

УДК 81.374

DOI: 10.25559/SITITO.15.201902.362-374

Алгоритм разработки и создания цифрового академического словарного корпуса русского языка

С. В. Лесников^{1*}, А. В. Лесников², Г. С. Лесников³, А. М. Фарина⁴¹ Институт лингвистических исследований Российской академии наук, г. Санкт-Петербург, Россия 199053, Россия, г. Санкт-Петербург, Тучков пер., д. 9

* serg@lsw.ru

² Московский государственный университет имени М.В. Ломоносова, г. Москва, Россия 119991, Россия, г. Москва, ГСП-1, Ленинские горы, д. 1³ Северный государственный медицинский университет, г. Архангельск, Россия 163000, Россия, г. Архангельск, пр. Троицкий, д. 51⁴ Сыктывкарский государственный университет имени Питирима Сорокина, г. Сыктывкар, Россия 167001, Россия, Республика Коми, г. Сыктывкар, Октябрьский пр., д. 55

Аннотация

В данной статье рассматриваются академические толковые словари русского языка /РЯ/ (прежде всего, БАС, СРНГ, МАС, СРЯ 1895-1937) как ядро (основа) академического словарного корпуса (АСК) [14, с.213-214; 15, с. 111-114; 18, с.25-28; 27, с.226-257; 32, с.76-83; 35, с.98-102] РЯ, а также алгоритм, концепция и принципы разработки и создания АСК РЯ. АСК конструируется в форме ГИЗАУРУСА (гипертекстового тезауруса) [5, с. 119-125; 22, с.268-269; 25, с.280-281] в нелинейной форме с учетом реляционных, иерархических и сетевых парадигматических связей, что позволит посредством соответствующей классификации и систематизации, дигитализации и ретроцифровизации академических толковых словарей РЯ структурировать и объединить лексикографические материалы, обеспечив тем самым их оперативный ввод в научный оборот с целью оптимизации научных исследований в современной лексикографии.

Научная значимость АСК определяется необходимостью централизованного описания русской лексики, так как многочисленные разработки в исследуемой области и каждая в отдельности имеют собственные задачи и направление исследования, что не позволяет увидеть полную языковую картину Российской Федерации, которую может показать информационно-поисковый /ИП/ АСК. Для этой цели предполагается смоделировать, разработать и создать работоспособный интерактивно пополняемый ИП АСК РЯ, позволяющий объединить в единую базу различные лексикографические источники.

Ключевые слова: база данных, гизаурус, гипертекстовый тезаурус, метаязык, искусственный интеллект, корпус, лексикография, лингвистика, поиск, русский язык, словарь, справочник, текст, энциклопедия.

Для цитирования: Лесников С. В., Лесников А. В., Лесников Г. С., Фарина А. М. Алгоритм разработки и создания цифрового академического словарного корпуса русского языка // Современные информационные технологии и ИТ-образование. 2019. Т. 15, № 2. С. 362-374. DOI: 10.25559/SITITO.15.201902.362-374

© Лесников С. В., Лесников А. В., Лесников Г. С., Фарина А. М., 2019



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.



Academic Vocabulary Russian

S. V. Lesnikov^a, A. V. Lesnikov^b, G. S. Lesnikov^c, A. M. Farina^d

^a Institute for Linguistic Studies of the Russian Academy of Sciences, Saint-Petersburg, Russia
9 Tuchkov Ln., St. Petersburg 199053, Russia

^b Lomonosov Moscow State University, Moscow, Russia
1, Leninskie gory, Moscow 119991, Russia

^c Northern State Medical University, Arkhangelsk, Russia
51 Troitskiy Ave., 163000, Arkhangelsk, Russia

^d Pitirim Sorokin Syktyvkar State University, Syktyvkar, Russia
55 Oktyabrsky Ave., Syktyvkar 167001, Komi Republic, Russia

Abstract

In this research project, it is planned to construct an information retrieval set of academic dictionaries of the Russian language in the form of information retrieval academic vocabulary. Academic dictionaries of the Russian language is developed in the form of hypertext information retrieval system in a nonlinear form, taking into account the relational, hierarchical and network paradigmatic relations through the implementation of syntagmatic relations in an interactive mode on a personal computer digitized academic dictionaries of the Russian language, which will allow on the basis of digital technology through appropriate classification and systematization of dictionaries, reference books and encyclopedias of the Russian language to combine lexicographic materials, to ensure their rapid entry into scientific circulation - to optimize scientific research in modern lexicography.

