. писатель, 1966. 656 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 71-537. Д 5"DS/DD 1 # 83 1 арх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1990 г.

2 Текст П Поэзия 19 в. Бенедиктов В.Г. Стихотворения. Л.: Сов. писатель, 1939. 334 с. (Б-ка поэта. Большая сер. 2-е изд.). С. 1-293. Д 5"DS/DD 2 # 10, 13 63 неарх., послед. ASCII (КОИ -7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Бестужев-Марлинский А.А. Стихотворения. Бестужев-Марлинский А.А. Полное собрание стихотворений. Л.: Сов. писатель, 1961. 311 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 51-256. Д 5"DS/DD 1 # 17 45 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19-20 в. Блок А.А. Стихотворения 1897-1904. Блок А.А. Собрание сочинений. В 8 т. Т. 1. Стихотворения. 1987-1904. М.-Л.: Худ. литература, 1960. 715 с. С. 3-556. Д 5"DS/DD 2 # 9, 11 64 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Блок А.А. Стихотворения и поэмы 1907-1921. Блок А.А. Собрание сочинений. В 8 т. Т. 3. Стихотворения и поэмы. 1907-1921. М.-Л.: Худ. литература, 1960. 714 с. С. 7-496. Д 5"DS/DD 1 # 84 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 18 в. Богданович И.Ф. Стихотворения и поэмы. Л.: Сов. писатель, 1957. 257 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 53-360 Д 5"DS/DD 1 # 49 30 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Боков В.Ф. Стихотворения. Боков В. Избранные произведения. В 2 т. Т. 1. Стихотворения. М.: Худож. лит., 1975. 270 с. С. 21-265. Д 5"DS/DD 3 # 86, 93, 99 20 арх. и неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Боков В.Ф. Стихотворения. Песни. Боков В. Избранные произведения. В 2 т. Т. 2. Стихотворения. Песни. М.: Худож. лит., 1975. 336 с. С. 7-190. Д 5"DS/DD 3 # 86, 93, 99 20 арх. и неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Бродский И.А. Стихотворения. Бродский И. Старые стихи. Шестивие. Чехов, 1970. Конец прекрасной эпохи. Ардис, 1971. Часть речи. Ардис, 1977. Новые стансы к Августе. Ардис, 1983. Урания. Ардис, 1988. Д 5"DS/DD 2 # 76, 78 2 арх., послед. ASCII (КОИ-7 альтернативный). Климова О.Г. 1990.

2 Текст П Поэзия 19-20 в. Брюсов В.Я. Стихотворения и поэмы. Л.: Сов. писатель, 1961. 910 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 69-666. Д 5"DS/DD 1 # 85 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Бунин И.А. Стихотворения. Бунин И. Стихотворения. Л.: Сов. писатель, 1956. 487 с. (Б-ка поэта. Большая серия, 2-е изд.). С. 31-377. Д 5"DS/DD 1 # 56 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Ваншенкин К.Я. Стихотворения 1945-1962. Ваншенкин К.Я. Избранные произведения. В 2 т. Т. 1. Стихотворения (1945-1962). М.: Худож. лит., 1975. 399 с. С. 11-391. Д 5"DS/DD 2 # 100, 101а 58 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Ваншенкин К.Я. Стихотворения 1963-1974. Ваншенкин К.Я. Избранные произведения. В 2 т. Т. 2. Стихотворения (1963-1974). М.: Худож. лит., 1975. 511 с. С. 7-499. Д 5"DS/DD 2 # 1S/DD 58 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Васильев П. Избранные стихотворения и поэмы. М.: Худож. лит., 1957. С. 23-450. Д 5"DS/DD 2 # 33, 34 76 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Винокуров Е.М. Стихотворения 1944-1969. Винокуров Е.М. Собрание сочинений. В 3 т. Т. 1. Стихотворения 1944-1969. М.: Худож. лит., 1983. 526 с. С. 11-515. Д S/DD 5"DS/DD 2 # 79, 86 2 арх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Винокуров Е.М. Стихотворения 1970-1983. Винокуров Е.М. Собрание сочинений. В 3 т. Т. 2. Стихотворения 1970-1983. М.: Худож. лит., 1983. 591 с. С. 2-576. Д 5"DS/DD 2 # 79, 86 2 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Вознесенский А.А. Стихотворения. Поэмы. Вознесенский А.А. Собрание сочинений: В 3 т. Т. 1. Стихотворения. Поэмы. Мне четырнадцать лет: Рифмы прозы. М.: Худож. лит., 1983. 463 с. С. 21-414. Д 5"DS/DD 1 # 109а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 20 в. Волошин М.А. Стихотворения. Волошин М.А. Стихотворения. Л.: Сов. писатель, 1977. 462 с. (Б-ка поэта. Малая сер. Изд. 3-е). С. 43-374. Д 5"DS/DD 1 # 111а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 20 в. Высоцкий В.С. Стихотворения. Высоцкий В. Нерв: Стихи. М.: Современник, 1982. 239 с. Д 5"DS/DD # 110а ASCII (КОИ-7 альтернативный).

2 Текст П Поэзия 19 в. Вяземский П.А. Стихотворения. Л.: Сов. писатель, 1958. 507 с. (Б-ка поэта. Большая серия, 2-е изд.). С. 49-403. Д 5"DS/DD 2 # 4, 18 58 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Галич А. Стихотворения. Галич А. Возвращение. ЛП ВТПО. Всесоюзное творческо-производ. объединение "Киноцентр". Лен. предприятие, 1989. Д 5"DS/DD 1 # 78 1 арх., послед. ASCII (КОИ - 7 альтернативный). Климова О.Г. 1990.

2 Текст П Поэзия 20 в. Гиппиус З.Н. Метафизика, личность, творчество. Сочинения. М.: Худ. лит. Д 5"DS/DD 1 # 208. 1 арх., послед. ASCII (КОИ-7 альтернативный). С перфолент (Ловля Е.Н., Шумилевская В.Н., 1992 г.).

2 Текст П Поэзия 20 в. Глазков Н.И. Стихотворения. Поэмы. Глазков Н.М. Избранное. М.: Худож. лит., 1989. 541 с. С. 12-528. Д 5"DS/DD 1 # 105а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 19 в. Гнедич Н.И. Стихотворения. Л.: Сов. писатель, 1956. 850 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 59-779. Д 5"DS/DD 3 # 29, 30,31 147 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Голодный М.С. Стихи. М.: Худож. лит., 1986. 286 с. С. 18-270 Д 5"DS/DD 1 # 86 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 20 в. Городецкий С. Стихотворения и поэмы. Л.: Сов. писатель, 1974. 624 с. (Б-ка поэта. Большая сер.). С. 55-417 Д 5"DS/DD 1 # 107а 1 арх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Гребенщиков В. Стихотворения. Д 5"DS/DD 1 # 78 1 арх., послед. ASCII (КОИ - 7 альтернативный). Климова О.Г. 1990.

2 Текст П Поэзия 19 в. Давыдов Д. Стихотворения. М.: Сов. Россия, 1979. С. 37-214 Д 5"DS/DD 1 # 20 9 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Дельвиг А. Стихотворения. Л.: Сов. писатель, 1951. 280 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 61-246. Д 5"DS/DD 1 # 16 45 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Дементьев А.Д. Стихотворения. Дементьев А.Д. Азарт. Стихи. М.: Сов. писатель, 1986. 222 с. (Б-ка произвед., удостоенных гос. премии СССР). С. 5-215. Д 5"DS/DD 1 # 114а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 18 в. Державин Г.Р. Стихотворения. Л.: Сов. писатель, 1957. 469 с. (Б-ка поэта. Большая сер. 2-е изд.). С. 77-360. Д 5"DS/DD 1 # 3 49 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 18 в. Дмитриев И.И. Стихотворения. Дмитриев И.И. Полное собрание стихотворений. Л.: Сов. писатель, 1967. 502 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 73 -414. Д 5"DS/DD 2 # 34, 35 79 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Друнина Ю.В. Стихотворения. Друнина Ю.В. Бабье лето: Новые стихи. М.: Сов. писатель, 1980. 102 с. С. 5-98. Д 5"DS/DD 1 # 113а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 19 в. Ершов П. Конек-горбунок. Стихотворения. Л.: Сов. писатель, 1951. 214 с. (Б-ка поэта. Малая сер., 2-е изд.). С. 57-193 Д 5"DS/DD 1 # 112а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Есенин С.А. Стихотворения. Д 5"DS/DD 1 # 7 63 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Жаров А.А. Стихотворения. Поэмы. Жаров А.А. Избранное. М.: Гослитиздат, 1950. 256 с. С. 3-253 Д 5"DS/DD 1 # 113а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 19 в. Жуковский В.А. Стихотворения. Том 1. Жуковский В.А. Стихотворения. Т. 1. Л.: Сов. писатель, 1939. (Б-ка поэта). С. 3-349. Д 5"DS/DD 3 # 1, 2, 8 153 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ, 1989-1990 гг.

2 Текст П Поэзия 19 в. Жуковский В.А. Стихотворения. Том 2. Жуковский В.А. Стихотворения. Т. 2. Л.: Сов. писатель, 1939. (Б-ка поэта). С. 5-462. Д 5"DS/DD 3 # 1, 2, 8 153 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ, 1989-1990 гг.

2 Текст П Поэзия 20 в. Заболоцкий Н.А. Стихотворения. Поэмы. Заболоцкий Н.А. Собрание сочинений: В 3 т. Т. 1. Столбцы и поэмы. 1926-1933. Стихотворения 1932-1958. Стихотворения разных лет. М.: Худож. лит. 1983. 655 с. Д 5"DS/DD 1 # 116а 1 арх., послед. ASCII (КОИ-7 альтернативный). USA. Кукушкина Е.Ю., 1990, Коробова М.М., 1991. ОСР. USA. D.Worth, R.Vroon, 1991.

2 Текст П Поэзия 20 в. Казин В.В. Стихотворения и поэмы. Казин В. Стихотворения и поэмы. М.: Худож. лит., 1964. 276 с. С. 13-271. Д 5"DS/DD 1 # 114а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 20 в. Каменский В. Стихотворения и поэмы. Каменский В. Стихотворения и поэмы. М.-Л.: Сов. писатель, 1966. 499 с. (Б-ка поэта. Больш. серия, 2-е изд.). С. 51-474. Д 5"DS/DD 1 # 63 72 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Кирсанов С.И. Стихотворения и поэмы 1923-1945. Кирсанов С. Избранные произведения. В 2 т. Т. 1. Стихотворения и поэмы 1923-1945. М.: Гослитиздат, 1961.

279 с. С. 5-277 Д 5"DS/DD 2 # 90, 108а 2 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Кирсанов С.И. Стихотворения и поэмы 1945-1960. Кирсанов С. Избранные произведения. В 2 т. Т. 2. Стихотворения и поэмы 1945-1960. М.: Гослитиздат, 1961. 287 с. С. 5-285 Д 5"DS/DD 2 # 90, 108а 2 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Клычков С.А. Стихотворения. М.: Худож. лит., 1985. 254 с. С. 23-248. Д 5"DS/DD 1 # 61 21 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Клжев Н.А. Стихотворения и поэмы. Архангельск: 1986. С. 23-238. Д 5"DS/DD 1 # 37 34 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Козлов И.И. Полное собрание стихотворений. Л.: 1960 г. С. 55-426. Д 5"DS/DD 1 # 118А 1 арх., послед. ASCII (КОИ-7 альтернативный). Сканирование.

2 Текст П Поэзия 19 в. Кольцов А.В. Стихотворения. Кольцов А.В. Полное собрание стихотворений. Л.: Сов. писатель, 1958. 305 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 45-255. Д 5"DS/DD 1 # 14 41 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Корнилов Б. Стихотворения и поэмы. Т. 1 и Т. 2. Пермь: Перм. кн. изд-во, 1986. Д 5"DS/DD 1 # 12 55 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 18-19 в. Крылов И.А. Стихотворения. Л.: Сов. писатель, 1954. 677 с. (Б-ка поэта. Большая сер. 2-е изд.). С. 49-602. Д 5"DS/DD 2 # 3, 6 79 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Кузмин М.А. Стихотворения. Кузмин М.А. Стихи и проза. М.: Современник, 1989. 431 с. С. 24-169. Д 5"DS/DD 1 # 114а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 19 в. Кюхельбекер В. Стихотворения. Л.: Сов. писатель, 1952. 426 с. (Б-ка поэта. Малая сер., 2-е изд.). С. 51-400. Д 5"DS/DD 1 # 98 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Лермонтов М.Ю. Стихотворения. Лермонтов М.Ю. Полное собрание сочинений. Т. 1. М.-Л.: Гослитиздат, 1947. 394 с. С. 21-321. Д 5"DS/DD 1 # 51 47 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1988-1989 гг.

2 Текст П Поэзия 19 в. Лермонтов М.Ю. Поэмы. Лермонтов М.Ю. Полное собрание сочинений. Т. 2. М.-Л.: Гослитиздат, 1947. 515 с. С. 7-494. Д 5"DS/DD 2 # 49, 50 65 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1988-1989 гг.

2 Текст П Поэзия 20 в. Луговской В.А. Стихотворения и поэмы. М.-Л.: Сов. писатель, 1966. 637 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 49-590. Д 5"DS/DD 2 # 64, 65 77 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Мандельштам О.Э. Стихотворения. Л.: Сов. писатель, 1973. 334 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 57-248 Д 5"DS/DD 1 # 59 21 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Мартынов Л. Стихотворения и поэмы. М.: Худож. лит., 1965. С. 104-657. Д 5"DS/DD 1 # 70 50 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Маяковский В.В. Стихотворения. Маяковский В.В. Избранные сочинения. В 2 т. Т. 1. М., Худож. лит. 1981. С. 53-514. Д 5"DS/DD 1 # 28 76 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Межиров А.П. Стихотворения. Межиров А. Прощание со снегом. Книга стихов. М.: Сов. писатель, 1964. 115 с. С. 5-112. Д 5"DS/DD 1 # 114а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 19 в. Михайлов М.Л. Стихотворения. Михайлов М.Л. Сочинения. В 3 т. Т. 1. Стихотворения. М.: Гослитиздат, 1958. 638 с. С. 51-527. Д 5"DS/DD 2 # 22, 26 71 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Некрасов Н.А. Стихотворения. Некрасов Н.А. Полное собрание стихотворений. В 3 т. Т. 2. Л.: Сов. писатель, 1976. (Б-ка поэта. Большая сер., 2-е изд.). 702 с. С. 7-590. Д 5"DS/DD 1 # 55 56 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Некрасов Н.А. Стихотворения. Некрасов Н.А. Полное собрание стихотворений. В 3 т. Т. 3 Л.: Сов. писатель, 1976. (Б-ка поэта. Большая сер., 2-е изд.). 512 с. С. Д 5"DS/DD 2 # 19, 44 99 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Никитин И.С. Стихотворения. Никитин И.С. Полное собрание стихотворений. М.-Л.: Сов. писатель, Ленингр. отд-ние, 1965. 613 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 52-564. Д 5"DS/DD 2 # 72, 73 # 88 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

3 Текст П Поэзия Никитина С.Е. (составитель корпуса). Русский фольклор: Духовн. стихи; Свад. причитания; Плачи. Сборник русских духовных стихов, составленный В.Варенцовым. Спб., 1860. Д 5"DS/DD 1 # 121 арх., послед. ASCII (КОИ - 7 альтернативный). МГУ. Реутт Т.Е.

3 Текст П Поэзия Никитина С.Е. (составитель корпуса). Русский фольклор: Духовные стихи. Калики переходные. Сборник стихов и исследование П.Безсонова. М., 1861. Д 5"DS/DD # 121 арх., послед. ASCII (КОИ-7 альтернативный). НИВЦ МГУ. Реутт Т.Е.

3 Текст П Поэзия Никитина С.Е. (составитель корпуса). Русский фольклор: Духовные стихи. Народная гравюра и фольклор в России 17-19 вв. М., 1976. Д 5"DS/DD # 121 ASCII (КОИ-7 альтернативный). НИВЦ МГУ. Реутт Т.Е.

3 Текст П Поэзия Никитина С.Е. (составитель корпуса). Русский фольклор: Духовные стихи. Русские письменные и устные традиции и духовная культура. Сборник. М., 1982. Д

5"DS/DD # 121 ASCII (КОИ-7 альтернативный). НИВЦ МГУ. Реутт Т.Е.

3 Текст П Поэзия Никитина С.Е. (составитель корпуса). Русский фольклор: Духовные стихи. Стихи духовные/ Сост. А.Ляцкий. Спб., 1912. Д 5"DS/DD # 121 арх., послед. ASCII (КОИ-7 альтернативный). НИВЦ МГУ. Реутт Т.Е.

3 Текст П Поэзия Никитина С.Е. (составитель корпуса). Русский фольклор: Плачи. Собрание Е.В.Барсова. М., 1872, ч. 1. Д 5"DS/DD # 121 арх., послед. ASCII (КОИ-7 альтернативный). НИВЦ МГУ. Реутт Т.Е.

3 Текст П Поэзия Никитина С.Е. (составитель корпуса). Русский фольклор: Плачи. Старообрядческие тексты. Рукописные сборники по материалам экспедиций. Д 5"DS/DD # 121 арх., послед. ASCII (КОИ-7 альтернативный). НИВЦ МГУ. Реутт Т.Е.

3 Текст П Поэзия Никитина С.Е. (составитель корпуса). Русский фольклор: Свадебные причитания. Рукописный сборник, составленные по материалам экспедиции МГУ 1974 г. в Пермскую обл. Д 5"DS/DD # 121 ASCII (КОИ-7 альтернативный). НИВЦ МГУ. Реутт Т.Е.

3 Текст П Поэзия Никитина С.Е. (составитель корпуса). Русский фольклор: Свадебные причитания. Русская свадьба. М., 1985. Д 5"DS/DD # 121 арх., послед. ASCII (КОИ-7 альтернативный). НИВЦ МГУ. Реутт Т.Е.

2 Текст П Поэзия 20 в. Окуджава В. Стихотворения. М.: Сов. писатель, 1985. С. 3-312. Д 5"DS/DD 1 # 77 1 арх., послед. ASCII (КОИ - 7 альтернативный). Климова О.Г. 1990.

2 Текст П Поэзия 19 в. Павлова К.К. Стихотворения. Павлова К. Полное собрание стихотворений. М.-Л.: Сов. писатель, 1964. 615 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 3-542. Д 5"DS/DD 1 # 94 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Пастернак Б.Л. Стихотворения и поэмы. Пастернак Б.Л. Избранное. В 2 т. М.: Худож. лит., 1985. Т. 1. Стихотворения и поэмы. 623 с. С. 30-525. Д 5"DS/DD 1 # 115а 1 арх., послед. ASCII (КОИ-7 альтернативный). USA. Кукушкина Е.Ю. (1990), Коробова М.М. (1991). ОСП. USA. D.Worth, R.Vroon 1991.

2 Текст П Поэзия 19 в. Плещеев А.Н. Стихотворения. Плещеев А.Н. Полное собрание стихотворений. М.-Л.: Сов. писатель, 1964. 430 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 53-360. Д 5"DS/DD 1 # 45 45 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Полежаев А.И. Стихотворения и поэмы. Л.: Сов. писатель, 1957. 473 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 39-420. Д 5"DS/DD 2 # 39, 41 65 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Полонский Я.П. Стихотворения. Л.: Сов. писатель, 1954. 547 с. (Б-ка поэта. Больш. серия, 2-е изд.). С. 50-492. Д 5"DS/DD 1 # 32 53 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Прокофьев А.А. Стихотворения. Том 1. Прокофьев А. Собрание стихотворений. В 2 т. Т. 1. М.-Л.:

Гослитиздат, 1961. 575 с. С. 19-563. Д 5"DS/DD 2 # 11, 15 103 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Прокофьев А.А. Стихотворения. Том 2. Прокофьев А. Собрание стихотворений. В 2 т. Т. 2. М.-Л.: Гослитиздат, 1961. 570 с. С. 7-558. Д 5"DS/DD 2 # 11, 15 103 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Пушкин А.С. Поэмы. Сказки. Пушкин А.С. Полное собрание сочинений. В 10 т. Изд. 4-е. Т. 4. Поэмы. Сказки. Л.: Наука, 1977. 447 с. С. 7-432. Д 5"DS/DD 1 # 111а 1 арх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 19 в. Пушкин А.С. Стихотворения 1813-1820. Пушкин А.С. Полное собрание сочинений. В 10 т. Изд. 4-е. Т. 1. Стихотворения 1813-1820. Л.: Наука, 1977. 480 с. С. 16-428. Д 5"DS/DD 1 # 46 35 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Пушкин А.С. Стихотворения 1820-1826. Пушкин А.С. Полное собрание сочинений. В 10 т. Изд. 4-е. Т. 2. Стихотворения 1820 (Юг) - 1826. Л.: Наука, 1977. 400 с. С. 7-352. Д 5"DS/DD 1 # 39 29 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Пушкин А.С. Стихотворения 1827-1836. Пушкин А.С. Полное собрание сочинений. В 10 т. Изд. 4-е. Т. 3. Стихотворения 1827-1836. Л.: Наука, 1977. 495 с. С. 7-432. Д 5"DS/DD 2 37, 38 40 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Рылеев К.Ф. Стихотворения. Рылеев К. Полное собрание стихотворений. Л.: Изд. Писатель. 1934. 575 с. С. 85-368. Д 5"DS/DD 1 # 4 58 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Сельвинский И.Л. Стихотворения. Поэмы. Сельвинский И.Л. Избранные произведения. В 2 т. Т. 1. Стихотворения. Поэмы. М.: Гослитиздат, 1956. 423 с. С. 29-417. Д 5"DS/DD 1 # 105а 1 арх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Симонов К.М. Стихотворения и поэмы. Л.: Сов. писатель, 1982. 623 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 71-546. Д 5"DS/DD 2 # 59, 60 52 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Случевский К.К. Стихотворения и поэмы. М.-Л.: Сов. писатель, 1962. 467 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 55-40. 4 Д 5"DS/DD 1 # 106а 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 20 в. Твардовский А.Т. Стихотворения. Поэмы. Твардовский А.Т. Собрание сочинений. В 4 т. Т. 2. Фронтовая хроника. Василий Теркин. М.: Гослитиздат, 1959. 375 с. С. 2-371. Д 5"DS/DD 3 # 84, 91, 108а 3 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Твардовский А.Т. Стихотворения. Поэмы. Твардовский А.Т. Собрание сочинений. В 4 т. Т.1. Из ранних стихотворений. Сельская хроника. Страна Муравия. М.: Гослитиздат, 1959. 367 с. С. 21-363. Д 5"DS/DD 3 # 84, 91,

108а 3 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Твардовский А.Т. Стихотворения. Поэмы. Твардовский А.Т. Собрание сочинений. В 4 т. Т.3. Дом у дороги. Стихотворения. За далью даль. М.: Гослитиздат, 1959. 359 с. С. 2-356. Д 5"DS/DD 3 # 84, 91, 108а 3 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Тихонов Н.С. Стихи и поэмы. Тихонов Н. Избранное. Стихи и поэмы. М.: Сов. писатель, 1948. 350 с. (Б-ка избранных произведений советской литературы 1919-1947). С. 7-342. Д 5"DS/DD 1 # 71 # 16 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Толстой А.К. Стихотворения. Толстой А.К. Полное собрание стихотворений. М.: Сов. писатель, 1937. 809 с. (Б-ка поэта). С. 55-714. Д 5"DS/DD 2 # 24, 26 73 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Тургенев И.С. Стихотворения, поэмы. Тургенев И.С. Собрание сочинений. Т. 10. М.: Правда, 1949 (Б-ка "Огонек"). С. 5-140. Д 5"DS/DD 1 # 90 1 арх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1990 г.

2 Текст П Поэзия 19 в. Тютчев Ф.И. Стихотворения. М.: Худож. лит., 1984. С. 49-435. Д 5"DS/DD 1 # 36 36 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Фет А.А. Стихотворения. Фет А.А. Полное собрание стихотворений. Л.: Сов. писатель, 1959. 897 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 81-681. Д 5"DS/DD 2 # 20, 21 81 неарх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 19 в. Фофанов К.М. Стихотворения и поэмы. М.-Л.: Сов. писатель, 1962. 334 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 49-290. Д 5"DS/DD 1 # 34 32 неарх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989-1990.

2 Текст П Поэзия 20 в. Хлебников В. Стихотворения. Поэмы. Хлебников В. Творения. М.: Сов. писатель, 1986. Д 5"DS/DD 1 # 129-132 арх., послед. ASCII (КОИ-7 альтернативный). УПЦ 1989. Цыб Е.А. 1989. ЕТС.

2 Текст П Поэзия 20 в. Черный С. Стихотворения. Л.: Сов. писатель, 1960. 627 с. (Б-ка поэта. Большая сер., 2-е изд.). С. 72-539. Д 5"DS/DD 1 # 89 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

2 Текст П Поэзия 19 в. Языков Н.М. Стихотворения и поэмы. Л.: Сов. писатель, 1988. 591 с. (Б-ка поэта. Большая сер., 3-е изд.). С. 47-498. Д 5"DS/DD 1 # 95 1 арх., послед. ASCII (КОИ - 7 альтернативный). УПЦ 1991.

Дополнение к Архиву Машинного фонда русского языка

Д. Г. Мальцева, Германия: страна и язык. Лингвострановедческий словарь. Москва, Русские словари, 1998, 384 стр., 40 а. л.

Е. М. Поспелов, Географические названия мира. Топонимический словарь. Москва, Русские словари, 1998, 504 стр., 40 а. л.

- Г. Я. Солганик, Стилистический словарь публицистики, Москва, Русские словари, 1998, 672 стр., 40 а. л.
- В. С. Елистратов, Словарь крылатых слов русского языка. Русский кинематограф. Москва, Русские словари, 1998, 200 с., 40 а. л.
- В. С. Елистратов, Словарь московского арго. Издание 2-е, переработанное и дополненное. Более 9 000 единиц. Объем более 50 а. л.
- Колл. авт., Материалы к большому словарю крылатых слов русского языка. Более 5000 единиц. Объем около 80 а. л.
- Ю. Карабчиевский, Воскресение Маяковского. Эссе. 20 а. л.
- М. М. Бахтин, Собрание сочинений. Т. 2. 40 а. л.
- М. М. Бахтин. Собрание сочинений. Т. 5, работы 1940-1960 гг. Объем вместе с комментариями порядка 50 а. л.
- В. В. Крестовский, Петербургские трущобы.
- П. И. Мельников-Печерский, В лесах, На горах.
- А. Н. Островский, Собрание сочинений, тт. 1, 3, 4.
- Е. А. Салиас де Турнемир, Пугачевцы.
- Н. Абрамов. Словарь русских синонимов и сходных по смыслу выражений, 40 а. л. формат Word5
- Н. А. Петровский. Словарь русских личных имен, 20 а. л.
- Ю. А. Федосюк. Русские фамилии, объем 12 а. л.
- С. М. Федченко. Словарь русских созвучий, объем порядка 30 а. л.
- Л. Г. Кочедыков. Словарь иноязычных фразеологизмов, объем 5 а. л.
- В. М. Мокиенко, А. М. Мелерович. Фразеологизмы в русской речи, объем 45 а. л., формат WinWord
- Орфоэпический словарь русского языка, 5 издание. Объем 70 а. л.
- Андрей Белый о Блоке. Воспоминания, статьи, речи. Подготовка текста, комментарий А. В. Лаврова, объем 45 а. л.
- Георгий Иванов. Распад атома. Избранные стихи, проза, мемуаристика, литературная критика. Комментарии С. Ивановой. Объем 45 а. л.
- Анна Вежбицкая. Язык, познание, культура. Сборник переводов. Объем 30 а. л.
- В. А. Плунгян. Почему языки такие разные. Объем 20 а. л., формат WinWord
- И. Прыжов. Двадцать шесть московских дур, дураков, пророков и юродивых и другие труды по русской истории и этнографии, объем 20 а. л.
- Кроме того, от агентства ПОСТФАКТУМ получен и предварительно обработан массив текстов объемом 2 млн. словоупотреблений.

III. Текстообработывающие системы UNILEX-T и UNILEX-M

Существуют несколько последовательных версий текстообработывающих систем семейства UNILEX:

- 1) **UNILEX-T** — с 1991 года;
- 2) **MAK** — с 1995 года;
- 3) **UNILEX-M** — с 1999 года.