Relevance of operation is caused by need of generalization of scientific results for the purpose of determination of the priority directions of research in the specified area.

Keywords: database, gizaaurus, hypertext thesaurus, the meta-language, artificial intelligence, corpus, lexicography, linguistics, search, Russian language, dictionary, reference, text, encyclopedia.

For citation: Lesnikov S.V., Lesnikov A.V., Lesnikov G.S., Farina A.M. Academic Vocabulary Russian. *Sovremennye informacionnye tehnologii i IT-obrazovanie* = Modern Information Technologies and IT-Education. 2019; 15(2):362-374. DOI: 10.25559/SITITO.15.201902.362-374



Введение

На современном этапе развития русского языка необходима фиксация, описание и репрезентация академических словарей, справочников и энциклопедий русского языка, так как многие культурные реалии уходят в прошлое и становятся частью истории. В связи с глобальной индустриализацией и урбанизацией лексика уходит в прошлое, и уносит за собой огромный массив истории русского народа. Необходимость создания академического словарного корпуса (АСК) РЯ для обобщающего исследования русской лексики как части исторической картины нашей страны очевидна. Кроме того, получение репрезентативного академического словарного корпуса позволит предоставить квалифицированный доступ к богатству материальной и духовной народной культуры как для специалистов, так и для всех социальных групп общества, включая студентов, аспирантов, магистрантов и школьников и всех интересующихся историей и культурой нашей страны.

Обзор работ по теме исследования

Под ГИЗаУРУСом понимается гипертекстовая реляционная база данных с гиперссылками (цифровой идеографический словарь, отображающий семантические отношения между лексическими единицами и предназначенный для поиска слов по их значениям) [33, с.29]. Один из основателей и идеологов Машинного фонда русского языка /МФ РЯ/ [1-4; 6-9; 21, с.58-60; 40-42; 44], Ю. Н. Караулов писал: «Гипертекст — или нелинейный текст, — такая организация записи отдельных отрывков, фрагментов множества текстов (контекстов), которая позволяет комбинировать их в том или ином порядке, основываясь на задаваемых заранее переменных связях между ними, с целью получения каждый раз разных линейных последовательностей (текстов)» [11, с.154].

Первая попытка сконструировать, точнее, описать «гипертекстовую машину» была предпринята в 1945 году [46]. Рассматривая историю развития счётных машин, начиная с Лейбница и Бэббиджа, В. Буш предложил проект электронно-механического устройства, названного «метех» (от английского MEMory EXtension), которое должно дать человеку инструмент для усиления мыслительных способностей. Метех позволяет человеку определённым способом организовать свои записи, статьи, книги. С помощью механических средств (рычагов, кнопок и т.п.) тексты, записанные на микрофильмы, могут быть связаны между собой так, что переход от одного к другому осуществляется непосредственно, даже если физически они расположены далеко друг от друга. При чтении таким способом возможны пропуски страниц, возврат назад или просмотр каждых десяти страниц. Одновременно на экран могут вызываться несколько микрофильмов. Дополнительно к существующим материалам могут записываться новые - статьи, заметки, а также комментарии, ссылки и т.п.

Следующее слово в истории гипертекста сказал Теодор Нельсон - программист, придумавший термин «гипертекст», дав следующее определение гипертекста: текст, не имеющий линейной последовательности («нелинейное письмо»). Нельсон стремился реализовать идеи удачного, по его мнению, использования компьютера в виде всеобщей системы электронных публикаций и архива. А в основе подобной системы должно лежать понятие гипертекста. По мнению Т. Нельсона, компьютер

призван сделать человека свободным, а поэтому нужно разработать систему, максимально облегчающую работу с информацией [55]. Литература (понимаемая в широком смысле как информация, предназначенная для сохранения) представляет собой непрерывную систему взаимосвязанных произведений, и это надо рассматривать не как определение, а как обнаруженную данность, т.е. «просто непоследовательное письмо (writing), текст, который разветвляется» и предоставляет читателю выбор [56]. Таким образом, под «непоследовательным письмом» Нельсон подразумевал организацию «нелинейного текста» посредством гиперссылок, меню или др. элементов дружественного интерфейса и дизайна информации на экране персонального компьютера в интерактивном режиме.