Система UNILEX-T

UNILEX-T — это пакет программ, обеспечивающих получение из текстов различной первичной информации по запросу пользователя. Если Вы начинаете работать с новым текстом, то прежде всего надо получить словник со статистикой, для чего текст должен быть оформлен соответствующим образом.

После этого с ним можно работать в так называемом пакетном режиме, когда запросы и результаты содержатся в файлах. В этом случае элементом запроса может быть лемма, грамматическая помета, словоформа, адрес. Адрес словоформы содержит имя модуля, маркеры фрагментов, в которые входит это слово, и номер слова внутри последнего нумеруемого фрагмента. Это дает возможность из полного словника делать выборки по группе модулей или фрагментов и получать по ним частотные словники и конкордансы.

Запрос состоит из одного или нескольких перечисленных элементов, причем любой из них может быть частично маскирован, т. е. задан с помощью шаблона. В частности, можно дать задание на получение полного словника по всему корпусу текстов. Элементы запроса, перед которыми стоит отрицание, указывают на те слова, которые не должны входить в выходной словник. В ответ на запрос создаются два файла: частотный словник и конкорданс.

Длина контекста определяется пользователем, при нулевой длине вместо конкорданса создается словоуказатель.

Пакетный режим предполагает последовательный просмотр всего исходного словника. Чтобы иметь возможность получать контексты быстро на экране дисплея по любому элементу словника, надо создать индекс словника. В этом случае можно обращаться к базе данных по лемме, части речи, словоформе, частоте, рангу.

Программное обеспечение автоматического конкорданса позволяет осуществлять следующие функции:

- получение списка текстформ с адресами;
- лемматизация;
- получение статистической информации;
- упорядочение словника по различным параметрам;
- получение полных или частичных словников, словоуказателей и конкордансов;
- загрузка словоуказателя в базу данных;
- диалоговая работа с базой данных.

Исходный текст представляет собой текстовый файл в кодах ASCII. Длина строки — не более 80 символов. Желательно, чтобы даже после вставки и нумерации маркеров длина строки не превышала 80.

Единицей обработки текста является модуль, длина которого не должна превосходить 60 тыс. символов, включая пробелы. Что представляет собой модуль содержательно — определяет пользователь. Это может быть отдельное произведение, глава, выборка.

Модуль должен быть оформлен следующим образом. В начале модуля, с начала строки ставится символ *, за которым без пробела идет имя модуля — не более 8 символов. В имени модуля могут быть зашифрованы какие-либо значимые для пользователя данные: жанр, автор, время, место создания текста и пр. Хорошо, если каждая позиция имени модуля будет предназначена для обозначения какого-либо значимого для пользователя признака. Тогда в дальнейшем, задавая в запросе имя модуля или его шаблон, можно будет получать выборки из полного словника в виде одного или группы модулей.

Признаком конца модуля служит последовательность символов &&& в начале строки.

Особым модулем является модуль, имя которого состоит из одних нулей. В этом модуле содержится схема модуля (см. п. **Схема модуля**), включающая в себя описание маркеров и атрибутов. В соответствии с описанием маркеров формируется адрес слова. Если маркеры не заданы, то адрес слова будет содержать только номер модуля и номер слова внутри модуля.

При помощи атрибутов задаются наименование корпуса текстов и знаки препинания. Если знаки препинания не заданы, то ими считаются все знаки, кроме букв и цифр.

Максимальное число модулей в одном корпусе — 1000.

Максимальное число строк в стоп-словаре — 80.

Максимальное число разных маркеров в модуле — 20.

Максимальная длина имени модуля — 30 символов.

Максимальная длина модуля — 60000 байтов.

Система используется в следующих организациях:

- Институт русского языка РАН (В. П. Григорьев);
- Институт языкознания РАН (С. Е. Никитина);
- НИВЦ МГУ (О. А. Казакевич);
- СПбГУ, кафедра математической лингвистики (г. Санкт-Петербург);
- СГУ (г. Сыктывкар, С. В. Лесников);
- СГУ (г. Саратов, В. Е. Гольдин);
- Лаборатория RAO (г. Париж, Ж. Мартиновский).
- Переводческий факультет Йёенсуунского университета (г. Савонlinna, Финляндия).

Система “Автоматический конкорданс МАК”

Автоматический конкорданс МАК предназначен для получения автоматических конкордансов на персональных компьютерах в среде DOS. При обработке текстов можно указать, какие слова и фрагменты текста не надо включать в словник. Можно также внести в текст различного рода пометы, которые не войдут в словник, но по ним также возможен доступ к тексту. При работе с конкордансом в него можно входить по словоформам, частоте и маркерам. Кроме того, запрос можно сделать в виде файла, тогда наряду с ключевым словом можно задать сочетания слов. В системе предусмотрена возможность осуществления автоматической лемматизации (с последующей коррекцией результатов), в этом случае вход будет возможен также по леммам и частям речи. В системе МАК возможно создание и работа с параллельными конкордансов. Полученные на экране контексты можно увеличивать и уменьшать, а также отсылать понравившийся контекст в файл. Можно также записывать контексты на отмеченные слова, не просматривая их предварительно.

В простейшем случае текст может быть безо всякой пользовательской разметки, но это всегда должен быть так называемый ASCII-файл, т. е. в нем не должно быть скрытого оформления, которое делают некоторые редакторы (например, Word). Можно использовать Multi-Editor или Лексикон, но без шрифтовых выделений.

При необходимости пользователь может отметить те фрагменты текста, которые не должны войти в словник. Для этого в качестве начального и конечного маркеров надо выбрать два символа, не встречающиеся в тексте, и вставить один из них в начале, а другой — в конце фрагмента. Эти же маркеры надо задать в стоп-словаре в таком виде: <начальный маркер><пробел><конечный маркер>. В качестве начального и конечного маркера может быть взят один и тот же символ. Расширение имени стоп-словаря — .stf. В этом же файле могут быть заданы слова, которые не надо включать в словник, причем может быть задано только начало или только конец слова. В этом случае на месте недостающей части слова ставится звездочка.

Кроме того, в текст можно вставить какие-либо пометы. Это могут быть, например, какие-то грамматические характеристики, номера предложений или любые пометы, которые должны войти в словник, но не должны смешиваться с обычными словами. Такие псевдо-слова также должны быть отмечены символом, не встречающимся в тексте. Этот символ должен стоять перед пометой, примыкая к ней без пробела. Концом пометы считается пробел. Длина пометы, как и обычного слова, не должна превышать 40 символов. Использование помет должно быть зафиксировано в файле конфигурации (расширение .cnf). Каждая строка файла конфигурации должна содержать один маркер пометы.

Если отвлечься от того, что при создании Автоматического конкорданса возможно при обнаружении какой-либо неправильности повторное выполнение одних и тех же шагов, то можно считать, что создание конкорданса производится однократно. И делается это для того, чтобы потом многократно можно было работать с Автоматическим конкордансом.

Для получения параллельного конкорданса надо соответствующим образом подготовить параллельные тексты. Параллельные тексты получаются, например, при переводе какого-либо текста с одного языка на другой или при пересказе текста на том же самом языке. Заметим, что параллельных текстов может быть не только два, но и больше, например, несколько переводов разных авторов одного и того же произведения. В Автоматическом конкордансе МАК предусмотрена возможность работы с параллельными текстами, количество которых не должно превышать 256. Главная предварительная работа, которая должна быть проделана с параллельными текстами, — это установление соответствия между фрагментами разных текстов.

Это достигается тем, что фрагменты нумеруются таким образом, что соответствующие фрагменты разных текстов получают одинаковые номера.

Поскольку номера фрагментов не должны входить в нормальный словарь, они помечаются маркерами, которые ставятся вплотную перед номером, не разделяясь пробелом (но после номера должен быть пробел обязательно). Например, номера фрагментов в тексте могут выглядеть так: ^1, |1a, &1.1, <1-2> и пр. При этом маркер, стоящий перед номером, должен быть указан в файле конфигурации (каждый маркер на отдельной строке). Каждый текст загружается независимо от других. В дальнейшем, получив Автоматический конкорданс размеченных таким образом текстов, можно работать с ним и как с обычным конкордансом, и как с параллельным. При работе с обычным конкордансом максимальный размер контекста достигает 20 строк экрана, а при работе с параллельным конкордансом верхнюю половину экрана занимает контекст первого текста, а в нижней половине можно просматривать соответствующие контексты остальных параллельных текстов.

Основные отличия нового варианта автоматического конкорданса (МАК) от прежнего (UNILEX-T) заключаются в следующем.

1. Текст может быть без всякой разметки.
2. Если все-таки в текст будут внесены какие-либо маркеры (а это допускается), то они не войдут в адрес слова, как это было в UNILEX-T, а будут иметь ссылку на текст, как все обычные слова.
3. Вход в конкорданс возможен не только по слову, частоте и части речи, но и по маркерам.
4. К существующему корпусу текстов можно добавлять новые тексты. Всего в корпусе текстов может быть до 256 текстов.
5. Возможно создание параллельного конкорданса, причем общее число параллельных текстов должно быть не более 256.

Система используется в следующих организациях:

- НИВЦ МГУ (Н. Н. Леонтьева);
- СПбГУ, кафедра математической лингвистики (г. Санкт-Петербург);
- Переводческий факультет Йёнсуунского университета (г. Савонлинна, Финляндия).

Система создавалась при поддержке Института Открытое Общество (грант 1176/94).

Система UNILEX-M

Текстоориентированная система UNILEX-M предназначена для получения конкордансов и словарей по текстам в операционной среде Windows.

Она состоит из двух частей:

- Программа TSelect позволяет компоновать корпус текста по заданным признакам из газетных текстов.
- Программа TAM включает в себя возможности поучать словники и конкордансы, осуществлять лемматизацию текстов и словников и генерацию словоформ.

Программа TSelect

При запуске программы TSelect появляется меню, с помощью которого можно выбрать тексты и задать параметры для отбора текстов:

- название газеты,
- тема публикации,
- дополнительные параметры,
- время публикации (период).

Списки этих параметров можно посмотреть в файле TSelect.cfg.

Программа TAM

Программа TAM предназначена для получения контекстов на заданные слова (конкордансов) и словников. Программа запускается файлом TAM.BAT, где сначала загружается в память морфологический словарь (используемый для лемматизации и генерации словоформ), а затем собственно программа TAM.EXE.

При получении конкорданса слово в начале и в конце может быть маскировано звездочкой. Можно задать поиск по одновременной встречаемости нескольких слов. Запрос может быть задан в файле. Там же может быть указан формат файла с текстом: заголовки (до 5 уровней), комментарии (в комментариях запрошенные слова не ищутся), и длина контекста. Заголовки указываются при контекстах в качестве адреса.

Можно получить следующие словники по тексту: алфавитный, частотный и обратный. Во всех случаях при слове указана его абсолютная частота.

Если файл запроса не задан, то для указания параметров запроса при нажатии кнопки “Start” выдается диалоговое окно, в котором можно их все задать.

Имеется возможность генерации словоформ в файле запроса. В этом случае в файле запроса задается лемма, а затем по ней генерируются все словоформы этого слова. Полученный результат используется в качестве запроса. Для генерации должен быть указан файл конфигурации, в котором указываются файлы словаря, используемого при генерации словоформ. Если он не указан, то файл с именем `Umllemma.cfg` берется из текущей директории. Этот же файл используется при лемматизации. Лемматизировать можно словники или конкордансы с длиной контекста, равной нулю.

Программа может запускаться с параметрами:

- 1) файл конфигурации,
- 2) файл(ы) текстов,
- 3) файл запроса,
- 4) выходной файл.

Признаки каждого файла при задании параметров такие:

- /g — файл конфигурации;
- /t — файл(ы) текстов;
- /r — файл запроса;
- /o — выходной файл, содержащий конкорданс.

Могут быть опущены от одного до всех параметров. Входные файлы могут быть заданы шаблоном (типа: `C:\TXT*.TXT`). Если заданы 2-ой и 3-ий параметры, то работа идет в пакетном режиме. Если параметры не заданы, то выдается меню и потом окно диалога. Через меню выбирается входной файл, можно выделить несколько файлов.

Если не задано имя выходного файла, то его имя будет последовательно `_concord.000`, `_concord.001` и т.д. Результат (контексты) записываются в этот файл.

Словники сохраняются в файлах с именами:

- алфавитный — `_voc_abc.000`, `_voc_abc.001` и т. д.
- частотный — `_voc_frq.000`, `_voc_frq.001` и т. д.
- обратный — `_voc_tur.000`, `_voc_tur.001` и т. д.

По умолчанию разделители заголовков `<H1>`, `<H2>` и т. д. В файле запроса их можно переопределить: `<H1>` <реальный заголовок>. Если в качестве длины контекста указать точку, то размер контекста будет равен предложению. Идентификатор длины контекста `<NS>`.

Пример файла запроса (справа — пояснения):

`<h1>_1` Заголовок первого уровня: `_1`
`<H2>_2` Заголовок второго уровня: `_2`

`<NS>` . Длина контекста — предложение

`*a` Слова, оканчивающиеся на “-а”

`слов*` Слова, начинающиеся на “слов-”

Литература

1. Аношкина Ж. Г. *Автоматизированная лексикографическая система УНИЛЕКС-2*. М.: Изд-во МГУ, 1989. 107 с.
2. Аношкина Ж. Г. *Текстовая база данных в автоматизированной лексикографической системе УНИЛЕКС* // Третья Всесоюзная конференция по созданию Машинного фонда русского языка. Тезисы докладов. М., 1989. С. 110–112.
3. Аношкина Ж. Г. *Один способ лемматизации для автоматического конкорданса* // Актуальные проблемы компьютерной лингвистики. Тезисы докладов Всесоюзной конференции в гор. Тарту 29–31 мая 1990. Тарту, 1990. С. 109 (соавтор: Е. А. Аношкин)
4. Аношкина Ж. Г. *Применение принципа аналогии при морфологическом анализе и синтезе* // Использование ЭВМ в научной и учебной работе гуманитарных вузов. Тезисы докладов. Минск, 1990. С. 75–76 (соавтор: Т. В. Марьян)
5. Аношкина Ж. Г. *Проблема лемматизации при автоматизированном получении словарей из текстов* // Материалы III Всесоюзной конференции по созданию Машинного фонда русского языка. М., 1990. С. 4–16.
6. Аношкина Ж. Г. *Омонимия словоформ в русском языке* // Проблемы создания машинных фондов национальных языков. Тезисы докладов. Киев, 1991. С. 5–6.
7. Аношкина Ж. Г. *Лингвистический программно-источниковый пакет УНИЛЕКС+*. *Тексто-ориентированная компонента УНИЛЕКС-Т* // Бюллетень Машинного фонда русского языка. Москва, 1992, вып. 2. С. 3–7.

8. **Аношкина Ж. Г.** *Подготовка частотных словарей и конкордансов на компьютере (пособие для филологов)* // Москва, 1995. 60 с.
9. **Аношкина Ж. Г.** *Тексто-ориентированная компонента АЛС УНИЛЕКС (УНИЛЕКС-Т)* // Альманах "Говор". Сыктывкар, 1995. С. 7–29.
10. **Аношкина Ж. Г.** *Морфологический процессор русского языка* // Бюллетень Машинного фонда русского языка. Москва, 1996, вып. 3. С. 53–57.
11. **Аношкина Ж. Г.** *О включении российских лингвистических ресурсов в систему EMIR (Европейский Многоязычный Информационный Поиск)* // Бюллетень Машинного фонда русского языка. Москва, 1996, вып. 3. С. 58–60.

**IV. Лексикографическая программная система
UNILEX-D:
процесс создания, принципы работы,
применение в словарных проектах**

Общая характеристика проекта

Лексикографическая программная система UNILEX-D, созданная в отделе Машинного фонда русского языка Института русского языка РАН под руководством В. М. Андриященко, позволяет конструировать словарные базы данных различных по структуре филологических словарей и производить над ними всевозможные автоматизированные операции: создавать указатели, индексы, подсловари, вычислять количественные параметры, конвертировать словарь из базы данных в оригинал-макет и т.д.

Идея универсальной автоматизированной лексикографической системы в отечественной вычислительной лингвистике впервые была изложена В. М. Андриященко в работе “Автоматизированная лексикографическая система UNILEX”. [Андриященко, 1982]

Основным принципом работы системы UNILEX-D является принцип управления словарными базами с помощью универсальной структуры словарной статьи. При создании базы данных в системе UNILEX-D структура филологического словаря описывается по единым формальным правилам. Любая операция над словарем в системе совершается с помощью описания его структуры, выступающей в роли своеобразного метаязыка.

По принципам UNILEX-D в Машинном фонде русского языка были созданы компьютерные базы данных для 7-ми словарей русского языка; для двух из них — “Русского орфографического словаря” и “Словаря языка русской поэзии XX века” — система UNILEX-D использовалась при подготовке оригинал-макета.

На систему программ UNILEX-D (под названием LEXPRO — лексикографический процессор) в 1995 году было выдано Свидетельство об официальной регистрации программы для ЭВМ за № 950266 в Российском агентстве по правовой охране программ для ЭВМ, баз данных и топологий интегральных микросхем (РосАпо).

Принципы формального описания структуры словарной статьи, принятые в UNILEX-D, применялись в диссертации М. Г. Пшеничной для описания статьи толкового словаря русского языка В. И. Даля [Пшеничнова, 1990].

Принципы создания автоматического словаря, на которых базируется система UNILEX-D, были положены в основу проекта “Автоматизированный словарь русских говоров” [Загоровская, Лесников и др., 1996; Лесников, 1996]

Описание системы UNILEX-D входит в каталог российских коллективов и разработок “Автоматизированная обработка речи и текстов на естественном языке”, созданный В. Э. Семеновой [Semёnova, Fluhr, 1997]. Информация о каталоге представлена также в сети Интернет, на сайте: <http://members.aol.com/fluhrsciper/catalogue>.

Этап электронно-вычислительных машин 3-го поколения

Проект создания универсальной лексикографической системы, общие контуры которой впервые были изложены в статье [Андриященко, 1982], реализовывался вначале в Лаборатории вычислительной лингвистики НИВЦ МГУ, а затем в отделе Машинного фонда русского языка ИРЯ РАН с начала 80-х годов.

Созданию программной системы UNILEX-D предшествовал этап изучения закономерностей структуры словарных статей филологических словарей различных жанров и типов. По результатам этих исследований в 1987 году была защищена кандидатская диссертация “Структура словарного текста в аспекте машинной лексикографии” [Колодяжная, 1986]. Основные результаты этого этапа опубликованы в статьях [Колодяжная 1987, Колодяжная 1988], в методической разработке [Колодяжная и др., 1986] и в монографии [Колодяжная, 1987].

Первая версия программной системы UNILEX-D была создана к 1987 году. Система была написана на языке программирования ПЛ-1 для электронно-вычислительных машин 3-го поколения. Система программ UNILEX- была протестирована на экспериментальных базах данных для различных филологических словарей.

Начиная с 1983 года, с даты начала осуществления проекта по созданию Машинного фонда русского языка [Машинный фонд русского языка: идеи и суждения, 1986], развитие системы

UNILEX-D продолжалось в рамках проекта “Автоматизация лингвистических исследований и создание Машинного фонда русского языка”.

Концепция и архитектура Машинного фонда русского языка подробно изложена в статьях и книге В. М. Андрущенко [Андрущенко, 1986, 1987, 1989]. В общей архитектуре Машинного фонда автоматизированные словарные системы являются, наряду с текстовыми конкордансами, грамматиками и т. п., одними из составляющих, оформленных как Лингвистические программно-источниковые пакеты. Словарный Лингвистический программно-источниковый пакет включает в себя базу данных словаря, организованную по единым для всех словарей правилам, и систему управляющих программ. Система UNILEX-D явилась одной из таких возможных систем управления словарями.

На этапе электронно-вычислительных машин 3-го поколения в рамках проекта был создан один такой Лингвистический пакет — многоязычный терминологический словарь “ТЕРМИН”. Уже на этом этапе при работе со Словарем под управлением системы UNILEX-D выполнялись практически все функции — автоматический ввод статей в базу данных из исходного, минимально размеченного текста, прямой доступ к информации в режиме диалога, редактирование статей в базе данных, выполнение различных операций над Словарем (выборки, проекции, вычисление количественных таблиц, объединение двух словарей). Подробное описание автоматизированного словаря “ТЕРМИН” можно найти в руководстве [Колодяжная и др., 1986].

Также на этом этапе были созданы экспериментальные словарные базы данных некоторых словарей русского языка, на которых была отработана методика построения формального описания словарных статей и создания словарных баз данных, работающих под управлением системы UNILEX-D. Была создана такая база данных для Малого академического словаря русского языка (МАС), в которую входило более 1000 статей (для буквы А). База данных Малого академического словаря описана в статье [Колодяжная, “Опыт создания...”, 1987]. Для “Грамматического словаря русского языка” А. А. Зализняка была создана база данных на материале 9000 статей (буквы А–В); результаты этой работы описаны в статье [Колодяжная, “Автоматический Генеральный словарь русских словарей”, 1988].

Работа по созданию словарных баз данных на электронно-вычислительных машинах 3-го поколения была прекращена с переводом работ в области вычислительной лингвистики в 1988 году на персональные компьютеры.

Этап персональных компьютеров

Реализация проекта по созданию Машинного фонда русского языка и, соответственно, его компонентов — словарно-ориентированных Лингвистических программно-источниковых пакетов на персональных компьютерах развивалась в нескольких направлениях. Во-первых, возникла необходимость перевода всех лингвистических источников на новые носители информации — дискеты и диски ПЭВМ. Во-вторых, пришлось заново программировать уже на языках персональных ЭВМ прежние программные системы, в том числе, UNILEX-D. Наконец, возникла необходимость в пересмотре некоторых положений, связанных с подготовкой и вводом информации в условиях новых технических возможностей.

В этих новых технических условиях были пересмотрены и модернизированы некоторые принципы системы UNILEX-D, в частности, упрощены принципы описания словарных статей. Изменилась структура программ системы UNILEX-D и сами программные модули были заново написаны на языке CLIPPER для персонального компьютера. Новая версия UNILEX-D для персонального компьютера была завершена к 1992-му году. В период с 1992 по 1995 год были внесены еще некоторые дополнения и изменения, связанные с применением системы к созданию базы данных “Русского орфографического словаря”.

В период с 1988 по 1999 годы были созданы словарно-ориентированные Лингвистические программно-источниковые пакеты для различных русских словарей. Они описаны ниже, в разделе “Применение в словарных проектах”.

Описание системы UNILEX-D

Введение

Среди задач, которые должна решать программная лексикографическая система, можно выделить следующие.

Во-первых – создание автоматических словарей в такой форме, чтобы компьютерные представления различных по структуре словарей основывались на одних и тех же принципах. Решение такой задачи осуществимо на пути изучения общих структурных особенностей словарных статей различных по жанру словарей.

Решение первой задачи подводит к решению второй задачи — всевозможным преобразованиям филологических словарей программным путем. При решении этой задачи необходимо определить для различных словарей общие типы их лексикографических преобразований.

Наконец, в круг задач универсальной словарной системы должна входить и подготовка компьютерного словаря к изданию: автоматизация процесса подборки шрифтов для различных зон словарной статьи и конвертирование базы данных словаря в издательский оригинал-макет.

В настоящей работе дается описание действующей универсальной программной лексикографической системы UNILEX-D, решающей вышеизложенные задачи.

Основные принципы работы универсальной лексикографической системы UNILEX-D изложены также в статьях [Колодяжная, 1990; Колодяжная, 1992; Колодяжная, 1995, Колодяжная, 1996].

Принципы построения универсальной лексикографической системы

Основным принципом системы UNILEX-D является принцип управления словарными базами с помощью структуры словарной статьи. Структура словарной статьи описывается списком элементарных и составных компонент статьи, идущих в последовательности, принятой в словаре. Каждая элементарная компонента описывается мнемоническим именем, задаваемым филологом и расшифровкой этого имени. Список составных компонент определяется филологом и, как правило, описывает те блоки, в которые естественно собираются элементарные компоненты.

Режим проектирования структуры статьи в системе UNILEX-D позволяет филологу использовать при описании нового словаря типовые структуры словарей (толкового, синонимов, грамматического и т. д.), а также редактировать описание структуры. Режим ввода новых статей осуществляется несколькими путями. При вводе статей с рукописи (картотеки) система обеспечивает филологу ввод по формату. При таком вводе в качестве подсказки на экране компьютера филологу высвечивается список компонент статьи, объявленный при создании словаря. Исходный формат филолог может при необходимости расширять (в случае повторяющихся компонент), или сжимать (в случае факультативных компонент). Другой возможностью пополнения компьютерного словаря является загрузка новых статей из текстового массива (файла), в котором каждая статья расписана покомпонентно, в соответствии с объявленной структурой.

Созданный компьютерный словарь можно редактировать, преобразовывать в новые словари, вычислять количественные параметры словаря и подготавливать его оригинал-макет для издания. Операции над словарями в лексикографической системе являются общими для словарей различных жанров и осуществляются для каждого словаря с помощью описания структуры его статьи.

В универсальной лексикографической системе выделяются четыре общих типа преобразований словаря. Общее количество преобразований (в рамках 4-х типов), которые можно связать с конкретным словарем, обусловлено структурой его статьи, количеством объявленных в словарной статье компонент.

Множество всех возможных преобразований словарной информации, определяющееся структурой словаря, образует лексикографическое пространство, позволяющее филологу исследовать лингвистические факты, содержащиеся в информационном массиве, с различных точек зрения (возможности автоматизации преобразований исходных словарей в новые посвящена статья [Kolođajzaja, Polikarov, 1990]; пример построения лексикографического пространства для словаря-модели можно найти в статье “Понятие лексикографического пространства” [Колодяжная, 1996]).

Преобразованием словаря называется такая совокупность словарных статей, которая построена из исходного словаря путем “рассыпания” его статей на составные компоненты, удаления ненужных компонент, сортировки по значениям какой-либо компоненты, объявленной как новый “вход”, и сборки при каждом новом “входе” оставшихся компонент в качестве новых статей.

Следуя этому определению, выделяются следующие типы словарных преобразований (для одного словаря): **таблица, инверсия, проекция и выборка**.

Таблицей называется такое преобразование словаря, при котором в каждой статье оставляется только одна заданная компонента, список значений этой компоненты сортируется по алфавиту и для каждого значения вычисляется, сколько раз оно встретилось в словаре. Таблицы показывают количественное распределение значений каждой компоненты, объявленной в структуре статьи. Потенциально для данного словаря можно составить столько таблиц распределения, сколько компонент объявлено в его структуре.

Инверсией называется такое преобразование словаря, при котором в каждой статье оставляются две компоненты, одна из них объявляется “входом”. Компоненты сортируются так, чтобы собрать при каждом значении нового входа список тех значений второй компоненты, которые входили с ним в одну и ту же исходную статью. К одному словарю потенциально можно составить столько инверсий, сколько различных пар компонент можно образовать для данной структуры словаря.

Проекцией называется такое преобразование словаря, при котором в каждой статье оставляются заданные типы компонент, но порядок компонент и порядок статей не меняются.

Выборкой называется такое преобразование, при котором в каждой статье, так же, как в проекции, оставляются заданные типы компонент, но одна из компонент объявляется новым “входом” и в новый словарь попадают статьи только с определенными значениями нового входа. Потенциально можно составить столько типов выборки, сколько типов компонент объявлено в структуре статьи.

Ниже приводится описание всех функций действующей программной лексикографической системы UNILEX-D.

Общие сведения о лексикографической системе

Лексикографическая система UNILEX-D является программной системой, позволяющей создавать филологический словарь в компьютерной форме, выполнять все возможные для данного словаря преобразования, объединять информацию из двух различных словарей, работать со словарями в интерактивном режиме, а также подготавливать для издания оригинал макет словаря.

Основным принципом системы UNILEX-D является принцип управления словарными базами с помощью структуры словарной статьи, которую проектирует сам филолог. Структура словарной статьи описывается списком элементарных и составных компонент. Режим создания позволяет филологу использовать при описании нового словаря структуры словарей известных типов и жанров, а также редактировать новый словарь с помощью информации базовых словарей. Словарную базу филолог может просматривать в диалоговом режиме, образуя входы в словарь от любой из объявленных компонент. При этом в другом окне экрана филолог может просматривать информацию из другого словаря или какие-либо преобразования данного словаря.

Операции создания новых словарей из базовых, а также исследовательские операции поддерживаются специальным блоком, включающим типовые функции словарных преобразований: инверсии, проекции, выборки, объединения и пересечения словарей. При создании нового словаря из базовых словарей UNILEX-D обеспечивает филологу возможность проектирования структуры нового словаря.

Работа филолога в UNILEX-D распределена по трем блокам: каждый из которых выполняет конкретные функции создания, редактирования, преобразования словаря, работы в диалоговом режиме, подготовки словаря к изданию. Ниже приведен перечень этих функций. Описание функции содержится в соответствующем разделе.

- Создание
- Структура
- Ввод статей
- Текст-база
- Индексация
- Каталог
- Редактирование1
- Редактирование2
- Просмотр словарей
- Таблицы
- Инверсия
- Выборка
- Проекция
- Объединение
- Пересечение
- База-текст
- База-макет

Функция Создание

Функция **Создание** позволяет проектировать структуру статьи новой словарной базы и инициализировать файл базы данных нового словаря. Структура словарной статьи описывается списком элементарных и составных компонент, идущих в последовательности, принятой в данном словаре. Каждая элементарная компонента задается мнемоническим именем (его изобретает филолог) и расшифровкой этого имени. Список составных компонент также определяется филологом и, как правило, описывает те блоки статьи, в которые естественно группируются элементарные компоненты. В процессе проектирования структуры статьи нового словаря филолог может использовать типовые структуры словарей известных жанров (толкового, синонимов, орфографического и т. д.), а также структуры других словарей, находящихся, в виде базы данных в UNILEX-D.

После выхода из функции **Создание** в системе появляется файл структуры нового словаря с именем NNNNstr.dbf и пустой файл будущей базы данных с именем NNNNdbf.dbf.

Заполнение новой словарной базы может происходить либо с использованием функции **Ввод**, в которой информация каждой словарной статьи вводится с экрана, либо с использованием функции **Текст-база**, в которой информация в базовый файл переносится из текста.

Кроме функции **Создание** новый словарь можно спроектировать и создать в таких операциях преобразования словаря, как **Выборка**, **Проекция**, **Объединение**, **Пересечение**.

Функция Структура

Функция **Структура** предназначена для редактирования уже созданной структуры словарной базы данных. Филолог выбирает имя словаря из списка, предлагаемого в меню. На экране редактора появляется список элементарных и составных компонент статьи. В процессе редактирования возможно выполнить следующие операции: отредактировать имя компоненты и его расшифровку, удалить компоненту, вставить компоненту, задав ее имя и его расшифровку, взять имя компоненты из другого словаря системы, перенести описание структуры из базового файла в текстовой.

Редактирование структуры словарной статьи не означает одновременного внесения изменений в имена компонент уже существующей базы данных словаря. Поэтому для обеспечения правильной работы с базой данных, необходимо заменить в базе старые имена компонент на новые, используя команду замены в функции **Редактирование2**. Если в структуру были добавлены новые компоненты, то необходимо внести их имена с нулевым значением в базу данных, используя функцию **Редактирование1**.

Функция Ввод статей

Функция **Ввод статей** позволяет филологу вводить текст статьи с экрана компьютера, используя описание структуры словарной статьи. Первоначально на экране задается карта ввода, являющаяся списком имен элементарных компонент статьи и связанных с ними пустых полей, в которые можно вводить текст (значение) компоненты.

Если текст значения компоненты занимает более чем 60 символов, то значение переносится в следующее поле с таким же именем, которое образуется при нажатии клавиши F2. Если в статье присутствует несколько компонент с одним и тем же именем, то поле компоненты повторяется нажатием клавиши F2, для возможности ввода в них значений компонент.

При занесении значений компонент новой статьи заданного словаря у филолога имеется также возможность взять значение из уже имеющейся в системе словарной базы данных. Для этого вызывается в нижнем окне нужный словарь и в нем ищется необходимое значение. После заполнения карты ввода одной статьи и выхода из поля ввода необходимо нажать клавишу F2, для запоминания введенного материала в базе данных. После завершения сеанса ввода статей, необходимо выполнить функцию **Индексация** для дальнейшей работы со словарем.

Если в процессе ввода филолог допустил ошибку, то исправить текст статьи можно, используя функцию **Редактирование**.

Функция Текст-база

Функция **Текст-база** позволяет добавить к уже созданному в системе словарю новые статьи, перенося их в базу данных из текстового файла. Имя словарной базы, к которой добавляются статьи из текста, должно уже присутствовать в Каталоге баз данных системы. Если имя словарной базы данных NNNNdbf.dbf, то имя текстового файла должно быть NNNN.txt.

Текстовое представление словарных статей можно подготовить в любом компьютерном редакторе. При этом необходимо выполнить следующие условия. Текст каждой статьи должен быть представлен покомпонентно, с 7-ой по 66 позиции строки редактора. Позиция с 1-ой по 5-ую заполняется мнемоническим именем компоненты, объявленным при создании словаря в структуре словарной статьи. Если количество символов в значении компоненты превышает 60, то запись этого значения занимает несколько строк. При подготовке покомпонентного текстового файла словаря обязательным является заполнение имени заглавной компоненты с мнемоникой "ЗАГЛ". Остальные имена филолог может проставить в системе, используя опцию **Структура** и функцию **Редактирование2**. Если информация из текстового файла переносится в пустой, только что инициализированный в функции **Создание** словарный файл, то прежде выполнения функции **Текст-база** надо выполнить в псевдо-режиме (то есть, не заполняя карту, функцию **Ввод статей**).

Функцию **Текст-база** можно также использовать для восстановления базового файла из его текстовой копии, полученной функцией **База-текст**. При этом необходимо заново инициализировать словарную базу данных с именем NNNNdbf.dbf.

После завершения сеанса ввода статей, необходимо выполнить функцию **Индексация** для дальнейшей работы со словарем.

Функция Индексация

Функция **Индексация** позволяет построить необходимые индексы к созданной словарной базе данных, позволяющих осуществлять доступ к информации словаря от различных входов (входом в компьютерный словарь в системе UNILEX-D может быть не только заглавное слово, но и значение любой компоненты, объявленной в структуре).

Для выполнения функции **Индексация** необходимо выбрать нужный словарь из системного Каталога. Для каждого словаря система строит индексы трех типов: по номеру статьи, полный индекс по имени и значению компоненты, уникальный индекс по имени и значению компоненты.

Как правило, выполнение функции **Индексация** требуется после каждого изменения базы данных, то есть после выполнения функций: **Ввод статей**, **Редактирование**, **Текст-база**, а также при создании новых словарных баз при выполнении операций **Проекция**, **Выборка**, **Объединение**, **Пересечение**.

Функция Каталог

Функция **Каталог** позволяет филологу вести и хранить в системе список имен тех словарей, с которыми он работает в данный период. Редактор функции **Каталог** дает возможность занести новое имя словарной базы данных в список **Каталога**, удалить имя словарной базы из **Каталога**, отредактировать имя словарной базы данных и его расшифровку. Каждая строка списка содержит два поля — поле имени словарной базы, состоящее из 4-х латинских символов, и расшифровку этого имени (поле из 20 символов).

Функция Редактирование1

Функция **Редактирование1** позволяет филологу вывести текст словарной статьи из базы данных на экран компьютера и внести в него необходимые изменения. При этом текст словарной статьи отражается как копия ее базового представления, то есть статья расписана покомпонентно и слева от значения каждой элементарной компоненты находится поле с ее именем, объявленным в структуре при создании.

В процессе редактирования имеется возможность: удалить элементарную компоненту, вставить в нужное место элементарную компоненту, изменить значение компоненты или ее имени. При редактировании словарной статьи у филолога имеется также возможность взять значение компоненты из другого словаря, уже имеющегося в системе в виде базы данных. Для этого вызывается в нижнем окне нужный словарь и в нем ищется необходимое значение. Когда это значение найдено, оно переносится в редактируемый текст нажатием клавиши **ENTER**. После выхода из окна редактирования нужно нажать клавишу **F2** для запоминания результата в промежуточном буфере. Накопленные в буфере в процессе редактирования статьи можно просмотреть, нажав клавишу **F3**, экрана, дающего список заглавных слов данного словаря.

Результат редактирования (массив накопленных в буфере статей) можно использовать далее тремя способами: 1) заменить в базовом файле старые статьи на новые, нажав клавишу **F4** экрана списка заглавных слов; 2) создать новый словарь из статей, накопленных в буфере, задав его имя по запросу системы в процессе выхода из функции **Редактирование1**. Структура статей нового словаря совпадает со структурой старого; 3) добавить накопленные в буфере статьи к другому словарю, уже имеющемуся в системе (например, к предыдущей выборке из этого же словаря).

После завершения сеанса редактирования необходимо выполнить функцию **Индексация** для дальнейшей работы со словарем.

Функция Редактирование2

Функция **Редактирование2** позволяет филологу просматривать базу данных компьютерного словаря подряд статью за статьей и вносить изменения прямо в базовый файл (в отличие от функции **Редактирование1**). При этом каждая словарная статья отражается как копия ее базового представления, то есть статья расписана покомпонентно и слева от значения каждой элементарной компоненты находится поле с ее именем, объявленным в структуре при создании.

В процессе редактирования имеется возможность: удалить элементарную компоненту, изменить значение компоненты или ее имени, упорядочить номера статей в файле, если они идут не по возрастанию, проставить имена компонент в полуавтоматическом режиме, используя описание структуры статьи, проставить имена компонент в командном режиме, если есть какие-либо формальные признаки в значении компоненты, заменить имя какой-либо компоненты во всем словаре в командном режиме, заменить значение компоненты для всего словаря в командном режиме. При редактировании словарной статьи у филолога имеется также возмож-

ность взять значение компоненты из другого словаря, уже имеющегося в системе в виде базы данных. Для этого вызывается в нижнем окне нужный словарь и в нем ищется необходимое значение. Когда это значение найдено, оно переносится в редактируемый текст нажатием клавиши ENTER.

После завершения сеанса редактирования необходимо выполнить функцию Индексация для дальнейшей работы со словарем.

Функция Просмотр словарей

Функция **Просмотр словарей** позволяет филологу организовать вход в словарь от любой элементарной компоненты и вызывать тексты словарных статей на экран компьютера в порядке значений компоненты входа.

В начале сеанса **Просмотра** филолог выбирает имя искомого словаря из списка **Каталога**. После этого на экран компьютера вызывается описание структуры статьи данного словаря в виде списка элементарных и составных компонент. Для организации просмотра филолог должен составить **Запрос**. Строка **Запроса** состоит из имен элементарных и составных компонент статьи, выбираемых последовательно из списка. Первая компонента **Запроса** определяет **Вход** в словарь. Остальные компоненты **Запроса** проектируют структуру статей, показываемых на экране. Для каждого уникального значения **Входа** в словарь организуется временная **Выборка** из словаря, которую филолог может просмотреть или перенести в текстовый файл. Количество статей **Выборки**, уже просмотренных филологом, высвечивается на экране. Кроме того, филолог может вызвать в нижнем окне экрана для просмотра другой словарь, находящийся в системе, или какое-либо преобразование данного словаря (**Таблицу**, **Выборку**, **Проекцию**, **Инверсию**), если до сеанса они были созданы.

Функция Таблицы

Функция **Таблицы** позволяет филологу определить те количественные параметры словаря, которые могут быть выражены в виде таблиц распределения значений какой-либо элементарной компоненты словарной статьи, объявленной в описании ее структуры при создании словаря.

В начале сеанса филолог выбирает имя искомого словаря из списка **Каталога**. После этого на экран компьютера вызывается описание структуры статьи данного словаря в виде списка элементарных и составных компонент. Для вычисления таблицы филолог должен составить **Запрос**. Строка **Запроса** состоит из имени той элементарной компоненты, для которой считается распределение. После нажатия клавиши F2 появляется поле, в которое филолог должен внести имя таблицы, состоящее из 4-х латинских символов. После вычисления таблица заносится в файл TTTTdbf.dbf (TTTT — имя, заданное филологом), структура которого совпадает со структурой базовых словарных файлов. После вычисления таблица появляется на экране компьютера. Во время просмотра таблицы ее можно скопировать в текстовый файл, чтобы использовать в дальнейших исследованиях. Имя текстового файла копии или определяется филологом, или остается стандартным — FILE.txt.

Имя вычисленной таблицы можно занести в **Каталог**, что дает возможность просмотра и редактирования таблицы в функции **Редактирование2**, или просмотра в нижнем окне совместно с просмотром данного словаря.

Функция Инверсия

Функция **Инверсия** позволяет филологу составлять различные указатели к словарю как инверсионные списки значений одной компоненты относительно другой, при этом для второй компоненты указателя вычисляется число ее значений относительно значения первой компоненты.

В начале сеанса филолог выбирает имя искомого словаря из списка **Каталога**. После этого на экран компьютера вызывается описание структуры статьи данного словаря в виде списка элементарных и составных компонент. Для получения инверсии филолог должен составить **Запрос**. Строка **Запроса** состоит из имен первой и второй компоненты инверсии (элементарные компоненты). После нажатия клавиши F2 появляется поле, в которое филолог должен внести имя файла инверсии, состоящее из 4-х латинских символов. После создания указатель-инверсия заносится в файл IIIdbf.dbf (III — имя, заданное филологом), структура которого совпадает со структурой базовых словарных файлов.

После создания файл указателя-инверсии появляется на экране компьютера. Во время просмотра базовый файл-инверсию можно скопировать в текстовый файл, чтобы использовать в дальнейших исследованиях. Имя текстового файла копии или определяется филологом, или остается стандартным — FILE.txt.

Имя полученного файла инверсии можно занести в **Каталог**, что дает возможность просмотра и редактирования в функции **Редактирование2**, или просмотра в нижнем окне совместно с просмотром данного словаря.

Функция Выборка

Функция **Выборка** позволяет создать новую словарную базу данных из статей уже существующего в системе словаря. Вначале сеанса филолог выбирает имя искомого словаря из списка **Каталога**. После этого на экран компьютера вызывается описание структуры статьи данного словаря в виде списка элементарных и составных компонент. Для получения выборки филолог должен составить **Запрос**. Строка **Запроса** состоит из имен элементарных и составных компонент статьи, выбираемых последовательно из списка. Первая компонента **Запроса** определяет **Вход** в словарь. Остальные компоненты **Запроса** проектируют структуру статей, которые будут составлять новый словарь. После составления **Запроса** и нажатия клавиши F2 появляется поле, в которое филолог должен внести имя выборки, состоящее из 4-х латинских символов. Выборка заносится в новый словарь VVVVdbf.dbf (VVVV — имя, заданное филологом). Структура статей нового словаря, определенная **Запросом**, заносится в файл VVVVstr.dbf. После определения имени выборки на экране компьютера появляется список значений компонент **Входа** в словарь. Выбирая нужные ему значения из этого списка и нажимая клавишу ENTER, филолог может переносить статьи из старого словаря в новый. Если словарь с именем VVVV уже существовал в системе, то статьи выборки будут просто добавляться к нему без инициирования новой словарной базы. Это дает возможность формирования путем выборки нового словаря за несколько сеансов работы.

Имя нового словаря, полученного путем выборки, можно занести в **Каталог**, что дает возможность работы с ним в других функциях: **Просмотр словарей**, **Таблицы**, **Инверсия**, **Проекция**, **Объединение**, **Пересечение**, **База-текст**, **База-макет**.

После завершения сеанса необходимо выполнить функцию **Индексация** для дальнейшей работы с новым словарем.

Функция Проекция

Функция **Проекция** позволяет создать новую словарную базу данных, в которой заглавные слова статей совпадают со старым словарем, но состав статей определяется заданным списком компонент. Вначале сеанса филолог выбирает имя искомого словаря из списка **Каталога**. После этого на экран компьютера вызывается описание структуры статьи данного словаря в виде списка элементарных и составных компонент. Для получения проекции филолог должен составить **Запрос**. Строка **Запроса** состоит из имен элементарных и составных компонент статьи, выбираемых последовательно из списка. Первой компонентой **Запроса** обязательно должно быть заглавное слово (имя “ЗАГЛ”). Остальные компоненты **Запроса** проектируют структуру статей, которые будут составлять новый словарь. После составления **Запроса** и нажатия клавиши F2 появляется поле, в которое филолог должен внести имя проекции, состоящее из 4-х латинских символов. Проекция заносится в новый словарь PPPPdbf.dbf (PPPP — имя, заданное филологом). Структура статей нового словаря, определенная **Запросом**, заносится в файл PPPPstr.dbf.

Имя нового словаря, полученного путем проекции, можно занести в **Каталог**, что дает возможность работы с ним в других функциях: **Просмотр словарей**, **Таблицы**, **Инверсия**, **Выборка**, **Объединение**, **Пересечение**, **База-текст**, **База-макет**.

После завершения сеанса необходимо выполнить функцию **Индексация** для дальнейшей работы с новым словарем.

Функция Объединение

Функция **Объединение** позволяет создать новую словарную базу данных, в которую будут входить все статьи из 2-х уже существующих баз, но состав новых статей будет определяться списком компонент, выбранных как из структуры первого, так и второго словаря. В результате работы функции **Объединение** в новый словарь будут входить статьи, содержащиеся либо только в первом, либо только во втором словаре. Из статей, имеющих одно и то же заглавное слово как в первом, так и во втором словаре, формируется новая статья, согласно списку компонент новой статьи. Выбор в качестве 1-го и 2-го одного и того же словаря позволяет сформировать новый словарь, отличающийся от старого другой последовательностью компонент. В начале сеанса **Объединение** филолог выбирает имена искомого словарей из списка **Каталога**. После этого на экран компьютера вызываются описания структур статей 1-го и 2-го словарей в виде списков их элементарных и составных компонент.

Для получения объединения филолог должен составить **Запрос**. Строка **Запроса** состоит из имен элементарных и составных компонент статей сначала 1-го, затем 2-го словаря. Первая

компонента списка для каждого словаря обязательно должна быть заглавным словом (имя “ЗАГЛ”), так как именно наличие одинаковых заглавных слов является критерием объединения 2-х статей в одну. Остальные компоненты Запроса проектируют структуру статей, которые будут составлять новый словарь. После составления Запроса и нажатия клавиши F2 появляется поле, в которое филолог должен внести имя словаря-объединения, состоящее из 4-х латинских символов. Объединение заносится в новый словарь JJJdbf.dbf (JJJ — имя, заданное филологом). Структура статей нового словаря, определенная Запросом, заносится в файл JJJstr.dbf. Имя нового словаря, полученного объединением, можно занести в Каталог, что дает возможность работы с ним в других функциях: Просмотр словарей, Таблицы, Инверсия, Выборка, Проекция, Пересечение, База-текст, База-макет.

После завершения сеанса необходимо выполнить функцию Индексация для дальнейшей работы с новым словарем.

Функция Пересечение

Функция Пересечение позволяет создать новую словарную базу данных, в которую будут входить слитые вместе статьи из 1-го и 2-го словарей, имеющие одно и то же заглавное слово. Состав новых статей будет определяться списком компонент, выбранных из структуры как 1-го, так и 2-го словарей. Выбор в качестве 1-го и 2-го одного и того же словаря позволяет сформировать новый словарь, отличающийся от старого другой последовательностью компонент.

В начале сеанса филолог выбирает имена искомых словарей из списка Каталога. После этого на экран компьютера вызываются описания структур статей 1-го и 2-го словарей в виде списков их элементарных и составных компонент. Для получения пересечения филолог должен составить Запрос. Строка Запроса состоит из имен элементарных и составных компонент статей сначала 1-го, затем 2-го словаря. Первая компонента списка для каждого словаря обязательно должна быть заглавным словом (имя “ЗАГЛ”), так как именно наличие одинаковых заглавных слов является критерием объединения 2-х статей в одну. Остальные компоненты Запроса проектируют структуру статей, которые будут составлять новый словарь. После составления Запроса и нажатия клавиши F2 появляется поле, в которое филолог должен внести имя словаря-пересечения, состоящее из 4-х латинских символов. Файл словаря-пересечения заносится в новый словарь CCCCdbf.dbf (CCCC — имя, заданное филологом). Структура статей нового словаря, определенная Запросом, заносится в файл CCCCstr.dbf.

Имя нового словаря, полученного пересечением, можно занести в Каталог, что дает возможность работы с ним в других функциях: Просмотр словарей, Таблицы, Инверсия, Выборка, Проекция, Объединение, База-текст, База-макет.

После завершения сеанса необходимо выполнить функцию Индексация для дальнейшей работы с новым словарем.

Функция База-текст

Функция База-текст позволяет создать текстовую копию словарной базы данных, существующей в системе. В начале сеанса филолог выбирает имя словаря из списка Каталога и нажимает клавишу ENTER. В процессе переноса статей в текстовый файл на экране компьютера высвечивается счетчик перенесенных записей.

Текстовая копия словарной базы данных представляет собой файл, в котором каждая статья словаря расписана покомпонентно, и слева от значения каждой элементарной компоненты указывается ее имя, объявленное в описании структуры словаря (имя располагается с 1-ой по 5-ую позиции строки текстового файла, значение — с 7-ой по 66). Если имя базы данных словаря было NNNNdbf.dbf, то имя его текстовой копии, созданной функцией База-текст будет NNNN.txt. Из текстовой копии можно заново создать словарную базу, используя функцию Текст-база.

Функция База-макет

Функция База-макет позволяет создать размеченную текстовую копию (макет-файл) базы данных компьютерного словаря, в которой слева от значения каждой элементарной компоненты статьи указан шифр шрифта.

В начале сеанса филолог выбирает имя словаря из списка Каталога. После выбора словаря на экране компьютера высвечивается описание структуры статьи данного словаря. В нижнем окне экрана при нажатии клавиши F5 высвечивается таблица типов шрифтов (прямой, полужирный, курсив; высота шрифта: 8, 9, 10, 12). Выбирая последовательно для каждой компоненты словарной статьи тип шрифта, филолог получает таблицу соответствия компоненты и шрифта.

После выхода из режима создания таблицы соответствия филолог может либо выйти из функции, либо продолжить работу. Во втором случае система формирует текстовый покомпонентный файл словаря, в котором вместо имени компоненты записан код ее шрифта.

Если имя словарной базы данных было NNNNdbf.dbf, то имя соответствующего текстового макет-файла будет NNNN.tex.

В последующих сеансах работы филолог может изменить типы шрифтов у компонент данного словаря.

Конвертирование макет-файла в оригинал-макет словаря происходит посредством специальной макро-программы на языке WORDBASIC под управлением операционной системы WINDOWS'95.

Применение в словарных проектах

Универсальная лексикографическая система UNILEX-D применяется для создания баз данных различных филологических словарей. Работа по созданию словарных баз данных проводится в отделе Машинного фонда русского языка (для двух словарей — совместно с Орфографическим отделом и отделом Стилистики языка и литературы).

К настоящему времени создано 7 словарных баз данных, которые хранятся в Архиве лингвистических источников Машинного фонда русского языка [Колодяжная, 1996].

При вводе информации в базы данных применялись различные технологии: ручной ввод в формат базы данных, конвертирование из текстовых файлов, конвертирование из текстовых баз данных. Словарные базы данных использовались (и используются) как в целях научного исследования (например, статистика слов русского языка по грамматическим категориям), так и для создания оригинал-макета словарей в издательских целях.

Описание словарных баз, созданных в отделе Машинного фонда русского языка, входит в каталог российских коллективов и разработок “Автоматизированная обработка речи и текстов на естественном языке”, созданный В. Э. Семеновой [Semenova, Fluhr, 1997]. Информация о каталоге представлена также в сети Интернет, на сайте <http://members.aol.com/fluhrsciper/catalogue>.

Словарь русского языка С.И. Ожегова

Общая характеристика

База данных создана на основе книги: **С. И. Ожегов. Словарь русского языка.** Под ред. Н. Ю. Шведовой. 22-е изд. М.: Рус. яз., 1990.

Компьютерная версия словаря С. И. Ожегова реализована как система словарных баз данных, каждая из которых содержит массив статей на одну букву (А–Я), работающих под управлением лексикографической программной системы UNILEX-D. База данных содержит около 35000 статей. В рабочем состоянии система баз данных для словаря С. И. Ожегова занимает 56 МБ. Компьютерная версия словаря описана в работе [Колодяжная, Пшеничникова, 1992].

Структура словарной статьи

Словарная статья состоит из заглавного слова, фонетической, грамматической, стилистической помет к нему, варианта заглавного слова, помет к варианту, дефиниции, примеров, фразеологических сочетаний с заглавным словом, их толкования и примеров с ними, производных слов, помет к производным и примеров с производными. Ниже приведено формальное описание структуры статьи словаря С. И. Ожегова в системе UNILEX-D

- 1 ЗАГЛ заглавное слово
- 2 ФПОМ фонетическая помета
- 3 ГПЗ грамматическая помета
- 4 ВАРЗ вариант заглавного слова
- 5 ГПВ грамматическая помета вариант
- 6 СТЗ стилистическая помета
- 7 ДЕФ дефиниция
- 8 ПРИМ пример
- 9 ФРАЗ фразеология
- 10 ТФРА толкование фразеологии
- 11 ПРИФ пример употребления фразеологии
- 12 ЧРП часть речи производного
- 13 ПРОИ производное
- 14 ГПП грамматическая помета производного

15	ПОМП	помета к производному	
16	ПРИП	пример к производному	
17	*СТАТ	вся статья	1 16
18	*ЗЗАГ	зона заглавного слова	1 11
19	*ЗПРО	зона производного	12 16
20	*ЗПОМ	заглавное с пометами	1 6
21	*ДЕПР	дефиниция с примерами	7 8
22	*ЗФРА	зона фразеологии	9 11

Знак * стоит перед именем составной компоненты; цифры обозначают номера элементарных компонент, ограничивающих составную. Элементарные компоненты могут быть обязательными, факультативными и рекурсивными (повторяющимися). Обязательной компонентой является в Словаре заглавное слово. Остальные компоненты — факультативные. Повторяющиеся могут быть зоны дефиниций, примеров, производных.

Особенности функционирования словаря в UNILEX-D

Создание базы данных

Книжный словарь был введен с клавиатуры на магнитные носители ЭВМ и затем конвертирован в текстовый файл с помощью специально написанной программы. При вводе словаря соблюдались некоторые формальные правила (расположение каждой компоненты с первой позиции новой записи, дополнительная разметка и т. п.)

Примеры статей в исходном текстовом файле:

МОДЕР'НЫЙ,

-ая, -ое.

\$МОДЕР'НОВЫЙ,

[дэ'],

-ая, -ое.

(прост.).

То же, что модерн (в 3 знач.).

*Модерные танцы. #

МОДИ'СТКА,

-и, ж.

(устар.).

Мастерица, изготавливающая дамские шляпы, а также портниха. #

МОДИФИЦИ'РОВАТЬ,

-рую, -руешь; -анный; сов. и несов., что.

(книжн.).

То же, что видоизменить (-нять).

// возвр.

модифици'роваться,

-руюсь, -руешься.

// сущ.

модифика'ция,

-и, ж. #

Примеры представления статей в базе данных:

ЗАГЛ А'ВГИЕВ,

ГПЗ -ы.

ФРАЗ @авгиевы конюшни (книжн.)

ТФРА — о крайне запущенном помещении, а также (перен.) о делах,

ТФРА находящихся в крайнем беспорядке [по древнегреческому мифу

ТФРА о конюшнях царя Авгия, к-рые не очищались 30 лет].

ЗАГЛ АГРЕ'ССОР,

ГПЗ -а, м.

ДЕФ Тот, кто производит агрессию.

ЧРП //прил.

ПРОИ агре'ссорский,

ГПП -ая, -ое.

ЗАГЛ А'ДСКИЙ,

ГПЗ -ая, -ое.

ДЕФ 1. см. ад.

ДЕФ 2. перен. Очень сильный, а также невыносимый (о чем-н.

ДЕФ неприятном) (разг.).
ПРИМ *А. холод. Адски (нареч.) устать.
ФРАЗ @Адская машина (устар.)
ТФРА — снаряд с часовым механизмом, взрывающийся в определенный
ТФРА момент.

ЗАГЛ БА'РХАТКА,
ГПЗ -и.
ВАРЗ \$БАРХО'ТКА,
ГПВ -и,ж.
ДЕФ Ленточка из бархата, кусочек бархата.

ЗАГЛ БА'РЧУ'К,
ГПЗ -а',м.
СТЗ (устар.).
ДЕФ Мальчик из барской семьи.

ЗАГЛ БОГАТЕ'ТЬ,
ГПЗ -е'ю,-е'ешь,несов.
ДЕФ Становиться богатым (в 1 знач., о земле, крае в 3 знач.),
ДЕФ богаче.
ЧРП //сов.
ПРОИ разбогате'ть,
ГПП -е'ю,-е'ешь.
ПОМП (по 1 знач. прил. богатый).