В 1980-е годы создаётся множество коммерческих гипертекстовых систем. Согласно Conklin J., гипертекстовые системы классифицируются так: 1) библиотечные макросистемы; 2) средства исследования проблем («инструментальные средства для поддержки неструктурированного мышления на ранней стадии исследования проблем, когда человек имеет дело с многочисленными, плохо связанными одна с другой идеями»); 3) системы для просмотра («browsing») базы данных (то есть системы, в которых добавление данных пользователями не предусматривается); 4) системы широкого назначения (для экспериментирования в различных областях применения гипертекста). Механизм, заключающийся в возможности связать отрывки текста, переходить от одного к другому, называется гипертекстом, или нелинейным текстом: а) «надтекст» - некая единица информации, частями которой являются тексты, и/или б) текст, части которого имеют «сверхсвязи», то есть соединены друг с другом не линейным отношением в одномерном пространстве (отношением следования, как в обычном тексте естественного языка), а множеством различных отношений, представляемых в многомерном пространстве [53]. Построение формальной грамматики словарной статьи АСК) РЯ базируется на теории формальных грамматик, разработанной Н.Хомским [47-52].

Правила грамматики записываются в обозначениях Бэкуса-Наура. Назначение правил: предоставить средство для адекватного описания абстрактных конструкций, принадлежащих некоторому формальному языку. При этом каждая из металингвистических формул задает правило порождения допустимых значений соответствующих метапеременных, которыми являются основные символы (1988 ЛСВ и ЗОВ в нашем случае — для автоматизированного словаря - это элементарные и/или составные параметры словарной статьи [9, с.64-70; 38, с.107-119]) или их цепочки, расположенные между разделителями [или получающиеся посредством последовательной замены в этих цепочках метапеременных с их допустимыми (порождаемыми) значениями].

В нашем случае Академический словарно-грамматический фонд /АСГФ/ «должен состоять из нескольких подфондов, число которых не следует заранее ограничивать. Одним из таких подфондов станет свод академических словарей - наиболее важных источников зафиксированных на сегодняшний день знаний о русском языке. Такой свод должен образовать однородное семейство баз данных, в пределах которого каждый из словарей продолжал бы существование в качестве отдельной базы данных и в то же время был бы связан средствами СУБД и номенклатурой своих составляющих со всеми другими словарями. Этот объединенный словарь мог бы не только вы-



полнять справочные функции, но и быть источников нового систематического изучения лексики» [1, с.32].

Л.В. Щерба утверждал, что в основе словарей академического типа «лежит единое (реальное) языковое сознание определенного человеческого коллектива в определенный момент времени» [45, с.266-267] и словарь академического типа рассматривал как нормативный, описывающий лексическую систему современного языка, т.е. в нём не должно быть фактов, противоречащих современному употреблению: «к нормативному (или академическому) словарю обращаются для самопроверки, а иногда и для нахождения нужного в данном контексте слова» [45].

«Словарь должен быть достаточно полным, включающим все богатство современного русского литературного языка. Словарь должен быть историческим... Словарь должен дать всестороннюю семантическую, стилистическую и грамматическую характеристику слов: правописания, всего многообразия значений, особенностей фразеологического и стилистического употребления слов, образующую роль слова в группах родственных слов... В научном отношении Академический Словарь русского языка должен представлять собою такой источник, на основе которого в дальнейшем мог бы быть создан ряд малых словарей-справочников разного типа... Полный охват Академическим Словарем всех слов русского языка оправдывает его многогранность ..., но мы не должны бояться такого объема издания, так как по своей научной значимости Словарь в этом случае будет таким фундаментальным документом, в котором будет отражена история развития русского языка за 200 лет и зафиксировано полностью все богатство содержания современного русского литературного языка» [43, с.3-4].