В случае грамматической омонимии книжная статья представлялась в базе данных как несколько статей. Например, книжная статья с заглавным словом АГА имеет вид:

АГА' [aha']. 1. межд. Восклицание с торжествующей интонацией. Ага! попался! 2. частица. То же, что да 1 (в 1 знач.) (прост.). Видишь? — Ага, вижу.

В машинном варианте эта статья разбивается на две:

ЗАГЛ АГА'
ФПОМ [aha'],
ГПЗ межд.
ДЕФ Восклицание с торжествующей интонацией.
ПРИМ *Ага! попался!

ЗАГЛ АГА'
ФПОМ [aha'],
ГПЗ частица.
ДЕФ То же, что да 1 (в 1 знач.) (прост.).
ПРИМ *Видишь? — Ага, вижу.

Работа со словарем С. И. Ожегова в UNILEX-D

При работе со словарной базой данных в UNILEX-D филологу-исследователю открываются новые, по сравнению с книжным словарем, возможности. Например, функция системы UNILEX-D Просмотр словарей позволяет организовывать Вход в Словарь от любой элементарной компоненты. При этом текст выходных статей формируется в соответствии с набором компонент, выбранных для просмотра, что позволяет решать целый ряд задач.

Например:

1. Вход в Словарь по компоненте “заглавное слово”.

Запрос вида: ЗАГЛ *СТАТ — обеспечивает быстрый выход на нужную статью и просмотр всей статьи.

Запрос вида: ЗАГЛ ГПЗ — позволяет просмотреть все заглавные слова и грамматические пометы к ним. При этом автоматически вычисляется распределение грамматических помет.

Словарь в данном случае выступает в роли грамматического справочника.

Запрос вида: ЗАГЛ ДЕФ — позволяет просмотреть только заглавные слова и их толкования (т. е. осуществлять работу со Словарем как с собственно толковым).

Запрос вида: ЗАГЛ ПРОИ — позволяет просмотреть все заглавные слова и их производные.

2. Вход в Словарь по компоненте, отличной от заглавного слова и просмотр произвольных компонент статьи.

Запрос вида: ФРАЗ ЗАГЛ — позволяет просмотреть только статьи, содержащие фразеологические сочетания с заглавным словом.

Запрос вида: ПРОИ ЗАГЛ — дает возможность просмотреть все производные слова и их производящие.

Запрос вида: ГПЗ ЧРП — позволяет получить распределение частей речи производных слов относительно частей речи производящих.

3. Вход в Словарь по компоненте, отличной от “заглавного слова” и просмотр всей статьи.

Запрос вида: ВАРЗ *СТАТ — позволяет найти и просмотреть все статьи, в которых у заглавного слова есть вариант.

Запрос вида: СТИЛ *СТАТ — позволяет найти и просмотреть все статьи, содержащие стилистические пометы к заглавному слову.

Научные исследования и публикации

Филологический словарь, существующий в структурированном формате базы данных, содержащий достаточно большое количество информации, является объектом научного исследования языка, его закономерностей.

По базе данных толкового словаря С. И. Ожегова были проведены различные исследования, результаты которых опубликованы в работах:

1) Тезисы доклада “Метасловарь количественных параметров Словаря русского языка С. И. Ожегова”. [Колодяжная, Пшеничнова, 1992]

В докладе дается описание параметров толкового словаря русского языка С. И. Ожегова. В качестве примеров приводятся результаты подсчетов, выполненные для массивов букв А и Б.

Книжные словари обычно характеризуются единственным параметром — числом словарных статей. Между тем количественному подсчету поддаются и совокупности различных признаков заглавного слова, зафиксированные в правой части словарной статьи. Очевидно, что эти количественные характеристики дают представление о структуре того лексического пространства, которое описано в данном словаре. В статье описывается три типа количественных параметров: элементарные, табличные и корреляционные.

В Словаре С. И. Ожегова были выделены следующие элементарные параметры:

Число статей	(А=842; Б=1617)
Число статей, содержащих фразеологию	(А= 40, Б= 86)
Число статей с дефиницией, выраженной фразеологическим сочетанием	(А= 10, Б= 25)
Число производных слов, имеющих самостоятельную статью	(А = 15, Б= 44)
Число заглавных слов, имеющих варианты	(А= 8, Б= 32)
Число заглавных слов, имеющих стилистические пометы	(А=180, Б= 378)

Параметры табличного типа можно разделить на два вида. К первому виду относятся количественные распределения каких-либо признаков заглавного слова по значениям (например, распределение слов по частям речи, распределение слов по числу значений и т. п.). Ко второму виду относятся количественные распределения заглавных слов по числу значений в одной статье какого-либо признака (например, распределение слов по числу синонимов, антонимов, производных и т. п.).

Распределение слов по грамматическим категориям:

	“А”	“Б”
сущ.	626	939
прил.	93	395
глагол.	65	185
межд.	9	8
нареч.	7	8
частица	5	7
союз	4	3
предлог		4
неизм.	3	
приставка	3	3
в знач. сказ.	2	6
нареч. и в знач. сказ.		2
союз и частица	2	
вводн. сл. и частица	1	1

К параметрам корреляционного типа относятся таблицы, в которых даны количественные распределения одного признака заглавного слова относительно другого признака (например, распределение грамматических классов слов-производных относительно грамматических классов производящих слов).

Корреляционная таблица классов производных и классов заглавных слов:

класс произв. класс и колич. слов,
дающих данное произв.

А

//возвр. глаг.(1)
//несов. глаг.(1)
//однокр. глаг.(3)
//уменьш. сущ. (5)
//сов. глаг.(7)
//сущ. глаг.(30) прил.(52)
//прил. нареч.(1) глаг.(7) сущ.(420)

Б

//многокр. глаг. (1)
//уменьш.-ласк. сущ. (1)
//уменьш.-унич. сущ. (1)
//уменьш.-шутл. сущ. (1)
//унич.и ласк. сущ. (1)
// межд собир. сущ. (2)
//унич. сущ. (2)
//возвр. глаг. (5)
//несов. глаг. (7)
//ласк. сущ. (8)
//однокр. глаг. (11)
//уменьш. нареч.(1) сущ.(55)
//сов. глаг. (80)
//ж. сущ. (94)
//сущ. глаг. (70) прил.(306)
//прил. глаг. (27) сущ. (471)

2) Тезисы доклада на 2-ой международной конференции по квантитативной лингвистике Qualico-94 (МГУ, Москва, 1994 г.): "Исследование количественных корреляций между стилистическими, грамматическими и полисемическими характеристиками слов (на основе словаря С. И. Ожегова)" [Kolodjaznaja, Polikarpov, 1994].

В докладе представлены результаты исследования, проведенного авторами на материале толкового словаря русского языка с целью исследования количественных зависимостей стилистики и полисемии.

В качестве материала для проведения исследования была взята база данных толкового словаря русского языка С. И. Ожегова (22-е издание), созданная в Машинном фонде русского языка. Инструментом исследования послужила программная лексикографическая система UNILEX-D, в среде которой функционирует данный компьютерный словарь.

На материале исследуемого толкового словаря была получена количественная таблица распределения слов по грамматическим категориям:

катег.	noun	verb	adj.	adv.	pron.	int.	par.	conj.	pred.	prep.	part.	всего
число	18829	5894	8407	806	132	83	78	93	175	112	163	34772

Кроме того, с использованием функций UNILEX-D были созданы 11 подсловарей, соответственно различным грамматическим категориям слов. Для каждого подсловаря была вычислена корреляционная таблица, показывающая количественную зависимость обобщенной стилистической пометы от числа значений (показателя полисемии). Ниже приведена одна из таких таблиц (для существительных).

Таблица распределения стилистических типов существительных относительно числа значений слова

стил. полисем.	Р-разг.	К-книжн.	У-устар.	О-обл.
1	1461	1423	578	61

2	145	169	33	9
3	14	31	4	3
4	2	4	1	
6	1			

Словарь синонимов Е. П. Евгеньевой

Общие характеристики

База данных создана на основе книги: *Словарь синонимов. Справочное пособие*, под ред. А. П. Евгеньевой. М.: Наука. 1975.

База данных Словаря синонимов была создана совместно с Лабораторией компьютерной лексикографии филологического факультета МГУ (зав. лаб.— А. А. Поликарпов).

Компьютерная версия словаря синонимов реализована как система 3-х лексикографических баз данных, соответствующих 3-м типам корпусов словарных статей:

ОСНОВНЫЕ СТАТЬИ, содержащие заглавное слово (его вариант), синонимы (варианты), антонимы;

СТАТЬИ ГЛАГОЛОВ, содержащие заглавное слово (из 1-го корпуса), помету о виде глагола (сов./несов.), глагол, по виду противоположный заглавному слову, его синонимы;

СТАТЬИ ПРОИЗВОДНЫХ СЛОВ, содержащие заглавное слово (из 1-го корпуса), производное от заглавного слова, синонимы к производному.

Общее число словарных статей в системах баз данных составляет 7902, из них:

- 5464 — групп основного корпуса,
- 1696 — групп глаголов противоположного вида,
- 742 — группы производных слов.

В рабочем состоянии (вместе с индексами) система баз данных для словаря синонимов занимает 2 МБ. Компьютерная версия словаря описана в работах [Колодяжная, 1992] (“Система компьютерный филологический словарь”) и [Колодяжная, 1995].

Структура словарной статьи в базе данных

Книжный словарь [Словарь синонимов, 1975], послуживший исходным материалом для создания компьютерной версии, организован следующим образом. Он включает корпус основных статей (5464), корпус статей глаголов, соотносительных по виду (1696), корпус статей, включающих синонимические группы слов, производных к словам основных групп (220 статей). Кроме того, в словаре более 20000 статей, являющихся отсылочными к основным.

Статья из основной группы возглавляется словом, являющимся доминантным (по значению) среди всех членов данной синонимической группы. Книжная статья основной группы может включать, кроме синонимического ряда, группу антонимов к доминантному слову, стилистические пометы, толкования слов, иллюстрации употребления в виде речений; кроме того, статья может включать варианты слов синонимического ряда.

Компьютерная версия синонимического словаря функционирует как система трех баз данных: корпуса основных статей; корпуса статей соотносительных по виду глаголов; корпуса статей синонимов слов производных. В книжном словаре статьи соотносительных по виду статьи производных слов являются подстатьями соответствующих групп основных статей. В соответствии с этим статьи глагольного корпуса и корпуса производных связаны со статьями основного корпуса общим заглавным словом.

В основные статьи компьютерной версии включена вся информация книжного словаря, кроме толкований, помет и иллюстраций.

Ниже приведены описания структур статей каждой из трех словарных баз и примеры самих статей.

Структура статьи основной базы:

- 1 ЗАГЛ заглавное слово
- 2 АТЗ антоним к заглавному
- 3 ВАРЗ вариант заглавного
- 4 СИН синоним
- 5 ВАРС вариант синонима
- 6 *СТАТ вся статья 1 5
- 7 *ЗЗАГ зона заглавн.с ант. 1 3
- 8 *ЗСИН зона синонима 4 5

Примеры статей основной базы:

ЗАГЛ АКТИВНЫЙ
АТЗ пассивный
АТЗ бездейственный
АТЗ бездеятельный
АТЗ инертный
СИН ДЕЯТЕЛЬНЫЙ
СИН ЭНЕРГИЧНЫЙ
СИН ИНИЦИАТИВНЫЙ

ЗАГЛ АД
АТЗ рай
АТЗ эдем
СИН ПРЕИСПОДНЯ
СИН ГЕЕНА
СИН ПЕКЛО
СИН ТАРТАР
СИН ТАРТАРАРЫ

ЗАГЛ АККУРАТНЫЙ
АТЗ неопрятный
АТЗ неаккуратный
АТЗ неряшливый
СИН ЧИСТОПЛОТНЫЙ
СИН ОПРЯТНЫЙ

Структура статей базы глаголов соотносительно вида:

1 ЗАГЛ заглавное слово
2 ПОМГ помета (сов./несов.)
3 СИНС синоним совершенного
4 ВSOB вариант совершенного
5 СИНН синоним несовершенного
6 ВNEC вариант несовершенного
7 *СТАТ вся статья 1 6

Примеры статей глагольной базы:

АДРЕСОВАТЬ
несов.
АДРЕСОВАТЬ
НАПРАВЛЯТЬ
ОБРАЩАТЬ

АМПУТИРОВАТЬ
несов.
АМПУТИРОВАТЬ
ОТНИМАТЬ
ОТРЕЗАТЬ

АРЕСТОВАТЬ
несов.
АРЕСТОВАТЬ
АРЕСТОВЫВАТЬ
БРАТЬ
ЗАБИРАТЬ
ЗААРЕСТОВЫВАТЬ
ЗАДЕРЖИВАТЬ
ХВАТАТЬ
СХВАТЫВАТЬ
ЗАЦАПЫВАТЬ
ЗАМЕТАТЬ

Структура статьи базы слов-производных:

1 ЗАГЛ заглавное слово
2 ПРСЛ производное заглавного
3 СИНП синоним производного
4 *СТАТ вся статья 1 3

Примеры статей базы слов-производных:

АККУРАТНЫЙ
АККУРАТНО
ЧИСТОПЛОТНО
ОПРЯТНО
АККУРАТНОСТЬ
ЧИСТОПЛОТНОСТЬ
ОПРЯТНОСТЬ

АКТЕР
АКТРИСА
АКТЕРКА
АРТИСТКА
ЛИЦЕДЕЙКА

АКТИВНЫЙ
АКТИВНО
ДЕЯТЕЛЬНО
ЭНЕРГИЧНО

Знак * в описании структуры стоит перед именем составной компоненты; цифры обозначают номера элементарных компонент, ограничивающих составную.

Особенности функционирования словаря в UNILEX-D

При вводе информации в базу данных использовались функции системы UNILEX-D: Структура статьи, Ввод, Редактирование.

При работе со словарем синонимов система обеспечивает прямой доступ к словарной статье по любому возможному входу:

- от заглавного слова,
- от варианта заглавного слова,
- от синонима,
- от варианта синонима,
- от антонима,
- от глагола противоположного вида,
- от производного слова.

Научные исследования и публикации

На материале компьютерной версии словаря синонимов была проведена исследовательская работа, результаты которой описаны в статье “Системно-квантитативное исследование русской синонимии” [Колодяжная, Поликарпов, 1997]. Ниже приведены основные положения статьи.

Системно-квантитативное исследование русской синонимии

В статье описаны элементарные и табличные количественные параметры синонимического словаря.

Элементарные количественные параметры являются характеристиками, отражающими в числовом виде самые общие свойства данной лексической системы, представленной в словаре. Ниже приведены элементарные количественные параметры Словаря синонимов.

	Параметр	Значение
1.	Число статей в основной базе (число доминант)	5464
2.	Число разных лексем, вступающих в синонимические отношения	16852
3.	Число всех лексико-семантических вариантов, вступающих в синонимические отношения	20580
4.	Число групп синонимов производных слов	220
5.	Число всех слов-производных, образующих синонимические группы	742
6.	Число статей глаголов, соотносительных по виду	1696
7.	Число всех слов в рядах глагольной базы	7440
8.	Число вариантов к заглавным словам	91
9.	Число вариантов к членам синонимических групп	751

Параметры табличного типа выражают количественно структуру этой совокупности, выявляют ее экстремальные точки, например такие, как максимальный объем синонимической группы, максимальное число “вхождений” слова в различные синонимические группы и т. п. В том

числе, представление этих данных в виде распределения дает возможность понять, как формируется то или иное среднее значение данной характеристики. Ниже приведены наиболее важные табличные параметры, характеризующие лексический массив данного Словаря синонимов.

Распределение основных групп по объему синонимического ряда

объем	2	3	4	5	6	7	8	9	10	11	12	13
группы												
число	1578	1618	959	563	296	152	101	69	46	22	11	16
групп												

объем	14	15	16	17	18	19	20	21	23	24	25	26	34
группы													
число	4	8	4	2	4	3	1	1	1	1	1	1	2
групп													

Словарь “Слитно или раздельно?”

Общие характеристики

База данных создана на основе книги: **Б. З. Букчина, Л. П. Калакуцкая** *Слитно или раздельно? (Опыт словаря-справочника)*, изд. 6-е, М.: Русск. яз., 1987.

Компьютерная версия словаря реализована как система словарных баз, каждая из которых содержит массив статей на одну букву (А–Я), работающих под управлением лексикографической программной системы UNILEX-D. База данных содержит около 82000 словарных статей.

Примеры словарных статей в базе данных

ЗАГЛ акваметр
ЧАСТ1 аква/етр

ЗАГЛ аккордноплачиваемый
ЧАСТ1 акк/ордно
ЧАСТ2 опл/ачиваемый

ЗАГЛ аккомпаниаторгитарист
ЧАСТ1 аккомпани/атор
СВЯЗК –
ЧАСТ1 гитар/ист
ПАРАД аккомпани/атора-гитар/иста

ЗАГЛ аллегромодерато
ЧАСТ1 алл/егро
ЧАСТ2 модер/ато
ГРАМ неизм.

Словарь управления в русском языке Д. Э. Розенталя

Общие характеристики

База данных словаря создана на материале книги: **Д. Э. Розенталь** *Управление в русском языке: Словарь справочник*. 2-е изд. М.: Книга, 1986. 304 с.

Число словарных статей в базе данных — 2484. В рабочем состоянии база данных словаря занимает 3 МБ. Компьютерная версия Словаря управления описана в статье [Колодяжная, 1992]

Структура словарной статьи

Словарная статья Словаря управления состоит из заглавного слова, зоны управления (вопросительное местоименное слово или падежная форма заглавного слова), примеров употребления. Для многозначных слов приводятся типы управления с примерами в каждом значении. Ниже приведен список элементарных и составных компонент. Для каждой компоненты дается ее условное имя и его расшифровка. Элементарные компоненты можно разделить на три группы: обязательные, факультативные и рекурсивные (повторяющиеся). В частности, для Словаря управления обязательными являются “заглавное слово” и “пример”, остальные компоненты факультативны. К рекурсивным относятся компоненты: “управление”, “пример”. Перед име-

нем составной компоненты в списке ставится символ *. Для каждой составной компоненты указываются номера ограничивающих ее элементарных компонент.

ЗАГЛ	заглавное слово
ВАРЗ	вариант заглавного
ЧРЕЧИ	часть речи
ОТС	отсылка
УПР0	управление (основн.)
ПОМ0	помета к УПР0
ПРИ0	пример к УПР0
УПР1	управление (1 знач.)
ПОМ1	помета к УПР1
ПРИ1	пример у УПР1
УПР2	управление (2 знач.)
ПОМ2	помета к УПР2
ПРИ2	пример к УПР2
УПР3	управление (3 знач.)
ПОМ3	помета к УПР3
ПРИ3	пример к УПР3
УПР4	управление (4 знач.)
ПОМ4	помета к УПР4
ПРИ4	пример к УПР4
УПР5	управление (5 знач.)
ПОМ5	помета к УПР5
ПРИ5	пример к УПР5
ПОЯС	пояснение
ПРПО	пример к пояснению
*СТАТ	вся статья 1 24
*ЗОСН	зона основная 1 7
*ЗОН1	зона 1 знач. 8 10
*ЗОН2	зона 2 знач. 11 13
*ЗОН3	зона 3 знач. 14 16
*ЗОН4	зона 4 знач. 18 20
*ЗОН5	зона 5 знач. 21 22
*ЗОНП	зона пояснения 23 24

Словарная статья Словаря управления состоит из заглавного слова, зоны управления (вопросительное местоименное слово или падежная форма заглавного слова), примеров употребления. Для многозначных слов приводятся типы управления с примерами в каждом значении.

Особенности функционирования словаря в UNILEX-D

Ввод информации в базу данных

Тексты книжных статей были введены в ПЭВМ в виде исходного текстового файла, размеченного по некоторым правилам. А именно:

1. каждая компонента словарной статьи располагалась с новой записи, заголовочное слово — с первой позиции, остальные компоненты — с отступом,
2. совокупность местоименных слов, обозначающих тип управления, выделялась символом @,
3. в статье многозначного слова перед каждым новым значением ставился символ \$
4. символ \$ ставился также перед пометой или толкованием,
5. символ * ставится перед началом примера.

Примеры статей в исходном текстовом файле:

обидный

@кому и для кого.@

*Обидное нам (для нас) обвинение. Обвинение в трусости – самое обидное для летчиков (Саянов).

обиженный (обижен)

@кем-чем и на кого-что.@

*Обижен вами. Обижен оказанным ему холодным приемом. Оробев, подходит ближе, чем-то словно виноват, чем-то будто бы обижен (Твардовский). Он обижен на приятеля.

обиловать

\$(устар.) @чем.@
*[Мы] спустились в долину р. Таймендры... ее русло обило-
вало крупными валунами (Обручев).

Для конвертирования информации статей из исходного файла в базовый была создана специальная программа.

Примеры представления статей в базе:

ЗАГЛ абонемент
УПРО @на что.@
ПРИО *Абонемент на пользование телефоном. Абонемент на цикл
ПРИО лекций.

ЗАГЛ агрессия
УПРО @куда.@
УПРО @где.@
УПР1 \$ 1. @куда.@
ПОМ1 (направление действия).
ПРИ1 * Вот и новая агрессия в соседнюю африканскую страну.
УПР2 \$ 2. @где.@
ПОМ2 (место действия).
ПРИ2 *Последствия агрессии в соседней стране.

ЗАГЛ адресат.
УПРО @адресата.@
ПОМ0 (Вин.пад.)
ПРИО *Найти адресата.

ЗАГЛ авария
УПРО @чего.@
УПРО @с чем.@
ПРИО * Авария машины. Авария с самолетом.

ЗАГЛ без — безо.
ПОЯС \$Добавление гласного _о_ к предлогу _без_ наблюдается в
ПОЯС сочетаниях с местоимениями %весь, всякий%, например: безо
ПОЯС всего,безо всяких колебаний
ПОЯС (наряду с "без всяких колебаний").

ЗАГЛ боязнь:
УПРО @из боязни.@
УПРО @из-за боязни.@
УПР1 \$ 1. @из боязни.@
ПОМ1 (при указании на сознательный характер действия).
ПРИ1 * Мы уже посадили последние хлеба в печь и, из боязни
ПРИ1 передержать их, не ложились спать (Горький).
УПР2 \$ 2. @из-за боязни.@
ПОМ2 (при указании только на причину).
ПРИ2 * Из-за боязни увидеть дурной сон ему не спалось (Федин).

Работа со словарем в системе UNILEX-D

При работе со Словарем управления система обеспечивает прямой доступ к словарной статье по каждому из входов:

- от заглавного слова,
- от варианта заглавного слова,
- от основного управления,
- от управления 1-ом, 2-ом и т.д. значений,
- от пометы,
- от примера,
- от пояснения.

Работа со Словарем управления с использованием функции **Просмотр словарей** позволяет решать следующие задачи.

1. Запрос: ЗАГЛ *СТАТ — обеспечивает быстрый выход на нужную статью по заглавному слову и просмотр всей статьи.
2. Запрос: ЗАГЛ ПРИ0 ПРИ1 ПРИ2 ПРИ3 ПРИ4 ПРИ5 — обеспечивает быстрый поиск по заглавному слову и просмотр примеров употребления этого слова в различных ситуациях.

3. Запрос: ЗАГЛ УПР0 — позволяет быстро просмотреть типы местоименных слов (что, за кем, кого и т. п.) для данного заглавного. При этом автоматически осуществляется генерация указателя, данного в конце книги-источника.
4. Запрос: УПР0 ЗАГЛ ПРИ0 — позволяет просмотреть все статьи, связанные с данным типом управления, выражаемым одним и тем же местоименным словом.
5. Запрос: ОТС *СТАТ — позволяет найти все отсылочные статьи.
6. Запрос: УПР1 *СТАТ или УПР2 *СТАТ — позволяет найти и просмотреть статьи многозначных слов, у которых тип управления зависит от значения слова.
7. Запрос: ПОЯС *СТАТ — позволяет найти и просмотреть статьи, в которых информация об управлении дается в виде пояснительной справки.

Научные исследования и публикации

По материалам компьютерной версии Словаря глагольного управления была написана статья [Колодяжная, 2000], основные положения которой изложены ниже.

Частотный словарь схем управления с относящимися к ним словами (на материале словаря управления Д. Э. Розенталя)

С помощью программы UNILEX-D был составлен первичный обратный указатель, каждая статья которого содержала уникальную схему управления и список относящихся к ней слов (например: во что к кому: заглянуть, залезть; из кого-чего: состоять). Всего таких уникальных схем оказалось — 212. Первичный указатель был преобразован в частотный словарь-указатель следующего вида. Для каждой схемы управления, выраженной местоименным словом, была вычислена “мощность” — число, показывающее, сколько слов соответствует данной схеме. Кроме того, была вычислена частота встречаемости схем данной мощности. Весь указатель был упорядочен по возрастанию “мощности” и разделен на новые словарные статьи, в заголовках которых были вынесены два числа: мощность схемы и частота схем данной мощности; содержанием статьи являлся список схем данной мощности с относящимися к ним словами. Всего в таком обратном словаре-указателе оказалось 48 статей.

Ниже приведены первые три статьи полученного словаря.

1. 1 (мощность схемы управления)
71 (частота встреч. схемы мощностью 1)

в ком: ходить; из кого-чего: состоять; из памяти: изгладиться, из-за чего: спорить; из: увертюра; к кому-л.: приверженный; к чему-л.: приверженный; к: увертюра; касательно чего: разговаривать; кого (вин. пад.): бояться; кого (род. пад.): бояться; кого до чего: довести; кого к кому-чему: антипатия; кого к кому: приворожить; кого от кого-чего: изолировать; кого с чем: поздравить кого-, что-л. где: выгрузить; кого-, что-л. куда: выгрузить; кого-, что-л. где: сажать; кого-, что-л. кому: предать; кого-, что-л. куда: сажать; кого-, что-л. чему: предать; кого-кого: у; кого-л. от чего: предостеречь; кого-л. против чего: предостеречь; кого-л. чем: упрекать; кого-чего в кого-что: превращение; кого-чего с кем-чем: взаимодействие; кого-что до чего: довести; кого-что за кем-чем: закрепить; кого-что куда: вселить; кого-что на что: вдохновить; кого-что от чего: предохранить; кого-что по чему: превосходить; кого-что с чего: своротить; кому в чем: отказать; кому к чему: присчитать; кому кого-что: поручить; кому что во что: вменить; кому-чему что: преградить; на (постели): лежать; на кого-л. по чему: походить; на кого-л. чем: походить; на ком-чем: отозваться; около кого-чего: увиваться; откуда: стартовать; относительно чего: разговаривать; подле кого-чего: увиваться; при ком-чем: трепетать; против кого-что: атака; с кем из-за чего: препираться; чего во что: претворение; чего на что: реакция; чего-что от кого: добиваться; чего-л. у кого: добиваться; чем за что: платить; чем на кого-что: похожий; чем по чему: водить; чем с кем: обменяться; что во что-кого: запустить; что и что: сравнить; что из чего: извлечь; что кем: заселить; что на кого: взвести; что на ком: выместить; что у кого: конфисковать; что чему: предпослать; что-л. во что: красить; что-л. подо что: приспособить; что-л. с чем: мыть; что-л. чему: посвятить;

2. 2 (мощность схемы управления)
25 (частота встреч. схемы мощностью 2)

во что к кому: заглянуть, залезть; во что кому: заглянуть, залезть; за чем: контроль, проследить; кого с кем: матч, недоразумение; кого чего: народить, стеречься; кого чему: обучить, учить; кого-л. за что: винить, упрекать; кого-л. на что: приговорить, уполномочить; кого-л. о чем: подозревать, предупредить; кого-что к кому-чему: привязать, причислить; кого-что с кем-чем: сравнить, уравнивать; между кем-чем: взаимодействие, между; на что к кому: броситься, кинуться; подо что: идти², подладиться; про что: разговаривать, рассказать; что кому: предоставить, создать; что куда: забросить, закинуть; что на кого-что: ввалить, призвать; что на чем: акцентировать, базировать; что-л. для кого: открыть, предназначить; что-л. для чего: приспособить, употребить; что-л. из чего: изъять, поливать; что-л. к чему: приспособить, устремить; что-л. на чем: повесить, уставить; что-л. с кого: брать, получить;

3. 3 (мощность схемы управления)
21 (частота встреч. схемы мощностью 3)

вокруг чего: обежать, обойти, объехать; до кого-чего: допускать, относиться, снизить; за кем: вдогонку, записать; охотиться; из-за кого-чего: переживать, страдать, убиваться; из себя: видный, изображать, собой; из чего: глава, исходить, фрагмент; кого кем: величать, гримировать, провозглашать; кого о чем: информировать, оповестить, осведомить; кого от чего: гарантировать, отвадить, страховать; кого-, что-л. где: опустить, положить, посадить; кого-, что-л. куда: опустить, положить, посадить; кого-л. к чему: представить, приговорить, уполномочить; кого-что в чем: ограничить, превосходить, уравнивать; кому-чему в чем: подражать, соболезновать, уподобиться; на что кому: броситься, кинуться, упасть; про кого-что: история, писать, упомянуть; собой: видный, изображать, собой; среди кого: болезни, пользоваться, случаться; что в кого-что: вдохнуть, вколотить, вонзить; что-л. к кому: подбросить¹, предъявить, снести; что-л. чем: красить, мыть, поливать;

Синтаксический словарь Г. А. Золотовой

Общие характеристики

База данных Синтаксического словаря создана на основе книги: Г. А. Золотова. *Синтаксический словарь: Репертуар элементарных единиц русского синтаксиса*. М.: Наука, 1988. Компьютерная версия Синтаксического словаря содержит статьи из первой части книги "Именные синтаксемы". База данных содержит 1374 статьи.

В рабочем состоянии (вместе с индексами) база данных Синтаксического словаря занимает 6 Мб. Компьютерная версия словаря описана в работе [Колодяжная, 1992].

Структура словарной статьи в базе данных

Каждая статья компьютерной версии Синтаксического словаря является уникальной совокупностью атрибутов, относящихся к данной синтаксической конструкции (синтаксеме), и иллюстративных примеров. К основным атрибутам синтаксемы относятся: название (номинатив, позитив, делиберат, комитатив и т. д.), тип (А — свободная, Б — обусловленная, В — связанная), функция и позиция в предложении (I, II, III, II.1 и т. д.), предложно-падежная форма, характеризующая именную синтаксему. Кроме того, в словарной статье, описывающей данную синтаксему, выделяются зоны: развернутого определения синтаксемы, пояснений, иллюстративных примеров, авторов примеров. Ниже приведена полная структура словарной статьи Синтаксического словаря в базе данных:

ЗАГЛ	номер статьи в книге	1
СИНТ	название синтаксемы	2
ФОРМА	падеж+предлог	3
ТИП	тип (свободная, обусловленная, связанная)	4
НОМТ	номер типа	5
ДЕФС	определение синтаксемы	6
ФУНКР	функция+позиция в предложении	7
ПОЯСФ	пояснение к функции	8
ВАРП	вариант позиции	9
ПРИМ	пример употребления синтаксемы	10
АВТОР	автор примера	11
*СТАТ	вся статья	1 11
*АТР	блок атрибутов	1 9
*ПРИМ	блок примеров	10 11

Пример словарной статьи в базе данных:

ЗАГЛ	0001
СИНТ	Номинатив
ФОРМА	ИМЕНИТЕЛЬНЫЙ
ТИП	A.
НОМТ	01.
ДЕФС	Номинатив (именительный наименования).
ФУНКР	I.
ПОЯСФ	В заголовках, на вывесках, на этикетках:
ПРИМ	Литературная газета.
ПРИМ	Орфографический словарь.
ПРИМ	Москва-сортировочная.
ПРИМ	Кафе "Ромашка".
ПРИМ	Галантерея.
ПРИМ	Молоко.

ПРИМ Чай грузинский.

Особенности функционирования словаря в UNILEX-D

При вводе информации в базу данных использовались функции системы UNILEX-D: Структура статьи, Ввод, Редактирование. При работе со Словарем в системе UNILEX-D обеспечивается доступ к словарной статье по любому из возможных “входов”:

- от названия синтаксемы,
- от предложно-падежной формы,
- от типа синтаксемы,
- от типа функции и позиции,
- от варианта позиции,
- от автора примета,
- от примера.

Научные исследования и публикации

Возможность совершать над компьютерным словарем в системе UNILEX-D различные преобразования позволила для Синтаксического словаря создать таблицы, содержащие количественные параметры и построить некоторые указатели к Словарю.

Некоторые количественные параметры Синтаксического словаря:

- количество статей — 1374;
- количество различных семантических категорий синтаксем — 294;
- количество различных наборов функций и позиций — 47;
- количество различных типов предложно-падежных форм — 40.

Распределение синтаксем по типам:

- А. (свободные) 701
- Б. (связанные) 595
- В. (обусловленные) 78

Наиболее частые синтаксемы:

- Локатив 132
- Директив 77
- Темпоратив 65
- Темпоратив 1 32
- Темпоратив 2 27

Наиболее частые типы функций и позиций:

- Ш.2. 257
- П. 2. 193
- Ш.1. 172
- Г. 144
- П.4. 139
- П.3. 80

Наиболее частые типы предложно-падежных форм:

- ПРЕДЛОЖНЫЙ В 102
- ВИНИТЕЛЬНЫЙ НА 89
- ВИНИТЕЛЬНЫЙ В 88
- ПРЕДЛОЖНЫЙ НА 71

Указатели к Синтаксическому словарю

Для синтаксического словаря, структура статьи которого состоит из 11 компонент, потенциально может быть построено столько указателей, сколько различных пар можно образовать из этого числа. Но, разумеется, не все указатели могут быть интересны в лексикографическом смысле. С лексикографической точки зрения наиболее интересны указатели для следующих пар компонент:

- СИНТ ФОРМА
- ФОРМА СИНТ
- СИНТ ФУНКР
- ФУНКР СИНТ

ТИП	ФОРМА
ФОРМА	ТИП
ТИП	СИНТ
СИНТ	ТИП.

В книжном варианте словаря приведен только один из возможных указателей — указатель одноименных синтаксем (инверсия: СИНТ ФОРМА); в книге этот указатель, составленный вручную, дается в неполном виде. Ниже указатель одноименных синтаксем дается в полном виде, полученном с помощью функции Инверсия системы UNILEX-D. Некоторые виды указателей к Синтаксическому словарю, полученные на основе его базы данных, опубликованы в статье [Золотова, Колодяжная, Шумарина, 1991].

УКАЗАТЕЛЬ: СИНТАКСЕМА ФОРМА (цифра перед формой обозначает частотность)

Авторизатор 1 РОДИТЕЛЬНЫЙ ДЛЯ, 1 РОДИТЕЛЬНЫЙ У
 Авторизатор квалификативной оценки 1 ДАТЕЛЬНЫЙ
 Авторизатор с посессивно-локат. оттенком 1 РОДИТЕЛЬНЫЙ У
 Адресат 7 ДАТЕЛЬНЫЙ, 7 ДАТЕЛЬНЫЙ К
 Адресат – каузатор отрицат. эмоций 1 РОДИТЕЛЬНЫЙ ПРОТИВ
 Адресат каузирующего обращения 1 РОДИТЕЛЬНЫЙ У
 В компликативной модели 2 ВИНИТЕЛЬНЫЙ, 2 ПРЕДЛОЖНЫЙ НА
 В модально-врем. модификации – предстояние действия 1 ПРЕДЛОЖНЫЙ В
 В страдат. модификациях 3 ТВОРИТЕЛЬНЫЙ
 “Второй дат.” 1 ДАТЕЛЬНЫЙ
 В экспр.-модальной модификации а) субъект обязат. занятий 1 ПРЕДЛОЖНЫЙ НА
 Вокатив 1 ИМЕНИТЕЛЬНЫЙ
 Вступл. в социальную зависимость 3 ВИНИТЕЛЬНЫЙ ПОД
 Выразитель связи с именем собственным 1 ДАТЕЛЬНЫЙ ПО
 Выход из социальной зависимости – с фаз. глаг. 1 РОДИТЕЛЬНЫЙ ИЗ-ПОД
 Генератив 9 РОДИТЕЛЬНЫЙ ИЗ,ИЗО
 Действие в аналитической пассивной конструкции 2 ДАТЕЛЬНЫЙ
 Действие в модально-стилистической модификации 1 ДАТЕЛЬНЫЙ
 Действие в фазисн. модификации 1 РОДИТЕЛЬНЫЙ С
 Действие при фазисных глаголах 1 ДАТЕЛЬНЫЙ К
 Действие, состояние в экспр.-модальной модификации 2 ДАТЕЛЬНЫЙ
 Делиберат борьбы 6 РОДИТЕЛЬНЫЙ ПРОТИВ
 Делиберат восприятия 1 ВИНИТЕЛЬНЫЙ, 1 ИМЕНИТЕЛЬНЫЙ
 Делиберат при глаголе верить, веровать 2 ВИНИТЕЛЬНЫЙ В
 Делиберат при глаголе играть 2 ВИНИТЕЛЬНЫЙ В
 Делиберат речевого воздействия 2 ВИНИТЕЛЬНЫЙ НА
 Делиберативно-каузативная 2 ВИНИТЕЛЬНЫЙ НА
 Дестинатив 4 ВИНИТЕЛЬНЫЙ В, 5 ВИНИТЕЛЬНЫЙ ПОД, 7 ДАТЕЛЬНЫЙ К, 4 РОДИТЕЛЬНЫЙ ДЛЯ, 1 РОДИТЕЛЬНЫЙ ОТ
 Дестинатив1 8 ВИНИТЕЛЬНЫЙ НА
 Дестинатив2 6 ВИНИТЕЛЬНЫЙ НА
 Дестинатив3 5 ВИНИТЕЛЬНЫЙ НА
 Деструктив 1 РОДИТЕЛЬНЫЙ ОТ
 Дименсив 2 ВИНИТЕЛЬНЫЙ, 3 ВИНИТЕЛЬНЫЙ В, 2 ВИНИТЕЛЬНЫЙ ЗА, 5 ВИНИТЕЛЬНЫЙ ПО, 2 ВИНИТЕЛЬНЫЙ ЧЕРЕЗ, 4 РОДИТЕЛЬНЫЙ ДО
 Дименсив1 5 ВИНИТЕЛЬНЫЙ НА
 Дименсив2 4 ВИНИТЕЛЬНЫЙ НА
 Дименсив локативный 6 РОДИТЕЛЬНЫЙ ОТ
 Дименсив при глаголах весить, стоять 1 ВИНИТЕЛЬНЫЙ
 Дименсив темпоративный 2 РОДИТЕЛЬНЫЙ ОТ
 Дименсивно-квантитативная (с оттенком приблизит.) 14 ВИНИТЕЛЬНЫЙ С
 Директ.-репродуктивная 3 РОДИТЕЛЬНЫЙ С
 Директив 10 ВИНИТЕЛЬНЫЙ В, 4 ВИНИТЕЛЬНЫЙ ЗА, 8 ВИНИТЕЛЬНЫЙ НА, 5 ВИНИТЕЛЬНЫЙ ПОД, 15 ДАТЕЛЬНЫЙ К, 6 РОДИТЕЛЬНЫЙ ИЗ-ЗА, 5 РОДИТЕЛЬНЫЙ ИЗ-ПОД, 6 РОДИТЕЛЬНЫЙ ОТ, 16 РОДИТЕЛЬНЫЙ С, ТВОРИТЕЛЬНЫЙ ЗА
 Директив1 8 РОДИТЕЛЬНЫЙ ИЗ,ИЗО, 3 РОДИТЕЛЬНЫЙ ПРОТИВ
 Директив2 9 РОДИТЕЛЬНЫЙ ИЗ,ИЗО, 2 РОДИТЕЛЬНЫЙ ПРОТИВ
 Директив с дименс. оттенком 8 РОДИТЕЛЬНЫЙ ДО
 Директивно-темпоративная 2 РОДИТЕЛЬНЫЙ ИЗ,ИЗО
 Дистрибутив 1 1 ВИНИТЕЛЬНЫЙ ПО 10 ДАТЕЛЬНЫЙ ПО 6 ВИНИТЕЛЬНЫЙ НА
 Дистрибутив 2 2 ВИНИТЕЛЬНЫЙ НА
 Дополняемое частью целое 4 ДАТЕЛЬНЫЙ К
 Изменение социального положения 3 РОДИТЕЛЬНЫЙ ИЗ,ИЗО
 Инструментив 2 ВИНИТЕЛЬНЫЙ В, 2 ВИНИТЕЛЬНЫЙ НА, 2 ПРЕДЛОЖНЫЙ НА, 3 РОДИТЕЛЬНЫЙ ИЗ,ИЗО, 2 РОДИТЕЛЬНЫЙ С, 4 ТВОРИТЕЛЬНЫЙ

Инструментив-каузатор 1 ИМЕНИТЕЛЬНЫЙ
 Интенсив 1 1 РОДИТЕЛЬНЫЙ ДО
 Интенсив каузированный моносубъектный 1 РОДИТЕЛЬНЫЙ ДО
 Интенсив каузированный полисубъектный 1 РОДИТЕЛЬНЫЙ ДО
 Интенсив каузированный моносубъектный 2 РОДИТЕЛЬНЫЙ ДО
 Интенсив каузированный полисубъектный 1 РОДИТЕЛЬНЫЙ ДО
 Интерпретация понятия 2 ПРЕДЛОЖНЫЙ В
 Интерпретируемое понятие 1 ТВОРИТЕЛЬНЫЙ ПОД
 Исходная причина или условие в компликативн. модели 1 РОДИТЕЛЬНЫЙ ИЗ,ИЗО
 Каузатив 3 ВИНИТЕЛЬНЫЙ ЗА, 2 ПРЕДЛОЖНЫЙ ПРИ, 4 РОДИТЕЛЬНЫЙ ИЗ,ИЗО, 10
 РОДИТЕЛЬНЫЙ ИЗ-ЗА, 2 ТВОРИТЕЛЬНЫЙ ЗА, 4 ТВОРИТЕЛЬНЫЙ ПОД
 Каузатив1 3 ДАТЕЛЬНЫЙ ПО, 8 РОДИТЕЛЬНЫЙ ОТ, 6 РОДИТЕЛЬНЫЙ С
 Каузатив2 3 ДАТЕЛЬНЫЙ ПО, 2 РОДИТЕЛЬНЫЙ ОТ, 1 РОДИТЕЛЬНЫЙ С
 Каузатор отвлеч. 1 ТВОРИТЕЛЬНЫЙ
 Каузатор отвлеченный 1 ИМЕНИТЕЛЬНЫЙ
 Каузатор потенциального действия 1 РОДИТЕЛЬНЫЙ ОТ
 Каузатор прошедшего воздействия 1 РОДИТЕЛЬНЫЙ ОТ
 Каузатор реактивного действия 5 ВИНИТЕЛЬНЫЙ НА
 Каузатор состояния или эмоц. отношения 1 ДАТЕЛЬНЫЙ
 Каузатор стихийного воздействия 1 ТВОРИТЕЛЬНЫЙ
 Каузатор-делиберат 1 ТВОРИТЕЛЬНЫЙ
 Каузатор-делиберат эмоц. сост. 2 ВИНИТЕЛЬНЫЙ ЗА
 Каузируемая ассоциация 1 ВИНИТЕЛЬНЫЙ
 Каузируемое действие или качество 1 ДАТЕЛЬНЫЙ
 Каузируемое действие — делиберат обучения 1 ДАТЕЛЬНЫЙ
 Качество в фазисно-каузат. модификациях 1 ВИНИТЕЛЬНЫЙ
 Квалитатив 2 ВИНИТЕЛЬНЫЙ В, 4 ТВОРИТЕЛЬНЫЙ С
 Квалитатив с вспомогат. глаг. 1 ТВОРИТЕЛЬНЫЙ
 Квалификация замещенного предмета 4 ВИНИТЕЛЬНЫЙ ЗА
 Квалификация генетического подобия 3 ВИНИТЕЛЬНЫЙ В
 Квантитатив 3 ВИНИТЕЛЬНЫЙ В, 4 РОДИТЕЛЬНЫЙ ДО
 Квантитатив (аппроксим.) 6 ВИНИТЕЛЬНЫЙ ЗА, 2 ВИНИТЕЛЬНЫЙ ПОД
 Класс предметов в компаративно-оцен. модели 1 РОДИТЕЛЬНЫЙ СРЕДИ
 Комитатив 3 ВИНИТЕЛЬНЫЙ ПОД, 2 ПРЕДЛОЖНЫЙ ПРИ, 3 ТВОРИТЕЛЬНЫЙ ЗА
 Комитатив 1 6 ТВОРИТЕЛЬНЫЙ С
 Комитатив 2 3 ТВОРИТЕЛЬНЫЙ С
 Комитатив 3 7 ТВОРИТЕЛЬНЫЙ С
 Компаративный предикат 2 ТВОРИТЕЛЬНЫЙ
 Компонент проявления действий, сост. нравств.-повед. сфер 4 ТВОРИТЕЛЬНЫЙ ПЕРЕД
 Коррелятив 5 ВИНИТЕЛЬНЫЙ В, 4 ВИНИТЕЛЬНЫЙ ПОД, 5 ДАТЕЛЬНЫЙ ПО, 1
 ПРЕДЛОЖНЫЙ ПО
 Коррелятив (лексически ограничен) 3 ДАТЕЛЬНЫЙ К
 Критерий 2 ТВОРИТЕЛЬНЫЙ
 Критерий квалификации 2 ДАТЕЛЬНЫЙ ПО
 Критерий квалификации и оценки 4 ВИНИТЕЛЬНЫЙ НА
 Критерий несоответствия 1 РОДИТЕЛЬНЫЙ ДЛЯ
 Критерий оценки 5 ПРЕДЛОЖНЫЙ В, 1 РОДИТЕЛЬНЫЙ ПРОТИВ
 Критерий признака 2 ПРЕДЛОЖНЫЙ НА
 Критерий сравнения 1 ДАТЕЛЬНЫЙ ПО
 Критерий сравнения, оценки 3 ТВОРИТЕЛЬНЫЙ ПЕРЕД
 Лимитатив 1 ПРЕДЛОЖНЫЙ В
 Лицо – целое по отношению к вертикальной мере 1 ДАТЕЛЬНЫЙ
 Лицо как целое по отн. к принадл. ему части 1 РОДИТЕЛЬНЫЙ У
 Лицо, по отн. к которому оценивается соответствие 1 ДАТЕЛЬНЫЙ
 Лицо-посессор – целое по отношению к части 1 ДАТЕЛЬНЫЙ
 Личный субъект каузированного состояния 1 ТВОРИТЕЛЬНЫЙ
 Логич. следствие в компликат. глаголах 1 ДАТЕЛЬНЫЙ К
 Логич. следствие в комплик. модели 1 ПРЕДЛОЖНЫЙ О
 Локализующая часть объекта напр. действия 2 ВИНИТЕЛЬНЫЙ В
 Локатив 4 ВИНИТЕЛЬНЫЙ ПО, 21 ПРЕДЛОЖНЫЙ В, 16 ПРЕДЛОЖНЫЙ НА, 5
 ПРЕДЛОЖНЫЙ ПРИ, 6 РОДИТЕЛЬНЫЙ МЕЖДУ, 5 РОДИТЕЛЬНЫЙ ПРОТИВ, 8
 РОДИТЕЛЬНЫЙ СРЕДИ, 16 РОДИТЕЛЬНЫЙ У, 13 ТВОРИТЕЛЬНЫЙ ЗА, 7
 ТВОРИТЕЛЬНЫЙ МЕЖДУ, 9 ТВОРИТЕЛЬНЫЙ НАД, 11 ТВОРИТЕЛЬНЫЙ ПЕРЕД, 11
 ТВОРИТЕЛЬНЫЙ ПОД
 Локатив – видимое в компликат. модели 2 ТВОРИТЕЛЬНЫЙ ЗА
 Локатив посессивный 11 РОДИТЕЛЬНЫЙ У
 Локатив с отнош. администрат. подчинения 4 ПРЕДЛОЖНЫЙ ПРИ
 Локативно-объектная 5 ТВОРИТЕЛЬНЫЙ НАД
 Медиатив 1 ВИНИТЕЛЬНЫЙ НА, 3 ДАТЕЛЬНЫЙ ПО 3 ВИНИТЕЛЬНЫЙ ЧЕРЕЗ, 6
 ПРЕДЛОЖНЫЙ НА, 3 ТВОРИТЕЛЬНЫЙ

Медиатив2 2 ВИНИТЕЛЬНЫЙ НА, 2 ВИНИТЕЛЬНЫЙ ЧЕРЕЗ, 4 ПРЕДЛОЖНЫЙ НА, 4
 ТВОРИТЕЛЬНЫЙ
 Медиатив3 5 ПРЕДЛОЖНЫЙ НА
 Модальная квалификация действия 2 ТВОРИТЕЛЬНЫЙ ПОД
 Модальная часть предиката 2 ПРЕДЛОЖНЫЙ В
 Неопред.-мн. субъект действий, реакций 2 ТВОРИТЕЛЬНЫЙ МЕЖДУ
 Неопред.-мн. субъект действия или реакции 3 РОДИТЕЛЬНЫЙ СРЕДИ
 Номинатив 6 ИМЕНИТЕЛЬНЫЙ
 Носитель качества 1 ВИНИТЕЛЬНЫЙ
 Носитель признака 1 РОДИТЕЛЬНЫЙ ДЛЯ
 Носитель приобретаемого признака 2 ДАТЕЛЬНЫЙ
 Объект – каузатор эмоц. отношения 2 ВИНИТЕЛЬНЫЙ
 Объект – каузируемый субъект потенц. действия 3 ДАТЕЛЬНЫЙ
 Объект – каузируемый субъект 1 ВИНИТЕЛЬНЫЙ
 Объект – каузируемый субъект состояния 1 ВИНИТЕЛЬНЫЙ
 Объект – субъект вторичного предиката 1 ВИНИТЕЛЬНЫЙ
 Объект действия при переходн. акцион. глаголе 1 ВИНИТЕЛЬНЫЙ
 Объект действия, отношения к лицу 3 ТВОРИТЕЛЬНЫЙ С
 Объект каузативного воздействия 2 ВИНИТЕЛЬНЫЙ НА
 Объект контактно-лок. действия 2 РОДИТЕЛЬНЫЙ
 Объект контактно-направл. действия 1 ВИНИТЕЛЬНЫЙ ПОД
 Объект наблюдения 2 ТВОРИТЕЛЬНЫЙ ЗА
 Объект непроизвольного воздействия 1 ВИНИТЕЛЬНЫЙ
 Объект при глаг. бить, стрелять 2 ДАТЕЛЬНЫЙ ПО
 Объект при глаг. отличать(ся) 1 РОДИТЕЛЬНЫЙ ОТ
 Объект при глаг. реализации превосходства 4 ТВОРИТЕЛЬНЫЙ НАД
 Объект при глаголах лексически огранич. круга 1 ДАТЕЛЬНЫЙ
 Объект при глаголах руководства 2 ТВОРИТЕЛЬНЫЙ
 Объект при глаголах проявления отношения 3 ТВОРИТЕЛЬНЫЙ НАД
 Объект при глаголе жениться 2 ПРЕДЛОЖНЫЙ НА
 Объект при отглагол. именах 1 РОДИТЕЛЬНЫЙ
 Объект при перех. глаг. с отрицанием 1 РОДИТЕЛЬНЫЙ
 Объект социально-этич. действия 2 ВИНИТЕЛЬНЫЙ ЗА
 Объект сравнения 3 ВИНИТЕЛЬНЫЙ НА
 Объект эмоц. отношения 3 ДАТЕЛЬНЫЙ К
 Объект-директив направл. действия 2 ВИНИТЕЛЬНЫЙ В
 Объект-каузатив при глаг. “горестного чувства” 2 ДАТЕЛЬНЫЙ ПО
 Объект-каузатор “горестного чувства” 2 ПРЕДЛОЖНЫЙ ПО
 Объект-каузатор отрицают. эмоций 4 ВИНИТЕЛЬНЫЙ НА
 Объект-каузатор оценочного отношения 1 ИМЕНИТЕЛЬНЫЙ
 Объект-каузатор эмоц. отношений 3 ТВОРИТЕЛЬНЫЙ
 Объект-субъект состояния в страд. обороте 1 ИМЕНИТЕЛЬНЫЙ
 Объектно-делиберативная 1 ТВОРИТЕЛЬНЫЙ
 Объектно-делиберативная при группе отвлеченных глаг. 1 ПРЕДЛОЖНЫЙ НА
 Объектно-количественная 2 РОДИТЕЛЬНЫЙ 2 ТВОРИТЕЛЬНЫЙ
 Объектно-орудийная 1 ТВОРИТЕЛЬНЫЙ
 Обществ. обусловленное состояние лица 3 ПРЕДЛОЖНЫЙ НА
 Ожидаемое действие, состояние 1 ТВОРИТЕЛЬНЫЙ ЗА
 Опосредов. субъект неактуального действия 2 РОДИТЕЛЬНЫЙ ОТ
 Опосредов. субъект потенциального действия 2 РОДИТЕЛЬНЫЙ ОТ
 Ориентирующая часть предмета 1 ТВОРИТЕЛЬНЫЙ
 Основание-соответствие 2 ДАТЕЛЬНЫЙ ПО
 Отправной предмет сопоставления, иерархии 5 ДАТЕЛЬНЫЙ
 Отсутствующий признак 10 РОДИТЕЛЬНЫЙ БЕЗ
 Оцениваемое действие в полипредикат. конструкциях 8 ПРЕДЛОЖНЫЙ В
 Ошибочная идентификация 1 ВИНИТЕЛЬНЫЙ ЗА
 Параметр измерения 1 ТВОРИТЕЛЬНЫЙ
 Партитив 3 ПРЕДЛОЖНЫЙ О, 4 РОДИТЕЛЬНЫЙ ИЗ, ИЗО
 Положение социальн. зависимости 5 ТВОРИТЕЛЬНЫЙ ПОД
 Поссесивн. предикат 2 ВИНИТЕЛЬНЫЙ
 Поссесивный объект 6 ТВОРИТЕЛЬНЫЙ
 Поссесивный субъект-отправитель 4 РОДИТЕЛЬНЫЙ ОТ
 Поссесивный субъект-отправитель 1 РОДИТЕЛЬНЫЙ ОТ
 Потенсив 8 ВИНИТЕЛЬНЫЙ НА, 3 ДАТЕЛЬНЫЙ К, 1 ПРЕДЛОЖНЫЙ НА, 1
 РОДИТЕЛЬНЫЙ ДЛЯ, 2 РОДИТЕЛЬНЫЙ ДО, 1 ТВОРИТЕЛЬНЫЙ
 Потенсив при каузат.-отложит. глаг. 3 РОДИТЕЛЬНЫЙ ОТ
 Потенсив при модально-отложит. глаг. 2 РОДИТЕЛЬНЫЙ ОТ
 Потенц. субъект-посессор в модальных модификациях 2 ДАТЕЛЬНЫЙ
 Предикат акциональный 1 ИМЕНИТЕЛЬНЫЙ
 Предикат именной 11 ТВОРИТЕЛЬНЫЙ