В конце 80-х гг. существенно активизировались исследования по русской филологии с применением ЭВМ в рамках проекта по созданию МФ РЯ. МФ РЯ формировался в виде нескольких систем комплексной автоматизации лингвистических исследований и предусматривал «накопление на машинных носителях всего лексического богатства русского языка, создание фонда лингвистических алгоритмов и программ, фонда полностью завершённых систем автоматического анализа и синтеза русского текста, нескольких информационно-справочных систем по языкознанию» [3, с.8].

Впервые мысль о МФ РЯ высказал академик А. П. Ершов в докладе «К методологии построения диалоговых систем: феномен деловой прозы» 26 сентября 1978 г. на научной конференции «Диалог-78»: «Любой прогресс в области построения моделей и алгоритмов останется, однако, академическим упражнением, если не будет решена наиважнейшая задача создания Машинного фонда русского языка. Это фундаментальная проблема, решение которой будет иметь очень большую научную, общекультурную и прикладную ценность. Не мне, конечно, составлять спецификацию такого фонда, но думается, что по крайней мере он должен содержать полный словарь и генератор словоформ, а также формализованный толковый словарь (тезаурус) русского языка. Очень хотелось бы видеть, что создание машинного фонда русского языка квалифицированными лингвистами опережало бы создание производственных лингвистических систем, потому что это не только бы позволило избежать дублирования больших усилий, но и защитило бы здоровую ткань русского языка от самоуправства и некачественного подхода» [8, с.115].

Проблемам создания МФ РЯ было посвящено три Всесоюзные

конференции (1983, 1987, 1989), на которых речь шла о предпосылках создания, основных проектных решениях и, следуя принципу интеграционного подхода [4].

В рамках проекта МФ РЯ разрабатывались 9 фондов-составляющих: 1) Генеральный словарь, 2) Словарный, 3) Текстовый, 4) Грамматический, 5) Терминологический, 6) Диалектологический, 7) Исторический, 8) Фонетический, 9) Лингвистический программно-источниковый.

В настоящий момент от подфондов фактически перешли к корпусам. Машинный фонд (корпус) - иерархически-разветвленная компьютерная система, предназначенная для решения как информационно-поисковые, так и лингвистические задачи в интерактивном режиме (напр., Интернете).

Учитывая макро и микроструктуры МФ РЯ [1-4; 21, с.58-60], система взаимосвязей и взаимозависимостей основных составляющих МФ РЯ представляется довольно сложной. Например, трудно выделить какой-либо из этих подфондов (корпусов) в качестве главного, определяющего или основного. Пожалуй, только относительно Генерального словаря /ГС/ будет меньше споров, если он будет рассматриваться как естественно создаваемый и пополняемый усилиями разработчиков всех других подфондов. Лингвистический программно-источниковый фонд (а это в большей степени программное обеспечение) может быть выделен особо как своеобразный интерфейс (гиперсистема), например, через ГС интеграции СловФ и ТекстФ, которые в свою очередь имеют в своем составе пересечения ТТ, ДФ и ИФ. Вероятно, проблема взаимосвязей и координации подфондов МФ РЯ — это принципиально важная и заслуживающая пристального внимания отдельная задача уже не на этапе проектирования, а в недалеком будущем, по мере формирования МФ РЯ. На наш взгляд, представление филологических материалов в компьютерной форме и внедрение современных методов научного анализа, основанных на применении цифровых технологий в русской лексикографии, позволяет взглянуть на эти проблемы по-новому и, в частности, создавать не только серию словарей, а один общий академический словарный корпус АСК - автоматизированный словарь (словарь-тезаурус, гизаурус).

Основа цифрового корпуса

В качестве основы АСК РЯ предполагается использовать оцифрованные академические толковые словари РЯ, изданные Императорской АН, АН СССР, РАН (315 тт., вып., частей):

Словарь Академии Российской (1789-1794, 6 частей);

Словарь Академии Российской по азбучному порядку (1806-1822, 6 частей);

Церковный словарь (1817-1819, 5 тт.);

Общий церковно-славяно-русский словарь (1834, 2 части);

Корнеслов русского языка (1842, 2 части);

Словарь церковно-славянского и русского языка (1847, 4 тт.);

Опыт областного великорусского словаря, Дополнения к Опыт (1852-1858, 2 тт.);