Предикат качественнй 1 ИМЕНИТЕЛЬНЫЙ
 Предикат квалиф.-оценочный 1 ИМЕНИТЕЛЬНЫЙ
 Предикат посессивного признака 1 ИМЕНИТЕЛЬНЫЙ
 Предикат состояния лица 1 ИМЕНИТЕЛЬНЫЙ
 Предикат состояния среды 1 ИМЕНИТЕЛЬНЫЙ
 Предмет локативно-контактного соотношения 1 ВИНИТЕЛЬНЫЙ В
 Предмет обмена 2 ВИНИТЕЛЬНЫЙ ЗА
 Предмет сопоставления с т.зр. социальных связей 1 ДАТЕЛЬНЫЙ
 Предмет хватат. движения 1 ВИНИТЕЛЬНЫЙ ЗА
 Предмет, локализующий столкновение 3 ВИНИТЕЛЬНЫЙ О
 Предпредикат в каузации превращения 5 РОДИТЕЛЬНЫЙ ИЗ,ИЗО
 Предстояние действия в экспр.-модал. модификации 1 ПРЕДЛОЖНЫЙ НА
 Предшеств. или сопутств. состояние 2 ВИНИТЕЛЬНЫЙ СКВОЗЬ
 Прежний владелец при изменении посессивных отношений 1 РОДИТЕЛЬНЫЙ У
 Прежний посессор при изменении посесс. отношений 2 РОДИТЕЛЬНЫЙ С
 При глаг. обойтись 1 РОДИТЕЛЬНЫЙ БЕЗ
 При глаголе поздравлять 1 ТВОРИТЕЛЬНЫЙ С
 Признак внешнего покрытия 6 ПРЕДЛОЖНЫЙ В
 Признак противоречащий предикативному 1 ПРЕДЛОЖНЫЙ ПРИ
 Признак, каузирующий логич. результат 2 ДАТЕЛЬНЫЙ ПО
 Причинно-следственные компоненты с компликативными глаголами 3 ТВОРИТЕЛЬНЫЙ
 Результат каузации разделения 2 ВИНИТЕЛЬНЫЙ В
 Результат каузации соединения 3 ВИНИТЕЛЬНЫЙ В
 Род. агентивный 2 РОДИТЕЛЬНЫЙ
 Род. даты 7 РОДИТЕЛЬНЫЙ
 Род. качественнй 4 РОДИТЕЛЬНЫЙ
 Род. количественнй 8 РОДИТЕЛЬНЫЙ
 Род. меры соответствия 1 РОДИТЕЛЬНЫЙ
 Род. модально-каузативный 4 РОДИТЕЛЬНЫЙ
 Род. носителя признака 1 РОДИТЕЛЬНЫЙ
 Род. посессивный 2 РОДИТЕЛЬНЫЙ
 Род. при отрицании 3 РОДИТЕЛЬНЫЙ
 Род. прикомпаративный 3 РОДИТЕЛЬНЫЙ
 Родовое понятие в классификат. моделях 2 ДАТЕЛЬНЫЙ К
 Рубрикатив 8 ДАТЕЛЬНЫЙ ПО
 С глаг. идти замуж, свататься 1 ВИНИТЕЛЬНЫЙ ЗА
 Ситуатив 5 ПРЕДЛОЖНЫЙ В, 3 ПРЕДЛОЖНЫЙ НА, 6 ТВОРИТЕЛЬНЫЙ ПОД
 "Скрытый" агенс 1 РОДИТЕЛЬНЫЙ У
 Состояние лица в экспр.-мод. модификации 1 РОДИТЕЛЬНЫЙ ДО
 Состояние, положение субъекта 1) физич. эмоцион. состояния 4 ПРЕДЛОЖНЫЙ В
 Состояние, положение субъекта 2) трудное, зависимое положение 3 ПРЕДЛОЖНЫЙ В
 Состояние, положение субъекта 3) временное состояние 5 ПРЕДЛОЖНЫЙ В
 Состояние, положение субъекта 4) состояние отношений с др. лицами 3 ПРЕДЛОЖНЫЙ В
 Состояние, положение субъекта 5) результативно-посессивное 1 ПРЕДЛОЖНЫЙ В
 Состояние, положение субъекта 6) состояние движения, занятия 5 ПРЕДЛОЖНЫЙ В
 Социальная квалификация лица в фаз.-каузативн. модификации 1 РОДИТЕЛЬНЫЙ С
 Социально-темпоральная характеристика лица 3 ПРЕДЛОЖНЫЙ НА
 Социально-темпоральная характеристика лица 1 ПРЕДЛОЖНЫЙ НА
 Способ действия 4 ВИНИТЕЛЬНЫЙ ПОД
 Способ или мера действия 3 ВИНИТЕЛЬНЫЙ В
 Способ именованя 4 ТВОРИТЕЛЬНЫЙ ПОД
 Субъект – владелец информации 1 РОДИТЕЛЬНЫЙ У
 Субъект – отправитель информации 4 РОДИТЕЛЬНЫЙ ОТ
 Субъект в модальной модификации 1 ДАТЕЛЬНЫЙ
 Субъект взаимных действий и отнош. 1 РОДИТЕЛЬНЫЙ МЕЖДУ 4 ТВОРИТЕЛЬНЫЙ
 МЕЖДУ
 Субъект взаимных отношений 1 РОДИТЕЛЬНЫЙ У
 Субъект владения 2 ТВОРИТЕЛЬНЫЙ ЗА
 Субъект возрастн. признака 1 ДАТЕЛЬНЫЙ
 Субъект каузированного состояния 1 ДАТЕЛЬНЫЙ
 Субъект каузируемых эмоций 2 РОДИТЕЛЬНЫЙ У
 Субъект качественнй 2 ДАТЕЛЬНЫЙ 3 РОДИТЕЛЬНЫЙ У
 Субъект непроизвольного действия-состояния 1 ДАТЕЛЬНЫЙ
 Субъект непроизвольного восприятия 2 ДАТЕЛЬНЫЙ
 Субъект ожидаемого действия 1 ТВОРИТЕЛЬНЫЙ ЗА
 Субъект посессивный 3 РОДИТЕЛЬНЫЙ У
 Субъект потенциального действия 1 ДАТЕЛЬНЫЙ
 Субъект предикативного признака 4 ИМЕНИТЕЛЬНЫЙ
 Субъект речи в диалогическом зачине (разг.) 1 РОДИТЕЛЬНЫЙ У
 Субъект состояния 1 ВИНИТЕЛЬНЫЙ 1 ДАТЕЛЬНЫЙ 2 РОДИТЕЛЬНЫЙ У

Субъект экспрессивно-модальной модификации 1 ДАТЕЛЬНЫЙ
 Субъект эмоц.-оценочного отношения 1 ДАТЕЛЬНЫЙ
 Субъект-носитель потенц. признака 1 РОДИТЕЛЬНЫЙ ИЗ,ИЗО
 Субъект-носитель признака 7 ПРЕДЛОЖНЫЙ В, 4 ТВОРИТЕЛЬНЫЙ ЗА
 Субъект-соучастник 10 ТВОРИТЕЛЬНЫЙ С
 Субъект-участник 2 РОДИТЕЛЬНЫЙ У
 Сурсив 4 РОДИТЕЛЬНЫЙ ИЗ,ИЗО, 2 РОДИТЕЛЬНЫЙ ОТ
 Тематив 2 ПРЕДЛОЖНЫЙ В, 15 ТВОРИТЕЛЬНЫЙ С
 Темпор.-директивная 4 РОДИТЕЛЬНЫЙ С
 Темпоратив 1 ВИНИТЕЛЬНЫЙ ПО, 5 ВИНИТЕЛЬНЫЙ ПОД, 3 ДАТЕЛЬНЫЙ К, 6
 ДАТЕЛЬНЫЙ ПО, 11 ПРЕДЛОЖНЫЙ В, 3 ПРЕДЛОЖНЫЙ НА, 8 РОДИТЕЛЬНЫЙ ДО,
 4 РОДИТЕЛЬНЫЙ ИЗ,ИЗО, 3 РОДИТЕЛЬНЫЙ МЕЖДУ, 7 РОДИТЕЛЬНЫЙ С, 5
 ТВОРИТЕЛЬНЫЙ, 3 ТВОРИТЕЛЬНЫЙ МЕЖДУ, 6 ТВОРИТЕЛЬНЫЙ ПЕРЕД
 Темпоратив1 4 ВИНИТЕЛЬНЫЙ В, 3 ВИНИТЕЛЬНЫЙ ЗА, 4 ВИНИТЕЛЬНЫЙ НА, 5
 ВИНИТЕЛЬНЫЙ ЧЕРЕЗ, 3 ПРЕДЛОЖНЫЙ ПО, 3 ПРЕДЛОЖНЫЙ ПРИ, 3
 РОДИТЕЛЬНЫЙ СРЕДИ, 6 ТВОРИТЕЛЬНЫЙ ЗА, 1 ТВОРИТЕЛЬНЫЙ С
 Темпоратив1 с оттенком кондицион. 4 ПРЕДЛОЖНЫЙ ПРИ
 Темпоратив2 16 ВИНИТЕЛЬНЫЙ В, 2 ВИНИТЕЛЬНЫЙ ЗА, 4 ВИНИТЕЛЬНЫЙ НА, 1
 ВИНИТЕЛЬНЫЙ ЧЕРЕЗ, 1 РОДИТЕЛЬНЫЙ СРЕДИ, 1 ТВОРИТЕЛЬНЫЙ ЗА, 2
 ТВОРИТЕЛЬНЫЙ С
 Темпоратив 2 приуроч. к лицу 2 ПРЕДЛОЖНЫЙ ПРИ
 Темпоратив 3 2 ВИНИТЕЛЬНЫЙ НА
 Темпоратив даты 2 РОДИТЕЛЬНЫЙ ОТ
 Темпоративно-локативная 7 ПРЕДЛОЖНЫЙ НА, 2 ПРЕДЛОЖНЫЙ ПРИ
 Транзитив 9 ВИНИТЕЛЬНЫЙ СКВОЗЬ, 14 ДАТЕЛЬНЫЙ ПО, 2 ТВОРИТЕЛЬНЫЙ
 Транзитив1 6 ВИНИТЕЛЬНЫЙ ЧЕРЕЗ
 Транзитив2 6 ВИНИТЕЛЬНЫЙ ЧЕРЕЗ
 Транзитив-темпоратив 1 ВИНИТЕЛЬНЫЙ СКВОЗЬ
 Трансгрессив 3 ВИНИТЕЛЬНЫЙ В
 Фабрикатив 5 РОДИТЕЛЬНЫЙ ИЗ,ИЗО
 Фаза движения, развития 5 ПРЕДЛОЖНЫЙ НА
 Фаза действия 4 ВИНИТЕЛЬНЫЙ В, 2 ВИНИТЕЛЬНЫЙ ЗА
 Фазисная характеристика лица по социальн. категории 8 ВИНИТЕЛЬНЫЙ В
 Фазисное состояние субъекта 2 ВИНИТЕЛЬНЫЙ В
 Финитив 5 ВИНИТЕЛЬНЫЙ ЗА, 2 ВИНИТЕЛЬНЫЙ ПО, 7 РОДИТЕЛЬНЫЙ ДЛЯ, 4
 ТВОРИТЕЛЬНЫЙ ЗА
 Финитно-фазисное состояние, положение 2 РОДИТЕЛЬНЫЙ ИЗ,ИЗО
 Функциональное положение предмета 1 ПРЕДЛОЖНЫЙ НА
 Характеризующее наличие предмета 2 ПРЕДЛОЖНЫЙ ПРИ
 Характеристика по бывшему содержанию 1 РОДИТЕЛЬНЫЙ ИЗ-ПОД
 Характеристика по принадл. к классу предметов 1 РОДИТЕЛЬНЫЙ МЕЖДУ
 Характеристика по принадлежности к классу лиц, предметов 2 РОДИТЕЛЬНЫЙ СРЕДИ
 Характеристика по принадлежности к социальной группе 6 ПРЕДЛОЖНЫЙ В
 Характеристика цвета или поверхности 4 ПРЕДЛОЖНЫЙ В
 Экспр.-модальная модификация б) обязательное занятие 1 ПРЕДЛОЖНЫЙ НА
 Эмоцион.-оценочная реакция 2 ВИНИТЕЛЬНЫЙ В
 Эмоцион., физического состояния 1 ПРЕДЛОЖНЫЙ НА

Информационная система Архив лингвистических источников Машинного фонда русского языка

Общие характеристики

Компьютерный Архив лингвистических источников создан в рамках общего проекта по созданию Машинного фонда русского языка, реализованного в отделе Машинного фонда русского языка Института русского языка РАН. Принципы организации архива источников подробно изложены в работе В. М. Андрущенко [Андрущенко, 1992].

Архив лингвистических источников Машинного фонда русского языка — это совокупность лингвистических источников (текстов, словарей, грамматик и т. п.), перенесенных на машиночитаемые носители информации.

Каждый источник описывается в специальной карте, совокупность которых составляет базу данных Архива лингвистических источников. База данных Архива является электронной картотекой, работающей под управлением программы UNILEX-D. В карте описания лингвистического источника содержится информация, связанная, с одной стороны, с характеристиками исходного текста (книги, рукописи и т. п.), с другой — с характеристиками магнитной копии источника.

Всего в базе данных Архива содержится более 200 карт описаний лингвистических источников, хранящихся в Машинном фонде русского языка.

Полное описание Архива лингвистических источников и его компьютерной версии содержится в работах [Колодяжная, 1995; Колодяжная, 1996]

Структура записи в базе данных

Карта описания лингвистического источника состоит из 23-х зон, часть из которых является обязательными (тип источника, жанр источника и т. д.), а часть — факультативными (адресат копии, пользователи и т. д.). Ниже приведен список полей карты описания с их mnemonicскими именами и пример одного из описаний.

1	НОМЕР	номер единицы описания
2	ТИП	тип источника
3	ХРАН	способ хранения
4	КЖАНР	код жанра
5	ЖАНР	жанр источника
6	ДАТА	век создания произведения
7	АВТОР	автор источника
8	РЕФ	название источника
9	БИБЛ	библиографическая справка
10	КОДН	код носителя источника
11	НОСИТ	носитель источника
12	НОММ	количество носителей
13	АРХН	архивный номер
14	ЧИСФ	число файлов
15	ТИПФ	тип файла
16	КОД	код файла
17	ЧИСС	число слов
18	ЧИСВ	число стихов
19	ПОСТ	источник поступления
20	КОРР	коррекция
21	РАЗМ	разметка
22	КОПИ	адресат копии
23	ПОЛЗ	пользователи

Пример карты описания источника:

НОМЕР	1
ТИП	2
ХРАН	Текст
КЖАНР	П
ЖАНР	Поэзия
ДАТА	18-19 вв.
РЕФ	Русская баллада
БИБЛ	Русская баллада. М.: Сов. писатель, 1936. 502 с. (Б-ка
БИБЛ	поэта). С. 3-404
КОДН	Д
НОСИТ	5"DS/DD
НОММ	2
АРХН	# 57, 58
ЧИСФ	50
ТИПФ	неарх., послед.
КОД	ASCII (КОИ-7 альтернативный)
ПОСТ	УПЦ 1989-1990

Особенности функционирования Архива в UNILEX-D

Ввод информации в базу данных лингвистических источников осуществляется в системе UNILEX-D с использованием функций Ввод статей и Редактирование. Распечатка текущего состояния Архива производится при помощи перевода содержимого базы данных в оригинал-макет с использованием функции База-макет. В работе [Колодяжная, 1996] содержится такая полная распечатка состояния Архива на 1995 год.

Функционирование базы данных лингвистических источников в системе UNILEX-D позволяет делать всевозможные выборки по заданным критериям: составить списки источников по жанрам, авторам, датам создания и т. п.; составлять количественные таблицы, характеризующие распределение источников по различным параметрам, например, по жанру произведения, как показано ниже (состояние базы на 1995 г.):

Таблица распределения лингвистических источников Архива по жанрам

Учебник	1
Драматургия	3
Публицистика	3
Словарная база данных	3
Разговорная речь	5
Словарь в текстовой форме	8
Поэзия, драматургия	12
Проза	32
Поэзия	117

Русский орфографический словарь (под. ред. В. В. Лопатина)

Общие характеристики

База данных Русского орфографического словаря создавалась на основе различных источников, представленных в компьютерной форме как совокупность текстовых файлов:

1. корпус статей (более 100000) “Орфографического словаря русского языка” [Орфографический словарь русского языка, 1991];
2. корпус новых статей (около 50000), являющихся дополнением к “Орфографическому словарю”;
3. корпус статей, представляющих собой отредактированные и измененные статьи (около 10000) “Орфографического словаря русского языка”.

База данных реализована как система словарных баз, каждая из которых содержит массив статей на одну букву (А–Я), работающих под управлением лексикографической программной системы UNILEX-D. База данных содержит около 160000 статей. В рабочем состоянии система баз данных Словаря занимает около 60 Мб.

Совокупность статей Русского орфографического словаря из базы данных была преобразована в оригинал-макет Словаря (с помощью функции База-макет), послуживший основой для книги “Русский орфографический словарь” [Русский орфографический словарь, 1999].

Структура словарной статьи

При выделении зон для формального описания словарной статьи в системе UNILEX-D учитывались следующие требования: необходимость выделения в отдельную зону каждого фрагмента статьи, могущего быть предметом особого шрифтового выделения при печати; необходимость выделения в отдельную зону каждого фрагмента статьи, содержащего признаки которые могут быть предметом особого исследования. Исходя из этого, была разработана следующая структура статьи Русского орфографического словаря:

1 ЗАГЛ	ключ заглавного
2 СЛОВО	заглавное слово
3 ФОН	фонетика
4 ПАРС	парадигма слова
5 ВИД	сов./несов.
6 ИПАРС	“и” в парадигме слова
7 УКП2	указание на парадигму 2 рода
8 ПАРС2	парадигма 2 рода
9 ПОМП	помета в парадигме
10 НЕСКЛ	несклоняемое
11 РОД	род
12 КАТЕГ	категория (грамматическая)
13 ИКАТ	“и” в КАТЕГ
14 НЕИЗМ	неизменяемое
15 ЕД	указание на форму ед. числа
16 ЕДСЛ	форма ед. числа
17 ПАРЕ	парадигма ед. числа
18 МН	указание на форму мн. числа
19 МНСЛ	форма мн. числа
20 ПАРМ	парадигма мн. числа
21 ИМН	“и” в зоне МН
22 РМН	указание на род. мн.
23 РМНСЛ	форма слова РМН
24 ПРМН	парадигма род. мн.

25 УСЧФ	указание на счетную форму
26 СЧФ	счетная форма
27 ПСЧФ	помета при СЧФ
28 УКРФ	указание на краткую форму
29 КРФСЛ	краткая форма
30 ПКРФ	парадигма краткой формы
31 ПКОТ	помета перед К и ОТ
32 К	указание на производящее
33 КСЛО	производящее при К
34 КИ	связка при К
35 КВАР	вариант производящего при К
36 ОТ	указание на производящее
37 ОТСЛО	производящее при ОТ
38 ОТИ	связка при ОТ
39 ОТВАР	вариант производящего при ОТ
40 ПОМС	помета при СЛОВЕ
41 ТОЛК	толкование
42 ТОЛК2	толкование 2 рода
43 РЕЧЕН	речение
44 ПОМР	помета к РЕЧЕН
45 ФРАЗ	фразеологическое сочетание
46 ТОЛКФ	толкование при ФРАЗ
47 ПОМФ	помета при ФРАЗ
48 РФРАЗ	расширение ФРАЗ
49 ИФРАЗ	“и” при ФРАЗ
50 ВФРАЗ	вариант ФРАЗ
51 ПОМВФ	помета ВФРАЗ
52 ТОЛВФ	толкование ВФРАЗ
53 ИСЛ	связка при СЛОВЕ
54 ППВАР	помета перед ВАР
55 ВАРС	вариант СЛОВА
56 ВАР2	вариант 2 рода
57 КАТЕВ	категория варианта
58 ПАРВ	парадигма варианта
59 ИПАРВ	“и” в ПАРВ
60 УКПВ2	указание на ПАРВ2
61 ПАРВ2	парадигма 2 для ВАРС
62 ЕДВ	указание на ед. число варианта
63 ЕДСЛВ	ед. число после варианта
64 ПАРЕВ	парадигма при ЕДСЛВ
65 УМНВ	указание на мн. число при ВАРС
66 МНВ	мн. форма при ВАРС
67 ПМНВ	парадигма мн. при ВАРС
68 УКРВ	указание на краткую форму ВАРС
69 КРФВ	краткая форма при ВАРС
70 ПКРФВ	парадигма краткой формы ВАРС
71 ТОЛКВ	толкование при ВАРС
72 ТОЛВ2	толкование 2-го рода
73 ПОМВ	помета при ВАРС
74 РЕЧВА	речение к ВАРС
75 ПКОТВ	помета перед К и ОТ
76 КВ	К при ВАРС
77 КВСЛО	слово при КВ
78 ВОТ	ОТ при ВАРС
79 ВОТСЛ	СЛОВО при ВОТ
80 НО	“но” при производящем

81 КАТНО	категория в НО
82 НОСЛ	СЛОВО при НО
83 ТОЛНО	толкование к НО
84 ТОЛН2	толкование 2 рода
85 РЕЧНО	речение после НО

Примеры словарных статей в базе данных:

ЗАГЛ баклуша
СЛОВО бакл/уша,
ПАРС -и,
УКП2 тв.
ПАРС2 -ей
ТОЛК (мелкое
ТОЛК озеро)

ЗАГЛ баклуши
СЛОВО бакл/уши,
ПАРС -/уш,
ЕД ед.
ПАРЕ -/уша,
ПАРЕ -и,
УКП2 тв.
ПАРС2 -ей
ТОЛК (чурки;
РФРАЗ б/ить
РФРАЗ бакл/уши)

ЗАГЛ баклушничать
СЛОВО бакл/ушничать,
ПАРС -аю,
ПАРС -ает

ЗАГЛ баковый
СЛОВО б/аковый

ЗАГЛ баксанский
СЛОВО бакс/анский
ОТ (от
ОТСЛО Бакс/ан)

ЗАГЛ баксанскоеущелье
СЛОВО Бакс/анское
СЛОВО уш/елье

ЗАГЛ баксы
СЛОВО б/аксы,
ПАРС -ов,
ЕД ед.
ЕДСЛ бакс,
ПАРЕ -а

ЗАГЛ бактериальный
СЛОВО бактери/альный

ЗАГЛ бактериемия
СЛОВО бактерием/ия,
ПАРС -и

ЗАГЛ бактеризация
СЛОВО бактериз/ация,
ПАРС -и

ЗАГЛ бактеризованный
СЛОВО бактериз/ованный;
УКРФ кр.
УКРФ ф.
ПКРФ -ан,
ПКРФ -ана

ЗАГЛ бактеризоваться
СЛОВО бактеризов/ать(ся),
ПАРС -з/ую,
ПАРС -з/ует(ся)

ЗАГЛ бактерии
СЛОВО бакт/ерии,
ПАРС -ий,
ЕД ед.
ПАРЕ -/ерия,
ПАРЕ -и

Особенности функционирования словаря в UNILEX-D

При создании базы данных и оригинал-макета Русского орфографического словаря система UNILEX-D применялась для выполнения следующих работ:

1. конвертирование исходного текстового корпуса статей в структурированное представление статей в базе данных;
2. пополнение базы данных;
3. редактирование информации в базе данных;
4. создание оригинал-макета Словаря как основы для издания.

Первый и второй виды работ выполнялись с использованием функции Текст-база2 системы UNILEX-D, являющейся модификацией основной функции Текст-база (см. раздел **Описание системы UNILEX-D**) и написанной специально для Орфографического словаря. Ниже приведен пример фрагмента исходного текстового файла:

бакл/уша, -и, тв. -ей (мелкое озеро)
бакл/уши, -/уш, ед. -/уша, -и, тв. -ей (чурки; б/ить бакл/уши)
бакл/ушничать, -аю, -ает
б/аковый
бакс/анский (от Бакс/ан)
Бакс/анское ущ/елье
б/аксы, -ов, ед. бакс, -а
бактери/альный
бактерием/ия, -и
бактериз/ация, -и
бактериз/ованный; кр. ф. -ан, -ана
бактеризов/ать(ся), -з/ую, -з/ует(ся)
бакт/ерии, -ий, ед. -/ерия, -и

Третий вид работы — редактирование базы данных — выполнялся с использованием операций Редактирование1, Редактирование2 и Индексация. Четвертый вид — создание оригинал-макета — с использованием функции База-макет. Ниже подробно описано создание оригинал-макета в системе UNILEX-D.

Создание оригинал-макета Русского орфографического словаря

Функция База-макет системы UNILEX-D позволяет создать такую текстовую копию базы данных компьютерного словаря, в которой слева от значения каждой элементарной компоненты статьи указан набор символов, определяющий тип шрифта, которым будет набрана эта компонента в оригинал-макете. После выбора словаря на экране компьютера высвечивается описание структуры статьи данного словаря. В нижнем окне экрана высвечивается таблица, задающая тип шрифта. В таблице заданы такие параметры шрифта, как: высота, жирность выделения буквы, наклон буквы (набор символов, задающий определенный тип шрифта заимствован из работы [В. М. Андрюченко, 1992] (“Об организации архива источников...”). Справа от каждой макрокоманды таблицы дана ее расшифровка. Ниже приведена эта таблица.

Код шрифта	Название шрифта
\sN	< 8п прямой
\sNi	< 8п курсив
\sNb	< 8п полужирный
\N	8п прямой
\Ni	8п курсив
\Nb	8п полужирный
\sT	9п прямой
\sTi	9п курсив
\sTb	9п полужирный
\T	11п прямой

\Ti	11п курсив
\Tb	11п полужирный
\sP	12п прямой
\sPi	12п курсив
\sPb	12п полужирный
\P	14п прямой
\Pi	14п курсив
\Pb	14п полужирный

Филолог последовательно обрабатывает список зон словарной статьи, выбирая из таблицы нужный шрифт и приписывая его очередной зоне. Ниже приведено описание структуры словарной статьи Русского орфографического словаря с приписанными каждой зоне типами шрифтов:

СЛОВО	\Tb
ФОН	\Nb
ПАРС	\Nb
ВИД	\Ni
ИПАРС	\N
УКП2	\Ni
ПАРС2	\N
ПОМП	\Ni
НЕСКЛ	\Ni
РОД	\Ni
КАТЕГ	\Ni
ИКАТ	\N
НЕИЗМ	\Ni
ЕД	\Ni
ЕДСЛ	\Nb
ПАРЕ	\N
МН	\Ni
МНСЛ	\Nb
ПАРМ	\N
ИМН	\N
РМН	\Ni
РМНСЛ	\N
ПРМН	\N
УСЧФ	\Ni
СЧФ	\N
ПСЧФ	\N
УКРФ	\Ni
КРФСЛ	\N
ПКРФ	\N
ПКОТ	\Ni
К	\Ni
КСЛО	\sTb
КИ	\N
КВАР	\sTb
ОТ	\Ni
ОТСЛО	\sTb
ОТИ	\N
ОТВАР	\sTb
ТОЛК	\sTi
ТОЛК2	\sT
ПОМС	\sTi
РЕЧЕН	\sTi
ПОМР	\sTi
ФРАЗ	\Tb
ТОЛКФ	\sTi
ПОМФ	\sTi
РФРАЗ	\sT
ИФРАЗ	\N
ВФРАЗ	\Tb
ПОМВФ	\sTi
ТОЛВФ	\sTi
ИСЛ	\N
ППВАР	\Ni
ВАРС	\Tb
ВАР2	\Tb
ПАРВ	\Nb
ИПАРВ	\N

УКПВ2	\Ni
ПАРВ2	\N
КАТЕВ	\Ni
ЕДВ	\Ni
ЕДСЛВ	\Nb
ПАРЕВ	\N
УМНВ	\Ni
МНВ	\Nb
ПМНВ	\N
УКРВ	\Ni
КРФВ	\N
ПКРФВ	\N
ТОЛКВ	\sTi
ТОЛВ2	\sT
ПОМВ	\sTi
РЕЧВА	\sTi
ПКОТВ	\Ni
КВ	\Ni
КВСЛО	\sTb
ВОТ	\Ni
ВОТСЛ	\sTb
НО	\N
КАТНО	\Ni
НОСЛ	\sTb
ТОЛНО	\sTi
ТОЛН2	\sT
РЕЧНО	\sTi

После обработки функцией **База-макет** словарного файла в базе данных как результат создается макет-файл, в котором каждой компоненте статьи приспан выбранный филологом тип шрифта. Конвертирование макет-файла в оригинал-макет словаря происходит посредством специальной макропрограммы на языке WORDBASIC под управлением операционной системы WINDOWS'95. Ниже приведены фрагменты макет-файла и оригинал-макета Русского орфографического словаря:

```
\Gb бакл/уша,
\nb -и
\sTi (мелкое
\sTi озеро)\
\Г баклуши
\Gb бакл/уши,
\nb -/уш,
\Ni ед.
\nb бакл/уша,
\N -и,
\Ni тв.
\N -ей
\sTi (чурки;
\sT б/ить
\sT бакл/уши)\
```

баклуша, -и, тв. -ей (мелкое озеро)
 баклуши, -уш, ед. -уша, -и, тв. -ей (чурки; бить баклуши)
 баклушничать, -аю, -ает
 баковый
 баксанский (от Баксан)
 Баксанское ущелье
 баксы, -ов, ед. бакс, -а
 бактериальный
 бактериемия, -и
 бактеризация, -и
 бактеризованный; кр. ф. -ан, -ана
 бактеризовать(ся), -зую, -зует(ся)
 бактерии, -ий, ед. -ерия, -и

Словарь языка русской поэзии XX века (под. ред. В. П. Григорьева)

Общие характеристики

Автором идеи и руководителем проекта "Словарь языка русской поэзии XX века" является д. филол. наук, проф. В. П. Григорьев (см. [Григорьев, 1973], [Григорьев, 1994]). Работа по

проекту ведется совместно двумя отделами Института русского языка (отделом Стилистики и языка литературы и отделом Машинного фонда русского языка) с 1995 года. Финансирование осуществляется по грантам РГНФ (№ 95-06-17344 и № 98-04-06054).

Компьютерная база данных Словаря создается на основе книжных источников по стихам 10-ти русских поэтов [Анненский, 1990; Ахматова, 1986, 1995; Блок, 1960; Есенин, 1961; Кузмин, 1990; Мандельштам, 1990; Маяковский, 1969; Пастернак, 1989; Хлебников, 1986; Цветаева, 1994] с использованием программных систем UNILEX-T [Аношкина, 1995] и UNILEX-D, созданными в отделе Машинного фонда русского языка.

База данных Словаря реализована как система словарных баз, каждая из которых содержит массив статей на одну букву (А–Я), работающих под управлением лексикографической программной системы UNILEX-D. К настоящему времени созданы базы данных Словаря для букв А–Н. Общее количество словарных статей в базе данных (буквы А–Н) составляет: 33728.

В 1998 году по материалам Словаря опубликована книга “Самовитое слово. Словарь русской поэзии XX века. Пробный выпуск: А – А-ю-рей”. Версия “Пробного выпуска” для сети Интернет представлена на сайте www.slovo.rema.ru

В 1999 году подготовлена к печати книга “Словарь языка русской поэзии XX века. Том I. А–В”, содержащая более 7400 статей.

В 2000 году подготовлен к печати II том “Словаря языка русской поэзии XX века” (буквы Г–Ж), содержащий 4180 словарных статей.

Структура словарной статьи

В структуре словарной статьи Словаря выделяется пять зон: ЗАГОЛОВОЧНОЕ СЛОВО (ЗАГЛАВНАЯ ФОРМА СЛОВА), ЗОНА ЗНАЧЕНИЯ, ЗОНА КОНТЕКСТОВ, ЗОНА КОММЕНТАРИЕВ и ЗОНА ШИФРОВ.

Зона значения является факультативной и следует сразу за заголовочным словом. Информация, содержащаяся в этой зоне, дается в квадратных скобках (кроме ссылок на другие статьи) прямым светлым шрифтом со строчной начальной буквы и предусматривает:

- а) сведения лингвистического характера;
- б) сведения энциклопедического и иного характера;
- в) ссылочную информацию.

Зона контекстов является основной и отсутствует лишь в ссылочных статьях. Ее составляют один или несколько контекстов, разъясняющий комментарий к контексту (факультативно); по существу к ней же относится зона шифров. Контексты внутри одной статьи располагаются в хронологическом порядке (дат написания произведений, с точностью до года), а внутри одной даты — по алфавиту авторов.

Зона комментариев является факультативной. Комментарий располагается после контекста, дается в квадратных скобках прямым светлым шрифтом со строчной начальной буквы. В отличие от информации в зоне значения (относящейся ко всем контекстам данного слова) комментарий относится только к конкретному единичному контексту, но также должен способствовать более глубокому раскрытию особенностей словоупотребления.

Зона шифров является обязательной и сопровождает каждый контекст. В этой зоне указываются автор и дата создания произведения, а также дается ссылка на страницу источника. Для каждого из 10-ти авторов вводятся краткие обозначения: Анн — Анненский, Ахм — Ахматова, АБ — Блок, Ес — Есенин, Куз — Кузмин, ОМ — Мандельштам, М — Маяковский, П — Пастернак, Хл — Хлебников, Цв — Цветаева. Для указания даты обычно используются три последние цифры года; дата печатается сразу, без пробела за шифром автора, курсивом: АБ898, Анн900, Ахм963. Иногда может указываться интервал между датами (или предположительный период) создания стихотворения: П913,28, АБ908–10, Анн900-е. Предположительная дата создания произведения заключается в квадратные скобки: Цв[911].

Формальное описание структуры статьи:

- 1 ЗАГЛ заголовочное слово,
- 2 ЗНАЧ зона значений,
- 3 КОНТ зона контекстов,
- 4 КОММ зона комментариев,
- 5 АВТОР зона шифров

Примеры словарных статей в базе данных:

ЗАГЛ ЛАБИРИНТ
КОНТ Помнишь думы? Они улетели.
КОНТ Отцвели завитки гиацинта.
КОНТ Мы провидели светлые цели

КОНТ В отдаленных краях лабиринта.
 АВТОР АБ904 (П,57)
 КОНТ Стихийный л., непостижимый лес,
 КОНТ Души готической рассудочная пропасть,
 КОНТ Египетская мощь и христианства робость,
 КОНТ С тростинкой рядом — дуб, и всюду царь — отвес.
 КОММ [о соборе Парижской Богоматери]
 АВТОР ОМ912 (83.2)
 КОНТ Когда за лиры л.
 КОНТ Поэты взор вперят,
 КОНТ Налево развернется Инд,
 КОНТ Правей пойдет Евфрат.
 АВТОР П913,28 (I,50)
 КОНТ Когда за лиры л.
 КОНТ Поэты взор вперят,
 КОНТ Налево глины слижет Инд,
 КОНТ А вправо уйдет Евфрат.
 КОММ [стих. является вариантом предыдущ.]
 АВТОР П913 (I,427.1)
 КОНТ Шагни, и еще раз, — твердил мне инстинкт
 КОНТ И вел меня мудро, как старый схоластик,
 КОНТ Через путанный, древний, сырой л.
 КОНТ Нагретых деревьев, сирени и страсти.
 АВТОР П916 (I,491.2)
 КОНТ "Ах, это вы? Зажмурьтесь и застыньте", —
 КОНТ Услышал он [Спекторский] в тот первый раз и миг,
 КОНТ Когда, сторонний в этом лабиринте,
 КОНТ Он сосвежу и точно стал в тупик.
 АВТОР П925-31 (I,358)
 КОНТ Как вдруг Спекторский обомлел и ахнул.
 КОНТ В глазах, уставших от чужих перин,
 КОНТ Блеснуло что-то яркое, как яхонт,
 КОНТ Он увидел Марии л..
 АВТОР П925-31 (I,369)
 КОНТ И в лабиринте влажного распева
 КОНТ Такая душная стрекошет мгла,
 КОНТ Как будто в гости водяная дева
 КОНТ К часовщику подземному пришла.
 АВТОР ОМ930 (165.1)

ЗАГЛ НИЖНИЙ

ЗНАЧ [прил. к низ]
 КОНТ Рухнула штукатурка в нижнем этаже.@
 КОНТ Нервы — @
 КОНТ большие,@
 КОНТ маленькие,@
 КОНТ многие! — @
 КОНТ скачут бешеные,
 АВТОР М914-15 (388)
 КОНТ В нижней церкви служили молебны
 КОНТ О моряках, уходящих в море,
 АВТОР Ахм914 (264)
 КОНТ Блаженна стрекоза, разбитая грозой,
 КОНТ Когда она прячется на нижней стороне
 КОНТ Древесного листа.
 АВТОР Хл915-19-22 (463)
 КОНТ Сердцу срединному солнечный хмель.
 КОНТ Произрастание — верхнему севу!
 КОНТ Воспоминание — нижним водам!
 АВТОР Куз922 (273)
 КОНТ И военной грозой потемнел
 КОНТ Н. слой помраченных небес -
 АВТОР ОМ922 (144)
 КОНТ Корсетная коробка@
 КОНТ этажей под шестьдесят. @<...>@
 КОНТ А в самом нижнем — @
 КОНТ "Дрогс сода,@
 КОНТ грет энд фэймус компани-нейшенал".
 АВТОР М925 (212)

ЗАГЛ НИЖНИЙ
 ЗНАЧ [то же, что Нижний Новгород]
 КОНТ ЮНКЕРАМ, УБИТЫМ В НИЖНЕМ _Загл.
 АВТОР Цв917 (1,363.2)
 КОНТ Ярмарки... там
 КОНТ В Нижнем, контракты, другие...
 КОНТ Пароходства... Волга!
 КОНТ Подумайте, Волга!
 АВТОР Куз922 (240)
 КОНТ Потом перрон и град шагов и фраз,
 КОНТ И чей-то крик: "Так, значит завтра в Нижнем?"
 КОНТ И у окна: "Итак, в последний раз,
 КОНТ Ступай. Мы больше ничего не выжжем". _РП
 АВТОР П925-31 (1,350)
 КОНТ Лед за пристанью за ближней,@
 КОНТ оковала Волга рот,@
 КОНТ это красный,@
 КОНТ это Н.,@
 КОНТ это зимний Новгород.
 АВТОР М927 (284)

Особенности функционирования словаря в UNILEX-D

База данных Словаря создается в два этапа. На первом этапе, на основе конкордансов, полученных в системе UNILEX-T (подробное описание системы UNILEX-T см. в работе [Аношкина, 1995]), создаются базы данных для каждого автора-поэта. Первичные базы данных содержат для каждого слова совокупности всех его контекстов. Первичные базы данных "сливаются" в системе UNILEX-D в Сводную базу данных и пересортировываются по алфавиту слов, внутри каждого слова — по дате создания контекста, внутри данной даты — по авторам контекстов. Ниже приводится фрагмент записей в Сводной базе данных.

Фрагмент совокупности исходных контекстов в Сводной базе данных для слова "ЛАЗУРНЫЙ":

ЗАГЛ ЛАЗУРНЫЙ 898БЛО~1,335.1#004-002
 КОНТ И самый мир сотрется в прах
 КОНТ Под тихим ужасом лазури...
 КОНТ Помедли, ночь! Небесный луч!
 КОНТ Не озаряй тюрьмы лазурной!
 КОНТ Пускай мешают нам сквозь туч
 КОНТ Лишь звезды — очи ночи бурной!
 АВТОР АБ898 (1,335.1)
 ЗАГЛ ЛАЗУРНЫЙ 898БЛО~1,379#001-009
 КОНТ Он тщетно силится прильнуть к устам
 КОНТ На поцелуй лобзанием отвечать,
 КОНТ Но вверх летит и в воздухе лазурном
 КОНТ Уста твои не может целовать...
 КОНТ И ты, коварная, надменной, строгой лаской
 КОНТ Закралась в душу мне и там зажгла огни.
 АВТОР АБ898 (1,379)
 ЗАГЛ ЛАЗУРНЫЙ 898БЛО~1,391.2#003-003
 КОНТ Щедрот Твоих и благи Твоей?
 КОНТ Надежды нет: вокруг и ветер бурный,
 КОНТ И ночь, и гребни волн, и дым небесных туч
 КОНТ Разгонят всё, и образ Твой лазурный
 КОНТ Затмят, как всё, как яркий солнца луч...
 КОНТ Но, если туча с молнией и громом
 КОНТ Пройдет, закрыв Тебя от взора моего,
 АВТОР АБ898 (1,391.2)
 ЗАГЛ ЛАЗУРНЫЙ 900-еАНН~107.1#002-001
 КОНТ А теперь влюбилось
 КОНТ В бездонность бирюзы,
 КОНТ В ее глаза усталые.
 КОНТ Всё, что есть лазурного,
 КОНТ Излилось в лучах
 КОНТ На зыби златошвейные,
 КОНТ Всё, что там безбурного
 АВТОР Анн900-е (107.1)
 ЗАГЛ ЛАЗУРНЫЙ 900-еАНН~119.1#001-004
 КОНТ Как чисто гаснут небеса,
 КОНТ Какою прихотью ажурной

КОНТ Уходят дальние леса
 КОНТ В ту высь, что знали мы лазурной...
 КОНТ В твоих глазах упрека нет:
 КОНТ Ты туч закатных догоранье
 КОНТ И сизо-розовый отсвет
 АВТОР Анн900-е (119.1)
 ЗАГЛ ЛАЗУРНЫЙ 900-еАНН~179.2#001-004
 КОНТ Что ни день, теплей и краше
 КОНТ Осенен простор эфирный
 КОНТ Осушенной солнцем чашей:
 КОНТ То лазурной, то сафирной.
 КОНТ Синью нежною, как пламя,
 КОНТ Горды солнцевы палаты,
 КОНТ И ревниво ключья ваты
 АВТОР Анн900-е (179.2)
 ЗАГЛ ЛАЗУРНЫЙ 900-еАНН~183.3#007-003
 КОНТ Дымится черный виксатин.
 КОНТ Когда бы бури пролетали
 КОНТ И все так быстро и светло...
 КОНТ Но не умчит к лазурной дали
 КОНТ Грозой разбитое крыло.
 АВТОР Анн900-е (183.3)
 ЗАГЛ ЛАЗУРНЫЙ 900АНН~163.2#001-023
 КОНТ Но не навеки,
 КОНТ Если б заснуть
 КОНТ Так, чтобы после проснуться,
 КОНТ Только под небом лазурным...
 КОНТ Новым, счастливым, любимым...
 АВТОР Анн900 (163.2)
 ЗАГЛ ЛАЗУРНЫЙ 900БЛО~I,038#001-001
 КОНТ Шли мы стезею лазурною,
 КОНТ Только расстались давно...
 КОНТ В ночь непроглядную, бурную
 КОНТ Вдруг распахнулось окно...
 АВТОР АБ900 (I,38)
 ЗАГЛ ЛАЗУРНЫЙ 900БЛО~I,038#001-013
 КОНТ Слезы мешают с дождем...
 КОНТ Хочешь обнять на прощание?
 КОНТ Прошлое вспомнить вдвоем?
 КОНТ Мимо, виденье лазурное!
 КОНТ Сердце сжимает тоской
 КОНТ В ночь непроглядную, бурную
 КОНТ Ветер, да образ быллой!
 АВТОР АБ900 (I,38)

Совокупности контекстов Сводной базы данных распределяются среди филологов-составителей итоговых словарных статей. После получения итоговых словарных статей они загружаются в итоговую базу данных, служащую основой для создания оригинал-макета словаря и его лексикографического исследования. Ниже приведен фрагмент статьи в итоговой базе данных для слова ЛАЗУРНЫЙ:

ЗАГЛ ЛАЗУРНЫЙ
 ЗНАЧ [_прил.; л. и Л.]
 КОНТ Ослепнем в царственных лучах,
 КОНТ Мы, знавшие лишь ночь да бури,
 КОНТ И самый мир сотрется в прах
 КОНТ Под тихим ужасом лазури... @@
 КОНТ Помедли, ночь! Небесный луч!
 КОНТ Не озаряй тюрьмы лазурной!
 КОНТ Пускай мешают нам сквозь туч
 КОНТ Лишь звезды — очи ночи бурной!
 АВТОР АБ898 (I,335.1)
 КОНТ в воздухе лазурном
 АВТОР АБ898 (I,379)
 КОНТ Надежды нет: вокруг и ветер бурный,
 КОНТ И ночь, и гребни волн, и дым небесных туч
 КОНТ Разгонят всё, и образ Твой л.
 КОНТ Затмят, как всё, как яркий солнца луч...
 АВТОР АБ898 (I,391.2)

КОНТ Всё, что есть лазурного,
 КОНТ Излилось в лучах
 КОНТ На зыби златошвейные,
 КОММ [_рфм.: безбурного]
 АВТОР Анн900-е (107.1)
 КОНТ Как чисто гаснут небеса,
 КОНТ Какою прихотью ажурной
 КОНТ Уходят дальние леса
 КОНТ В ту высь, что знали мы лазурной...
 АВТОР Анн900-е (119.1)
 КОНТ Что ни день, теплей и краше
 КОНТ Осенен простор эфирный
 КОНТ Осушенной солнцем чашей:
 КОНТ То лазурной, то сафирной.
 АВТОР Анн900-е (179.2)
 КОНТ Когда бы бури пролетали
 КОНТ И все так быстро и светло...
 КОНТ Но не умчит к лазурной дали
 КОНТ Грозой разбитое крыло.
 АВТОР Анн900-е (183.3)
 КОНТ Если б заснуть
 КОНТ Так, чтобы после проснуться,
 КОНТ Только под небом лазурным...
 КОНТ Новым, счастливым, любимым...
 АВТОР Анн900 (163.2)
 КОНТ Шли мы стезею лазурною,
 КОНТ Только расстались давно...
 КОНТ В ночь непроглядную, бурную
 КОНТ Вдруг распахнулось окно...
 АВТОР АБ900 (1,38)
 КОНТ Хочешь обнять на прощание?
 КОНТ Прошлое вспомнить вдвоем?
 КОНТ Мимо, виденье лазурное!
 КОНТ Сердце сжимает тоской
 КОНТ В ночь непроглядную, бурную
 КОНТ Ветер, да образ былой! ib.
 АВТОР АБ900 (1,38)

Создание оригинал-макета Словаря языка русской поэзии

Для создания оригинал-макета Словаря использовалась функция База-макет системы UNILEX-D. Была составлена таблица соответствия зон словарной статьи и шрифтов, получен макет-файл и затем оригинал-макет. Ниже приведена таблица соответствия и фрагмент оригинал-макета Словаря.

Таблица соответствия зон и шрифтов:

ЗАГЛ	\sTb
ЗНАЧ	\N
КОНТ	\sT
АВТОР	\sTi
КОММ	\N

Фрагмент статьи ЛАЗУРНЫЙ в оригинал-макете Словаря:

ЛАЗУРНЫЙ [прил.; л. и Л.] Ослепнем в царственных лучах, Мы, знавшие лишь ночь да бури, И самый мир сотрется в прах Под тихим ужасом лазури... // Помедли, ночь! Небесный луч! Не озаряй тюрьмы лазурной! Пускай мешают нам сквозь туч Лишь звезды – очи ночи бурной! АБ898 (1,335.1); в воздухе лазурном АБ898 (1,379); Надежды нет: вокруг и ветер бурный, И ночь, и гребни волн, и дым небесных туч Разгонят всё, и образ Твой л. Затмят, как всё, как яркий солнца луч... АБ898 (1,391.2); Всё, что есть лазурного, Излилось в лучах На зыби златошвейные, [рфм.: безбурного] Анн900-е (107.1); Как чисто гаснут небеса, Какою прихотью ажурной Уходят дальние леса В ту высь, что знали мы лазурной... Анн900-е (119.1); Что ни день, теплей и краше Осенен простор эфирный Осушенной солнцем чашей: То лазурной, то сафирной. Анн900-е (179.2); Когда бы бури пролетали И все так быстро и светло... Но не умчит к лазурной дали Грозой разбитое крыло. Анн900-е (183.3); Если б заснуть Так, чтобы после проснуться, Только под небом лазурным... Новым, счастливым, любимым... Анн900 (163.2); Шли мы стезею лазурною, Только расстались давно... В ночь непроглядную, бурную Вдруг распахнулось окно... АБ900 (1,38); Хочешь обнять на прощание? Прошлое вспомнить вдвоем? Мимо, виденье лазурное! Сердце сжимает тоской В ночь непроглядную, бурную Ветер, да образ былой! ib.; <...>

Научные исследования и публикации

По материалам Словаря языка русской поэзии было опубликовано 5 научных статей [Колодяжная, Шестакова, 1998; Григорьев, Колодяжная, Шестакова, 1999]. Кроме того, был подготовлен к публикации словарь “Религиозная лексика в стихах русских поэтов Серебряного века”, являющийся подсловарем основного Словаря [“Религиозная лексика...”, 2000].

Литература

1. **Андрющенко В. М.** *Автоматизированная лексикографическая система UNILEX (Основные проектные решения)* // Вычислительная лингвистика. Теоретические вопросы. Вопросы автоматизации лексикографических работ. М.: Изд-во Моск. ун-та. 1982. С. 104–119.
2. **Андрющенко В. М.** *Концепция и архитектура Машинного фонда русского языка* // Машинный фонд русского языка: идеи и суждения. М.: Наука. 1986. С. 26–44.
3. **Андрющенко В. М.** *Некоторые конструктивные черты Машинного фонда русского языка* // Доклады Второй Всесоюзной конференции по созданию Машинного фонда русского языка. М., 1987. С. 5–27.
4. **Андрющенко В. М.** *Машинный фонд русского языка. Интеграционный подход.* М., 1989. 77 с.
5. **Андрющенко В. М.** *Концепция и архитектура Машинного фонда русского языка.* М.: Наука, 1989. 197 с.
6. **Андрющенко В. М.** *Об организации источников Машинного фонда русского языка, их разметке и комментировании* // Бюллетень Машинного фонда русского языка. Вып. 1. Москва, 1992. С. 3–44.
7. **Аношкина Ж. Г.** *Подготовка частотных словарей и конкордансов на компьютере.* М.: Русские словари, 1995. 60 с.
8. **Григорьев В. П.** *Введение* // Поэт и слово. Опыт словаря. М.: Наука, 1973. С. 13–150.
9. **Григорьев В. П.** *Самовитое слово и его словарное представление* // Изв. РАН. Серия лит-ры и языка. Т. 53, № 4. 1994.
10. **Григорьев В. П.** *Анненский, Блок, Хлебников, Мандельштам: слово “ветер”* // Русистика сегодня, № 3, 1994. С. 57–65.
11. **Григорьев В. П., Колодяжная Л. И., Шестакова Л. Л.** *Словарь русской поэзии XX века “Самовитое слово”. Трудные случаи при заполнении зоны значения словарной статьи.* // Альманах “Говор”. Материалы II-ой Всероссийской конференции “Русский язык: прошлое, настоящее, будущее”. Ч. II. Сыктывкар, 1999. С. 18–22.
12. **Григорьев В. П., Колодяжная Л. И., Шестакова Л. Л.** *Статья Пушкин в словаре “Самовитое слово”* // А. С. Пушкин и русский литературный язык в XIX–XX вв. Тезисы докладов научной конференции. Нижний Новгород, 1999. С. 97–100.
13. **Григорьев В. П., Колодяжная Л. И., Шестакова Л. Л.** *Имя “Пушкин” в стихах русских поэтов XX века (Опыт словарной статьи)* // А. С. Пушкин. К 200-летию со дня рождения. Российская академия наук. Институт научной информации по общественным наукам. М., 1999. С. 33–43.
14. **Григорьев В. П., Колодяжная Л. И., Шестакова Л. Л.** *Имя собственное ПУШКИН и его производные в Словаре русской поэзии XX века “Самовитое слово”* // Сборник, посвященный 200-летию А. С. Пушкина. М.: Наука, 1999.
15. **Загоровская О. В., Лесников С. В., Маркова З. М., Сергиева Н. С.** *Автоматизированный Словарь Русских Говоров Республики Коми и сопредельных областей* // Альманах “Говор”. Часть 3. Саратов, Сыктывкар, 1996. С. 58–110.
16. **Золотова Г. А., Колодяжная Л. И., Шумарина И. В.** *Использование информации Синтаксического словаря русского языка в филологических исследованиях* // Труды Машинного фонда русского языка. Том I. М., 1991. С. 87–95.
17. **Колодяжная Л. И.** *Структура словарного текста в аспекте машинной лексикографии.* Автореф. дисс. ... к. филол. н. М., 1986.
18. **Колодяжная Л. И., Лавошникова Э. К., Мошкович Ж. Г., Салий А. Д.** *Руководство по использованию программного обеспечения многоязычного автоматического словаря “ТЕРМИН”.* Методическая разработка. М.: НИВЦ МГУ, 1986. 75 с.
19. **Колодяжная Л. И.** *Автоматическое выделение фрагментов размеченной словарной статьи в филологических словарях* // Вопросы кибернетики. Прикладные аспекты лингвистической теории. М., 1987. С. 117–130.
20. **Колодяжная Л. И.** *Автоматизированная лексикографическая система UNILEX.* М.: Изд-во Моск. ун-та, 1987. 115 с.
21. **Колодяжная Л. И.** *Опыт создания машинной версии фрагмента словаря русского языка* // Доклады Второй Всесоюзной конференции по созданию Машинного фонда русского языка. М. 1987. С. 163–175.
22. **Колодяжная Л. И.** *Автоматический Генеральный словник русских словарей* // Машинный фонд русского языка: предпроектные исследования. М., 1988. С. 41–64.

23. Колодяжная Л. И. Принципы создания филологического словаря на персональном компьютере // Материалы III Всесоюзной конференции по созданию Машинного фонда русского языка. М.: изд-во Моск. ун-та, 1990. С. 17–29.
24. Kolodjaznaja L., Polikarpov A. A system for compilation and analysis of computerized philological dictionaries using a personal computer // Terminology and knowledge engineering. Vol. 2. Index Verlag, Frankfurt/M., 1990. P. 473–479.
25. Колодяжная Л. И. Система “Компьютерный филологический словарь” (UNILEX-D) // Бюллетень Машинного фонда русского языка. Вып. 2. М., 1992. С. 7–14.
26. Колодяжная Л. И. Компьютерный словарь управления в русском языке. Описание // Бюллетень Машинного фонда русского языка. Вып. 2. М., 1992. С. 14–18.
27. Колодяжная Л. И. Компьютерный синтаксический словарь. Описание // Бюллетень Машинного фонда русского языка. Вып. 2. М., 1992. С. 19–21.
28. Колодяжная Л. И., Марон М. Е. Интегрированная обработка филологических баз данных // Бюллетень Машинного фонда русского языка. Вып. 2. М., 1992. С. 31–42.
29. Колодяжная Л. И., Пшеничнова М. Г. Компьютерная версия толкового словаря русского языка. Описание // Бюллетень Машинного фонда русского языка. Вып. 2. М., 1992. С. 26–31.
30. Колодяжная Л. И., Пшеничнова М. Г. Метасловарь количественных параметров Словаря русского языка С. И. Ожегова // Инженерная лингвистика и оптимизация преподавания языков. Республиканская конференция. Тез. докл. Самарканд, 1992. С. 140.
31. Kolodjaznaja L., Polikarpov A. Study of quantitative correlations between stylistics, grammar and polysemy of words // Qualico-94. Moscow, 1994. P. 110–112.
32. Колодяжная Л. И. Пространства преобразований филологических словарей и их создание в процессе UNILEX-D // Лингвистика на исходе XX века: итоги и перспективы. Тезисы межд. конференции. Том 1. М.: Филол. фак-т МГУ, 1995. С. 240–241.
33. Колодяжная Л. И. Компьютерный словарь русских синонимов // Альманах “Говор”. Сыктывкар, май 1995. С. 57–60.
34. Колодяжная Л. И. База данных описаний лингвистических источников в архиве Машинного фонда русского языка // Альманах “Говор”. Сыктывкар, июль 1995. С. 9–29.
35. Колодяжная Л. И. Архив лингвистических источников Машинного фонда русского языка // Бюллетень Машинного фонда русского языка. Вып. 3. М., 1996. С. 16–53.
36. Колодяжная Л. И. Понятие лексикографического пространства // Альманах “Говор”. Часть 3. Саратов, Сыктывкар, 1996. С. 41–47.
37. Колодяжная Л. И., Поликарпов А. А. Системно-количественное исследование русской синонимии // Альманах “Говор”. Часть 1. Сыктывкар, 1997.
38. Колодяжная Л. И., Шестакова Л. Л. Словарь русской поэзии XX века “Самовитое слово”: зоны словарной статьи и их типология // Современные проблемы лексикографии. Труды Межд. конф. Гродно, 1998.
39. Колодяжная Л. И. Частотный словарь схем управления с относящимися к ним словами (на материале словаря управления Д. Э. Розенталя) // Интернет-сайт филологического факультета МГУ: www.humlang.newmail.ru/ (раздел: Наши публикации). М., 2000. 17 с.
40. Лесников С. В. Конструирование гипертекстового словаря русского говора с. Лойма // Альманах “Говор”. Часть 1. Саратов, Сыктывкар, 1996. С. 57–91.
41. Машинный фонд русского языка: идеи и суждения. М.: Наука, 1986. 240 с.
42. Пшеничнова М. Г. Лексикографические параметры и способы их формализации (на материале Словаря В. И. Даля). Автореф. дисс. ... к. филол. н. М., 1990.
43. Религиозная лексика в стихах русских поэтов Серебряного века. По материалам Словаря русской поэзии XX века. Под ред. В. П. Григорьева. // Интернет-сайт Web-центра “Омега”: www.wco.ru/biblio/ (раздел: Справочная литература). М., 2000. 100 с.; также интернет-сайт “Мир Марины Цветаевой”: www.ipmce.su/~tsvet/ (раздел “Словари”).
44. Semelova V., Fluhr. Les Industries de la Langue dans les Pays de l'ex-URSS: repertoire des acteurs et des produits. Vol. 2. Paris, 1997. С. 52–68.