Словарь церковно-славянского и русского языка (1867-1868, 4 тт.);

Словарь русского языка (1891-1937, 53 вып.);

Материалы для словаря древне-русского языка по письменным памятникам (1893-1912, 6 тт.);

Словарь древнего славянского языка (1899, 1 том);

Толковый словарь русского языка (1934-1940, 4 тт.);



Материалы для терминологического словаря Древней России (1937, 1 том);

БАС1 Словарь современного русского литературного языка (1948-1965, 18 тт.);

Словарь русского языка (1949, 1 том);

Словарь языка Пушкина (1956-2000, 4 тт.);

Рукописный лексикон первой половины XVIII века (1964, 1 том);

СРНГ Словарь русских народных говоров (1965-2018, 51 вып.);

Словарь русского языка XI-XVII (1975-2015, 30 вып.);

Новые слова и значения (1977-2014, 24 тт.);

Словарь русского языка XVIII века (1984-2013, 20 вып.);

МАС (1985-1988, 4 тт.);

Словарь древнерусского языка XI-XIV (1988-2012, 9 тт.);

БАС2 Словарь современного русского литературного языка (1991-1994, 6 тт.);

Русский толковый словарь (1994, 1 том);

Славянские древности. Этнолингвистический словарь (1995-2012, 5 тт.);

Русский семантический словарь (16300 слов) (1998-2015, 4 тт.);

Большой толковый словарь русского языка (1998-2000, 2 тт.);

Толковый словарь русского языка конца XX века. Языковые изменения (1998, 1 том);

Новый словарь русского языка. Толково-словообразовательный (2000, 2 тт.);

Толково-понятийный словарь русского языка (2003, 1 том);

Словарь обиходного русского языка Московской Руси XVI-XVII (2004-2016, 7 тт.);

БАС3 (2004-2017, 24 тт.);

Толковый словарь русского языка начала XXI века. Актуальная лексика (2007, 1 том);

Толковый словарь русского языка с включением сведений о происхождении слов (2011, 1 том);

Словарь русской ментальности (2014, 2 тт.);

Толковый словарь ключевых слов русского языка (2014, 1 том);

Академический толковый словарь русского языка (2016, 2 тт.). Отдельно выделим:

1) **САР1** - Словарь Академии Российской, в 6 частях, от А до Я. СПб. 1789-1794.

2) **САР2** - Словарь Академии Российской, в 6 частях, от А до Я. СПб. 1806-1822.

3) **СЦРЯ1** - Словарь церковно-славянского и русского языка, составленный II Отделением Академии Наук, в 4 томах, от А до Я. СПб. 1847.

4) **СЦРЯ2** - 2-ое издание СЦРЯ. СПб. 1867-1868.

5) **СРЯ 1891-1895** - Словарь русского языка, составленный II Отделением Академии Наук (под редакцией акад. Я. К. Грота), т. I, вып. 1, 2 и 3 (от А до Дя). СПб. 1891-1895. СРЯ 1891-1895 по существу незакончен, так как его продолжение построено было уже на иных принципах и должно приниматься как новое издание.

6) **СРЯ 1897-1929** - издание, начатое по плану акад. А.А.Шахматова и в значительной части им же разработанное, также осталось незаконченным. В него вошли: т. II, вып. 1-9 (Е, Ж, З), т. III, вып. 1-2 (И-Издергивать), т. IV, вып. 1-10 (К-Крошечный), т. V, вып. 1-3 (Л-Лисичий), т. VI, вып. 1-2 (М-Маститый), т. VIII, вып. 1-2 (Не-Недодорубщик). Это издание, 6 по счету, выходило в 1897-1916, 1922, 1926-1929 гг.

7) **СРЯ 1929-1937** - Новое издание СРЯ, начатое с конца 1929 г., является в серии словарных академических изданий седьмым, представляя собою совершенно новое предприятие: это — толковый словарь современного русского языка, взятый в историческом развитии (СРЯ 1933. Т.1, вып.2. Ажно-Аллотриология. Предисловие). С 1934 г. нумерация томов «СРЯ» меняется. Впредь каждой букве будет соответствовать свой номер тома, а именно: А - том 1, Б - том 2, В - том 3, Г - том 4, Д - том 5, Е - том 6, Ж - том 7, З - том 8, И - том 9, К - том 10, Л - том 11, М - том 12, Н - том 13, О - том 14 и т. д. (СРЯ 1934. Т.11, вып.3).

Основная задача, концепция и принципы разработки корпуса

Основной задачей создания академического словарного корпуса русского языка на основе академических толковых словарей является развернутое описание употребления русской лексики, а также показ лексического и семантического богатства русского языка во всем его многообразии за последние несколько веков, что безусловно будет способствовать повышению культуры в русской речи не только в нашей стране, но и во всем мире, что налагает особую ответственность на авторов-составителей и редакторов АСК РЯ на отбор академических толковых словарей РЯ (изданных Императорской АН, АН СССР, РАН), и при этом должен быть безупречным: язык толкований значений слов, иллюстраций, фразеологизмов и речений; отбор цитатного материала, показ сочетаний слов и т. д. В основе концепции разработки и создания АСК лежат следующие принципы:

1) при отборе лексикографического материала - максимально полный охват слов и выражений в терминологических значениях из разнообразных корпусов текстов с использованием современных информационных ресурсов;

2) на уровне макроструктуры АСК - тезаурусный, тематический, идеографический и концептный принципы систематизации материалов по категориям, разделам и рубрикам;

3) на уровне микроструктуры АСК - расширенные зоны: вокабулы (заглавного слова, лемм, вариантов лексических единиц); значения (дефиниции, изъяснения, интерпретации) и полисемии (многозначности, оттенков значения, переносных значений, отсылочных значений); комментарии и иллюстрации семантики рассматриваемых терминов терминосистемы русского языка.

В основе проекта разработки и создания АСК лежит идея комплексного многоаспектного лексикографического описания системы русского языка, включающей предметный, географический, грамматический, исторический, источниковедческий, прагматический, семантический, стилистический, фонетический, этимологический и другие аспекты и отражающей реальную лексико-грамматическую структуру лексических единиц во всем многообразии.

При формировании АСК выдерживаются следующие принципы:

а) функциональность (совместное рассмотрение структуры и системы многоаспектного поиска с приоритетом функций над структурами);

б) глобальная цель (оптимальность подсистем не является оптимумом для системы поиска);

в) единство (совместное рассмотрение системы как целого и как совокупности частей);



г) развитие (учёт изменяемости системы, её способности к развитию, накоплению информации с учётом динамики окружающей среды, напр., использование статистики пользовательских запросов и новых лексикографических источников);

д) иерархия (учёт использования соподчинения и ранжирования частей, напр., систематизация по тематическому, хронологическому и частотному параметрам);

е) агрегирование (объединение и укрупнение значений);

ж) децентрализация (распараллеливание запросов по распределенным на разных серверах лексическим базам данных);

з) неопределенность (по возможности учёт вероятностного наступления события, напр., за счёт накопления статистики).

Алгоритм разработки и создания корпуса

Кратко алгоритм разработки и создания АСК формулируется следующим:

1) Выбор источников и кластеризация предметной области. При выборе источников учитывается: - наиболее точное соответствие лексического материала предметной области [25, с.280-281]; - технологическая насыщенность и важность источников.

2) Составление базового и генерального словарей (аскрипторы, дескрипторы, лексические единицы, одиночные слова (существительные, прилагательные, глаголы, наречия); именные словосочетания; лексически значимые компоненты сложных слов; сокращения слов и словосочетаний; реляторы; термины и терминологизмы) [31, с.37-45].

3) Выявление семантических полей посредством определения кортежей условной эквивалентности [37, с.84-89], использование фреймового подхода [30, с.51-54] для реализации гизауруса (гипертекстового тезауруса).