Источники

1. Анненский И. Ф. Стихотворения и трагедии. Л.: Советский писатель, 1990. С. 51–286.
2. Ахматова А. А. Сочинения в двух томах. Том 1. М.: Художественная литература, 1986.
3. Ахматова А. А. Реквием // “В то время я гостила на земле...”. Стихотворения и поэмы. СПб.: Лениздат, 1995. С. 349–357.
4. Блок А. А. Собрание сочинений в восьми томах. Тома 1–3. М.–Л.: Художественная литература, 1960.
5. Есенин С. А. Собрание сочинений в пяти томах. Тома 1–3. М.: Художественная литература, 1961.
6. Кузмин М. А. Избранные произведения. Л.: Художественная литература, 1990. С. 17–328.
7. Мандельштам О. Э. Сочинения в двух томах. Том 1. М.: Художественная литература, 1990. С. 65–440.
8. Маяковский В. В. Стихотворения. Поэмы. Пьесы. М.: Художественная литература, 1969. С. 25–606.

9. *Орфографический словарь русского языка*. Под ред. В. В. Лопатина. 29-е издание. М.: Русский язык, 1991. 415 с.
10. **Пастернак Б. Л.** *Сочинения в четырех томах. Тома 1–2*. М.: Художественная литература, 1989.
11. **Пастернак Б. Л.** *Сочинения в четырех томах. Том 3*. М.: Художественная литература, 1989. С. 511–538.
12. *Русский орфографический словарь*. Под ред. В. В. Лопатина. М.: Азбуковник, 1999. 1260 с.
13. **Хлебников Велимир.** *Творения*. М.: Советский писатель, 1986. С. 30–504.
14. **Цветаева М. И.** *Собрание сочинений в семи томах. Тома 1–3*. М.: Эллис Лак, 1994.

V. Текстообработка и словарная система UNILEX-A

Ниже приведена распечатка стандартного Help-файла (C++ Help System). Подстраничные примечания — это гипертекстовые ссылки.

SKA+ Unilex A — это система подготовки частотных словарей и конкордансов для корпусов текстов, составленных из произвольной коллекции текстов. Результаты своей работы система сохраняет в словарной базе данных, в которой наряду с ними могут храниться также и обычные словарные данные, например, из “Словаря русского языка” С. И. Ожегова.

SKA+ Unilex A — система, управляемая посредством меню. Главное меню системы состоит из следующих пунктов:

<u>Objects</u>	<u>Texts</u>	
	<u>Dictionaries</u>	
	<u>Grammatical Tables</u>	Load
		Edit
<u>Tasks</u>	<u>Processing</u>	<u>Text Corpus</u>
		<u>Dictionary DB</u>
	<u>Import</u>	
	<u>Export</u>	
	<u>View/Edit</u>	
Help	About	
	Contents	
Exit	Quit	

#SAK+ Имеется три типа объектов: *тексты*, *словари* и *грамматические таблицы*. При выборе объекта “**Тексты**” пользователю предоставляется окно, в котором он может сформировать корпус из текстов, хранящихся в директории *Unilex_A\Texts*. При выборе объекта “**Словари**” пользователю предоставляется такое же окно, в котором можно выбрать нужные для работы словари из директории *Unilex_A\Dics*. Если они еще не были загружены в базу данных, они загружаются. При выборе объекта “**Грамматические таблицы**” пользователю предоставляется окно, в котором из директории *Unilex_A\Gramm* выбирается одна из существующих таблиц и она либо загружается в память (**Load**), либо она выбирается в редактор (**Edit**).

#SAK Имеется 4 типовых задачи: **Processing** (Обработка выбранного объекта), **Import** и **Export** (Копирование выбранных объектов в систему и из системы), а также **View/Edit** (Просмотр или

#IDH_Welcome

§ Welcome to Unilex_A

K частотные словари; конкордансы; корпус текстов; коллекция текстов; словарная база данных; словарные данные;

A частотные словари; конкордансы; корпус текстов; коллекция текстов; словарная база данных; словарные данные;

+

IDH_MainMenu

§ ?????? ?????

K Unilex_A; system; Objects; Texts; Dictionaries; Grammatical Tables; Load; Edit; Tasks; Processing; Import; Export; View; Help; About; Content; Exit; Quit;

A Unilex_A; system; Objects; Texts; Dictionaries; Grammatical Tables; Load; Edit; Tasks; Processing; Import; Export; View; Help; About; Content; Exit; Quit;

+

IDH_Objects

§ ???????

A system; Objects; Texts; Dictionaries; Grammatical Tables; Load; Edit; Tasks; Processing; Import; Export; View; Help; About; Content; Exit; Quit;

K system; Objects; Texts; Dictionaries; Grammatical Tables; Load; Edit; Tasks; Processing; Import; Export; View; Help; About; Content; Exit; Quit;

+

IDH_Tasks

§ ???????

A Objects; Texts; Dictionaries; Grammatical Tables; Load; Edit; Tasks; Processing; Import; Export; View; Help; About; Content; Exit; Quit;

редактирование выбранных объектов). Результаты обработки выбранного объекта помещаются в директорию `\Unilex_A\Results`. Задачи **Import** и **Export** могут выполняться как с/на локальные, так и с/на сетевые диски. Для просмотра и/или редактирования используется стандартный редактор операционной системы или любой другой редактор, заданный в конфигурационном файле.

\$ Система помощи будет описана позднее.

#SA> Объектами в системе Unilex_A являются файлы и совокупности файлов (директории, базы данных), содержащие лингвистические объекты, такие как тексты, словари и грамматические таблицы.

#SA> Текстами в системе Unilex_A являются файлы с расширением имени **.txt**, расположенные в директории `\Unilex_A\Texts`. Файлы могут быть отобраны для обработки как индивидуально (по одному), так и группами (вся директория). Множество текстов, отобранных для обработки, образует корпус текстов.

#SA> Корпус текстов — это множество текстов, отобранных для обработки в некоторой задаче (Tasks.Processing).

#SA> Словарями в системе Unilex_A являются файлы с расширением имени **.dic**, расположенные в директории `\Unilex_A\Dics`. Файлы могут быть отобраны для обработки как индивидуально (по одному), так и группами (вся директория). Множество словарных текстов, отобранных для обработки, образует словарный корпус, который хранится в виде словарной базы данных.

#SA> Грамматическими таблицами в системе Unilex_A являются файлы с расширением имени **.tbl** или **.doc**, расположенные в директории `\Unilex_A\Gramm`, представляющие краткое изложение словообразования русского языка в табличной форме, зависящей от версии системы. Грамматическая таблица **.doc** может быть вызвана для использования в процессе обработки (*load*) или интерактивно отредактирована.

#SA> Задачей обработки в системе Unilex_A является последовательность определенных в системе процедур над корпусом текстов или над словарным корпусом, приводящая к построению лингвостатистических объектов для корпуса текстов или лексикографических объектов для словарного корпуса.

Лингвостатистическими объектами являются различным образом упорядоченные (по алфавиту, по убыванию частот, по алфавиту с конца слова) частотные словари и конкордансы (или словоуказатели — при нулевой длине контекста).

Лексикографическими объектами являются словарные базы данных.

#SA> Импорт (копирование) файлов или директорий может быть произведено в любую субдиректорию директории `\Unilex_A` с любого локального или сетевого диска.

^К Objects; Texts; Dictionaries; Grammatical Tables; Load; Edit; Tasks; Processing; Import; Export; View; Help; About; Content; Exit; Quit;

```
# IDH_Help
$ ??????
# IDH_Objects1
$ ??????
A Objects
> PopUp
# IDH_Texts
$ ??????
A Texts
> PopUp
# IDH_Corpus
$ ??????
A ??????
> PopUp
# IDH_Dics
$ ??????
A Dictionaries
> PopUp
# IDH_Gramm
$ ?????????????? ??????
A Grammatical Tables
> PopUp
# IDH_Proc
$ ?????? ??????????
A Processing, ??????, ??????
> PopUp
# IDH_Import
```

#\$A>Экспортирование (копирование) файлов или директорий может быть произведено из любой субдиректории директории \Unilex_A на любой локальный или сетевой диск.

#\$A> Для просмотра и/или редактирования используется стандартный редактор операционной системы или любой другой редактор, заданный в конфигурационном файле.

Система Unilex_A строит и использует и использует два вида баз данных: текстовую и словарную. Текстовая база данных используется для представления конкордансов, а словарная — для представления словарей.

Все базы данных должны быть отображены в компоненте ODBC Data Sources Windows 95/98/NT. Для этого нужно вызвать с панели управления Администратор ODBC Data Sources и внести соответствующие записи в вкладки User DSN, System DSN и File DSN. Используя кнопки Add и Configure, нужно выбрать Microsoft Access Driver (*.mdb) и далее, следуя логике Администратора, нужно выйти в окно ODBC Microsoft Access Setup и ввести записи Data Source Name (Conc и Dic соответственно), Database: Select (\Unilex_A\Results\Conc.mdb и \Unilex_A\Dic.mdb соответственно) и System Database (отметить Database и кнопкой System Database выбрать из директории \Windows\System файл system.mdv).

```
$ ??????????????
A Import
> PopUp
# IDH_Export
$ ??????????????
A Export
> PopUp
# IDH_ViewEdit
$ View-Edit
A View,Edit
> PopUp
```

VI. Участие в проектах Института русского языка РАН

Сотрудники отдела Машинного фонда русского языка участвовали в издании журнала “Русистика сегодня” (Е. А. Гришина), в проекте издания Орфографического словаря, в подготовке нового Орфографического словаря, а также в проекте отдела стилистики и языка художественной литературы — в создании “Словаря языка русской поэзии XX века Самовитое слово” (Л. И. Колодяжная, Ж. Г. Аношкина). Подробности см. в разделе IV.

VII. Участие в проектах TELRI и ELAN

Отдел Машинного фонда русского языка является участником европейских проектов TELRI (Trans-European Language Resources Infrastructure) и ELAN (European Language Activity Network).

Эти проекты имеют цели, аналогичные целям Машинного фонда русского языка, т. е. разработку и/или сбор языковых ресурсов (корпусов текстов, словарей, программ их обработки), хранение этих ресурсов и предоставление их в пользование международным научным коллективам.

На вопросах формирования текстовых корпусов, их кодификации и маркирования был в последнее время сосредоточен интерес европейских ученых, занятых обработкой текстов на естественных языках.

С 1990 г. разрабатывается международный проект TEI (Text Encoding Initiative), в рамках которого опубликован документ For the Encoding and Interchange of Machine-Readable Texts. В 1994–1996 гг. состоялись несколько международных симпозиумов по текстовым корпусам и по проекту TEI, на которых были рассмотрены вопросы формирования, кодификации, маркирования и обработки текстовых корпусов, создания на их основе лексических баз данных и связанные с этим проблемы.

Все эти материалы доступны в Internet, составляют десятки тысяч страниц и содержат сведения как о конкретных разработках, так и о новых международных стандартах по разметке и кодификации словарей и текстов.

В 1995 г. под эгидой Комиссии Европейского Сообщества была создана Европейская ассоциация лингвистических ресурсов ELRA, ставящая своей целью взаимодействие ученых западноевропейских стран в области создания и распространения источников, необходимых для изучения языков и конструирования систем обработки данных на естественных языках. ELRA распространяет через Internet лингвистические ресурсы, создаваемые в рамках различных проектов и ассоциаций, таких как ELSNET (European Network in Language and Speech), проекты программы Copernicus RELATOR, MULTTEXT (Multilingual Texts), EAGLES и TELRI (Trans-European Language Resources Infrastructure):

ELRA-сервер: <http://www.icp.grenet.fr/elra/home.html>

ELSNET-сервер: <http://www.cogsci.ed.ac.uk/elsnet/home.html> (различные корпуса текстов на английском языке)

RELATOR-сервер: <http://www.relator.research.ec.org/relator/papers> (программные средства, такие как языковые процессоры)

MULTTEXT-сервер:

EAGLES-сервер: <http://www.ilc.cnr.it/EAGLES/home.html> (проекты стандартов на лингвистические ресурсы)

TELRI-сервер <http://www.ids-mannheim.de/telri/telri.html> (корпусы текстов на немецком языке)

В 1995 г. начал издаваться международный журнал “International Journal of Corpus Linguistics”, членом редколлегии которого от Машинного фонда русского языка стал нынешний заведующий Отделом А. Я. Шайкевич.

В последние годы созданы и функционируют в Internet несколько файл-серверов, содержащих русские тексты:

<http://iaiwww.uni-muenster.de/cgi-bin/simplex/lat/lit.html>

rsl.ox.ac.uk:70/11/lib-corn/hunter

<http://anxiety-closet.mit.edu:8001/activities/russian-club/dolina.html>

<http://www-personal.umich.edu/~yegor/kharmis/lat/pisma.html>

<ftp.yt.cache.waseda.ac.jp>

<ftp.funet.fi/pub/culture/russian>

<ftp.pitt.edu/dept/slavic/aatseel>

<http://rudolph.psu.edu/slavic/welcome>

<http://sunsite.oit.unc.edu/sergei/Software/Software.html>

<http://www.auburn.edu/~mitrege/summer-programs/>

<http://www.amu.edu.pl/~sipkadan/mig.htm>

<http://www.amu.edu.pl/~sipkadan/vuk1.html>

<http://www.bucknell.edu/departments/russian/>

<http://www.duke.edu/web/slavic/>

<http://www.elvis.msk.su/koi8install.html>

<http://www.free.net/Docs/cyrillic/notes.en.html>

<http://www.helsinki.fi/~jslindst/>

http://www.nar.com/tag/koi8_explained.html

http://www.osc.edu/ukraine_nonpubl/htmls/macukr.html
<http://www.paragraph.com/PARATYPE/LIBRARY/library.htm>
<http://www.pitt.edu/mapst57/rus/russian.html>
<http://www.pitt.edu/djbpitt>
<http://www.yt.cache.waseda.ac.jp>
<http://www.auburn.edu/korotan>
<ftp.cica.indiana.edu>
<gopher://infomeister.osc.edu:74/11/ukrainian/software/windows/fonts>
<gopher://infomeister.osc.edu:74/59/ukrainian/software/windows/fonts/bkkoi8>
<http://www.ling.helsinki.fi/dept/slav/ccmh.html>
<http://home.aol.com/Tolstoy28>

и множество других.

Имеющиеся сведения о развитии технологии обработки текстов в Европе и в Америке (файл- и WWW-серверы в Internet, ориентация на WYSIWYG-системы типа Microsoft Windows и стандарт SGML) свидетельствуют, что развитие Машинного фонда русского языка в последние годы, хотя и приблизилось к уровню технологий этих стран, но все-таки существенно отстает от доступных там возможностей и несколько отклоняется в методологическом отношении. Это отклонение связано в основном со все еще не преодоленной нами технологической отсталостью и необходимостью создавать национальные лингвистические ресурсы (в то время как западные филологи вполне могут пользоваться стандартизованными американско-европейскими разработками). Велением времени для нас является восстановление собственных файл-серверов в Internet.

Конкретными работами, выполненными Отделом в рамках проектов TELRI и ELAN были:

1. Подготовка и разметка в стандарте SGML/TEI книги Платона "Республика" и
2. Подготовка и разметка в стандарте SGML/PAROLE корпуса газет объемом 5 млн. словоупотреблений.

Автоматическая разметка "Республики" Платона

В рамках TELRI существовала задача создания параллельных размеченных корпусов текстов на различных европейских языках. Наша задача была — передать размеченный русский перевод "Республики" Платона для включения в такой корпус.

Текст (русский перевод А. Н. Егунова) был отсканирован по изданию сочинений Платона в 3 томах, т. 3, стр. 89–454. М.: Мысль, 1971.

Разметка текста была сделана в двух вариантах. Прилагаются отрывки из описаний обоих вариантов.

```
<encodingdesc>
<samplingdecl>
  <p>
    Plato's Republic in Russian
  </p><p>
    The Russian text of Plato's Republic (translated by
    A.N.Egunov), is that of Plato's Works in 3 volumes, v.3,
    p.1, pp.89-454, published by "Mysl" publishing house in
    1971.
  </p><p>
    This edition was scanned by Victor Presnov with the
    help of "Fine Reader" program. The electronic text was
    proof-read by Anatole Shaikevich.
  </p><p>
    Moscow, May 19, 1997
  </p>
</samplingdecl>
<editorialdecl>
  <p>
    The following editorial policy has been implemented:
  </p>
  <list>
    <item>Typographical mistakes found while up-translating
    have been corrected. Very rare diacritic marks,
    denoting word accent, were omitted. All endnote numbers
    were omitted. Translator's
    parenthetical additions were retained and put into
    {}. Translator's headlines, introduced into the
    Russian text, were retained and SGML encoded as
```

```

    subsections;
  </item>
  <item>The Russian text contains standard Cyrillic ASCII
  characters. Latin characters appear in SGML entities
  and in Stephanus edition paging;
  </item>
  <item>When hyphen was used for intra-sentential
  punctuation, it was substituted by &mdash;
  </item>
  <item>P elements were sometimes omitted in order to
  keep grammatical sentence intact;
  </item>
  <item>Top level quotes introducing the speeches of the
  participants of the dialogue were marked by <quote>
  in the present version (platoru1.sgm). However,
  in another version (platoru2.sgm) those speeches
  were treated as parts of piece of drama and marked
  with <sp> signs;
  </item>
  <item>Embedded quotes have been marked by <q>;
  </item>
  </item>
  <item>(Fragments of) poems appearing in the text have been
  marked up by <poem> and <l> signs. As a rule they were
  subordinated to sentence. To comply with TEI practice
  they should be changed to separate paragraph level, but
  that would create serious grammatical problems further
  on;
  </item>
  </list>
</p>
</editorialdecl>
<tagsdecl>
</tagsdecl>
</encodingdesc>
<samplingdecl>
  <p>
  Plato's Republic in Russian
  </p><p>
  The Russian text of Plato's Republic (translated by
  A.N.Egunov) is that of Plato's Works in 3 volumes, v.3,
  p.1, pp.89-454, published by "Mysl" publishing house in
  1971.
  </p><p>
  This edition was scanned by Victor Presnov with the
  help of "Fine Reader" program. The electronic text was
  proof-read by Anatole Shaikevich.
  </p><p>
  Moscow, May 19, 1997
  </p>
</samplingdecl>
<editorialdecl>
  <p>
  The following editorial policy has been implemented:
  <list>
    <item>Typographical mistakes found while up-translating
    have been corrected. Very rare diacritic marks,
    denoting word accent, were omitted. All endnote numbers
    were omitted. Translator's
    parenthetical additions were retained and put into
    {}. Translator's headlines, introduced into the
    Russian text, were retained and SGML encoded as
    subsections;
    </item>
    <item>The Russian text contains standard Cyrillic ASCII
    characters. Latin characters appear in SGML entities
    and in Stephanus edition paging;
    </item>

```

```

    <item>When hyphen was used for intra-sentential
    punctuation, it was substituted by &mdash;
  </item>
  <item>P elements were sometimes omitted in order to
  keep grammatical sentence intact;
  </item>
  <item>Socrates' text as well as speeches of other
  participants of the dialogue were marked by <sp>
  in the present version (platoru2.sgm). The
  names of the speakers were introduced (e.g.
  who=Glaucon);
  </item>
  <item>Embedded quotes have been marked by <q>;
  </item>
</item>
  <item>(Fragments of) poems appearing in the text have been
  marked up by <poem> and <l> signs. As a rule they were
  subordinated to sentence. To comply with TEI practice
  they should be changed to separate paragraph level, but
  that would create serious grammatical problems further
  on;
  </item>
  </list>
</p>
</editorialdecl>

```

Далее Е. Н. Морозовой была сделана программа SGML-разметки InSGML (на языке Pascal), расставляющая маркеры книг, подразделов (subsections), заголовков, абзацев, концов предложений, а также стихотворных цитат. Кроме того, данная программа проводила замену некоторых символов (кавычек, тире) на коды SGML. Кроме того, в отсканированном тексте присутствовали отсылки на расположение текста по страницам и столбцам в эталонном издании, эти отсылки были также переведены в формат SGML.

Были подготовлены два варианта текста с SGML-разметкой (см. описание А. Я. Шайкевича по-английски выше). Для второго варианта была сделана полуавтоматизированная вставка маркеров реплик и авторов реплик (с ручной доразметкой А. Я. Шайкевичем). В обоих вариантах было сделано ручное постредактирование для различения типов кавычек.

Для разметки корпуса газетных текстов Н. А. Ребечкой был создан программный пакет ELA.BAT (платформа MS DOS, язык C) в соответствии с методикой SGML/PAROLE. В задачи проекта входил этап разметки текстов на 1-ом уровне, включающем абзацы. Программа выделяет абзацы по формальным признакам (знакам переноса строки (n)) и расставляет в нужном месте символы начала и конца абзаца (<p>, </p>). Кроме того, программа размечает заголовки текстов по специфическим кодам заголовков, заданным составителями текстового корпуса (<head> — </head>). И, наконец, программный модуль перекодирует все кавычки в текстах на знаки <q> и </q> — открывающая и закрывающая кавычка. Из-за невозможности автоматически добавлять вложенные кавычки включается этап постредактирования текстов для поиска вложенных кавычек и добавления нужного символа (открывающей или закрывающей кавычки) вручную. Дополнительно создан модуль, позволяющий размечать текст на 2-м уровне (уровне предложений (<s>, </s>)), который при необходимости может быть добавлен в базовый пакет.

На входе программы в интерактивном режиме задается имя исходного текста с расширением .txt. На выходе задается имя выходного текста — под этим именем будет записан размеченный на уровне абзацев текстовый файл. Программные модули: ELPRE2, ELANMOD, ELANM2.

Приложение

С. В. Лесников
(Сыктывкарский государственный университет)

Словарный подфонд Диалектологического фонда русского языка

Сотрудниками научно-исследовательского сектора (НИС) и преподавателями, а также аспирантами и студентами филологического факультета Сыктывкарского государственного университета был подготовлен и реализуется проект Словарного подфонда (СПФ) Диалектологического фонда русского языка. Этот подфонд создается в виде гипертекстового словаря из следующих источников:

1. *Словарь русских народных говоров* (33 выпуска СРНГ имеются как в полностью отсканированном, так и фрагментами в текстовом варианте);
2. *Архангельский* (10 тт),
3. *Вологодский* (5 тт) и
4. *Новгородский* (12 тт) областные словари;
5. *Словарь русских говоров Карелии и сопредельных областей* (4 тт),
6. *Словарь с. Лойма Прилузского района Республики Коми* (от А до Я) и
7. *Словарь русских говоров Республики Коми* (от А до Г);
8. *Словарь говоров Подмосковья*,
9. *Словарь брянских говоров* (4 тт),
10. *Словарь говора деревни Акчим Красновишерского района Пермской области* (2 тт),
11. *Словарь современного русского народного говора* (д. Деулино Рязанского района Рязанской обл.),
12. *Мотивационный диалектный словарь: говоры Среднего Приобья* (2 тт) и
13. *Толковый словарь живого великорусского языка В. И. Даля* (несколько вариантов) объемом свыше 250 тысяч словарных статей.

Кроме того имеются:

- компьютерные версии некоторых словарей и энциклопедий (Даля, Ушакова, Ожегова, синонимов, Грамматический, Синтаксический, Орфографический, Иностранных слов, Сводный, БСЭ, Энциклопедии Брокгауза и Эфрона, Словаря Пушкина, отраслевые словари (медицинский, туристический и др.);
- текстовые источники (объемом более 2 Гб, 25 тыс. наименований);
- текущая периодика в Интернете.

Проект поддержан РФФИ грантом № 96-06-80691 “*Гипертекстовый диалектный словарь*”.

Адреса в Интернет:

<http://www.komitex.ru/gowor>,
<http://gowor95.chat.ru/>,
<http://lesnikow.chat.ru/>,
<http://glossa.chat.ru/>,
<http://www.komi.com/Gowor>,
<http://www.online.ru/~gowor>,
<http://lsw.chat.ru>,
<http://vault.agava.ru>,
<http://subscribe.ru/catalog/science.humanity.hypervault> (рассылка)

В интернетовской рассылке предполагается интерактивное объединение данных о русских словарях (XIX–XX вв) в виде гипертекстовой библиографической системы (с размещением в Интернет и на компакт-диске). Проект поддержан РФФИ грантом 2000-06-80176 (научный руководитель С. В. Лесников. Одной из целей проекта является анализ энциклопедий и словарей русского языка для гипертекстовой интеграции этих словарей в нелинейной форме с учетом реляционных, иерархических и сетевых парадигматических связей посредством реализации синтагматических связей в интерактивном режиме на ЭВМ оцифрованных лексикографических материалов, что позволит на основе новых информационных технологий объединить лексикографические материалы, обеспечить их оперативный ввод в научный оборот — с целью оптимизации научных исследований в современной лексикографии.

Литература

1. *Машинный фонд русского языка: идеи и суждения*, М.: Наука, 1986.
2. *Лингвистические задачи и обработка данных на ЭВМ*, Институт русского языка РАН, М., 1987.
3. *Машинный фонд русского языка: Предпроектные исследования*, Институт русского языка РАН, М., 1988.
4. *Бюллетень Машинного фонда русского языка, вып. 1*, Институт русского языка РАН, М., 1992.
5. *Бюллетень Машинного фонда русского языка, вып. 2*, Институт русского языка РАН, М., 1992.
6. *Бюллетень Машинного фонда русского языка, вып. 3*, Институт русского языка РАН, М., 1996.
7. *Труды Машинного фонда русского языка, том I*, Институт русского языка РАН, М., 1991.
8. *Труды Машинного фонда русского языка, том II*, Институт русского языка РАН, М., 1996.
9. *Вторая Всесоюзная конференция по созданию Машинного фонда русского языка (тезисы докладов)*, Институт русского языка АН СССР, М., 1987.
10. *Третья Всесоюзная конференция по созданию Машинного фонда русского языка (тезисы докладов)*, ч. 1–2, Институт русского языка АН СССР, М., 1989.
11. *Материалы III Всесоюзной конференции по созданию Машинного фонда русского языка (тезисы докладов)*, Институт русского языка АН СССР, НИВЦ МГУ, М., 1990.
12. *Доклады Третьей Всесоюзной конференции по созданию Машинного фонда русского языка (тезисы докладов)*, Институт русского языка АН СССР, М., 1990.
13. **Андрющенко В. М.** *Концепция и архитектура Машинного фонда русского языка*, М.: Наука, 1989.
14. **Аношкина Ж. Г.** *Подготовка частотных словарей и конкордансов на компьютере (пособие для филологов)*, Институт русского языка РАН, М.: Русские словари, 1995.
15. **Исаев И. А., Черкасова Г. А.** *Информационная система "ДАРЯ" // Вторая Всесоюзная конференция по созданию Машинного фонда русского языка (Тезисы докладов)*. Институт русского языка АН СССР, М., 1987.
16. **Мошкович (Аношкина) Ж. Г.** *Автоматизированная лексикографическая система Унилекс-2*, НИВЦ МГУ, М., 1989.
17. **Пшеничнова Н. Н.** *Автоматизированный вариант Диалектологического атласа русского языка (АВ ДАРЯ) // Лингвистические задачи и обработка данных на ЭВМ*, М., 1987.
18. **Пшеничнова Н. Н.** *Диалектологический атлас русского языка и его Автоматизированный вариант // Общеславянский лингвистический атлас. Материалы и исследования. 1985–1987*, М., 1989.

Труды, связанные с программой создания Машинного фонда русского языка

1. *Бюллетень Фонетического фонда русского языка*, Herausgeber: Christian Sappok (Bochum), Лия Васильевна Бондарко (Ленинград) № 1–4, 1988–.
2. *Проблеми створення машинних фондів мов*, Український мовно-інформаційний фонд (тези доповідей), Київ, 1991.
3. *Лексикографическая разработка фразеологизмов для словарей различных типов и для Машинного фонда русского языка (Материалы к методической школе-семинару)*, Институт русского языка АН СССР, М., 1988.
4. *Фразеологические словари и компьютерная фразеология*. Тезисы сообщений школы-семинара, Научный совет по лексикологии и лексикографии АН СССР. Орел, 1990.
5. *Макет словарной статьи для автоматизированного толково-комбинаторного словаря русских фразеологизмов. Образцы словарных статей*. Институт языкознания АН СССР, М., 1991.
6. *Лингвистическая концепция терминологического банка данных Машинного фонда русского языка (Проект)*, под ред. А. С. Герда. Институт русского языка АН СССР, МГПИИЯ им. М. Тореза, М.: 1989.