По мнению профессора Массачусетского технологического института, специалиста в области представления данных, искусственного интеллекта и робототехнике Марвина Минского, «основные структурные элементы, образующие фундамент для развертывания процессов восприятия, хранения информации, мышления и разработки языковых форм общения, должны быть более крупными и иметь более четкую структуру; их фактическое и процедуральное содержание следует более тесно увязывать друг с другом с тем, чтобы получить возможность объяснить феномен силы и «быстродействия» человеческого мышления» [54, с.6]. И именно поэтому Минский в своей работе делает попытку создать единую и стройную теорию, отправным моментом для которой «служит тот факт, что человек, пытаясь познать новую для себя ситуацию или по-новому взглянуть на уже привычные вещи, выбирает из своей памяти некоторую структуру данных (образ), называемую нами фреймом, с таким расчетом, чтобы путем изменения в ней отдельных деталей сделать ее пригодной для понимания более широкого класса явлений или процессов. Фрейм является структурой данных для представления стереотипной ситуации. С каждым фреймом ассоциирована информация разных видов. Одна ее часть указывает, каким образом следует использовать данный фрейм, другая - что предположительно может повлечь за собой его выполнение, третья - что следует предпринять, если эти ожидания не подтвердятся. Фрейм можно представлять себе в виде сети, состоящей из узлов и связей между ними. «Верхние уровни» фрейма четко

определены, поскольку образованы такими понятиями, которые всегда справедливы по отношению к предполагаемой ситуации. На более низких уровнях имеется много особых вершин-терминалов или «ячеек», которые должны быть заполнены характерными примерами или данными.

Каждым терминалом могут устанавливаться условия, которым должны удовлетворять его задания. Простые условия определяются маркерами, напр., в виде требования, чтобы заданием терминала был какой-либо субъект, или предмет подходящих размеров, или указатель на субфрейм определенного типа. Более сложными условиями задаются отношения между понятиями, включенными в различные терминальные вершины.

Группы семантически близких друг к другу фреймов объединены в систему фреймов. Результаты существенных действий представляются в виде трансформаций между фреймами системы. Это дает возможность моделировать такие понятия, как внимание и ценность информации, сделать более экономичными некоторые типы вычислений, а также показать эффективность использования фреймов в системах ИИ» [54, с.7-8. ИИ - искусственный интеллект].

Фрейм можно определить как поименованную структуру, которая состоит из слотов. В свою очередь, слот - это конструкция из имени и значения слота. Значение слота - шпация.

Достоинством фреймового подхода является то, что слоты фрейма могут группироваться как самостоятельные структурные единицы (в частности, в качестве шпаций могут указываться имена других фреймов), а поэтому возможна их обработка как единого целого, т.е. как очередной фрейм и т.д. Таким образом, механизм организации ссылок, а теперь мы скажем ГИПЕРТЕКСТОВЫХ ссылок, позволяет создавать в интерактивном режиме на современных компьютерах из отдельных фреймов сложные сетевые структуры и системы именно благодаря установленным связям между понятиями, объектами и событиями [30; 37].

4) Установление парадигматических (способы выражения: аналитический, графический, лексикографический, табличный), иерархических, ассоциативных, грамматических, межязыковых, семантических... связей [16, с.269-279].

5) Создание словарных статей и лексико-систематических (тематические, категориальные, смешанные), пермутационных, иерархических и др. указателей и списков специальных категорий лексических единиц /ЛЕ/ (общие категории: названия дисциплин и отраслей деятельности; методы, материалы; методы, процессы, операции, явления; свойства, величины, параметры, характеристики; отношения, структуры, модели, законы, правила, абстрактные понятия) [24, с.190-203].

6) Автоматизация процессов построения АСК: - частотный, статистический, компьютерный, математический анализ; б) коррекция статей; в) сортировка словарей; г) проверка взаимности и непротиворечивости ссылок; д) составление указателей; е) конвертация в требуемых форматах и кодировках [12, с.622-630]. Построение гизауруса с использованием компьютеров [19, с.200-202] позволяет автоматизировать следующие процессы построения гизауруса [20, с.39-53]: частотный анализ, коррекция статей, алфавитная сортировка словаря, проверка взаимности и непротиворечивости ссылок, составление указателей, распечатка в требуемых формах [23, с.30-34].

7) Экспертиза, регистрация, ввод АСК в научный оборот и технологическое внедрение [25, с.51-54].



При моделировании АСК предполагается учесть следующие обстоятельства:

1) Компьютерные словари (наследуя недостатки книжных словарей) имеют следующие особенности и проблемы: - неполнота генерального словника и в целом словаря; - трудности поддержания словаря в актуальном состоянии; - требуют наличие специфичных программных и аппаратных средств. При этом одним из главных достоинств электронных версий словарей является то, что электронные словари многофункциональны, доступны, и неуклонно становятся дешевле, чем их бумажные аналоги.

2) Словник и содержимое книжных словарей в процессе постепенного и длительного их издания требует значительного дополнения, расширения и уточнения. В период моделирования академического словарного корпуса естественно будут появляться новые словари русского языка, а также множество аспектных, терминологических словарей и энциклопедий, которые могут существенно обогатить АСК.

3) Дифференциальный принцип отбора словников (нормативный, терминологический, аспектный и др.) для академических словарей, оказавшийся продуктивным при описании лексики в книжных версиях, для цифровой версии АСК не является принципиальным, т.к. в компьютерной форме лексическая база данных автоматически становится полной и зависит лишь от полноты введенных источников (текстов записей информантов, картотек и любых источников лексикографических материалов), что позволяет в компьютерной форме конструировать АСК в тезаурусной (наиболее полной) форме.

4) Порядок расположения фразеологизмов и устойчивых словосочетаний в академических словарях, как правило, подчинён избранному для лексики словаря алфавитному или гнездовому принципам, что затрудняет поиск нужного словосочетания не только в объёмных вокабулах, но и в словаре или энциклопедии в целом. Именно поэтому в АСК предполагается реализовать несколько входов: авторский, алфавитный, библиографический, через веб-интерфейс, посредством генерального словника, географический и территориальный, гнездовой, источниковедческий, систематический и пермутационный, тематический, предметный, хронологический, а также, по ключевым словам, и словосочетания и др., что безусловно по возможностям отличает АСК от книжных версий академических словарей русского языка.

Среди областей применения АСК РЯ можно выделить следующие:

- составление, уточнение и оптимизация поисковых запросов;
- автоматическое выделение из текстов документов наиболее содержательных фрагментов, обеспечивающее пользователю возможность быстро просматривать массивы документов больших объемов;
- установление гипертекстовых ссылок (связей) между похожими документами и/или их фрагментами, что дает возможность пользователю, найдя один релевантный документ, запросить похожие на него документы;
- реферирование и аннотирование, редактирование текстов;
- повышение полноты и точности поиска в полнотекстовых БД на естественном языке;
- автоматизированное индексирование текстов, документов и запросов;

- анализ терминологического состава той или иной отрасли знания;
- терминологический контроль и нормализация лексики конкретной отрасли знания, обеспечение единого и формализованного представления информации в ИПС и БД/БЗ;
- явное выражение парадигматических отношений, существующих между лексическими единицами.

Кроме этого, АСК РЯ может быть использован как средство:

- описания (инвентаризации) существующего уровня знаний в той или отрасли наук;
- описания истории развития специальных знаний в определенной предметной области;
- исследования особенностей развития отраслевых знаний, включая количественное исследование вклада отдельных ученых и научных школ в развитие общей теории с учетом как количества, так и новизны (в зависимости от статуса вводимого термина - основной термин или синоним уже существующего наименования понятия) и иерархического статуса понятий, вводимых соответствующим термином;
- систематизации терминологии (в тезаурусе, где система понятий эксплицитна, т.е. явно выражена, иерархия понятий всегда представлена наглядно);
- стимулирования развития знаний (напр., выявить пробелы (пропуски, лакуны), малоизученные понятия и «точки сгущения» в терминологических и понятийных системах, что может служить ориентиром в последующих исследованиях и позволит выявить аналогии подсистем понятий, а также определить тенденции выделения подсистем понятий и рубежи знаний, от которых возможно дальнейшее развитие данной области).

Заключение

Практическая значимость и ценность академического словарного корпуса /АСК/ толковых словарей русского языка /РЯ/ определяется следующими факторами и потенциальными возможностями: а) возможность в интерактивном режиме составлять, уточнять, тем самым оптимизировать поисковые запросы; б) автоматически/автоматизированно выделять из АСК РЯ наиболее содержательные фрагменты, выборки и проекции, обеспечивающее пользователю возможность быстро просматривать массивы больших объемов лексикографических материалов (в частности, по некоторым зонам и полям словарных статей или их комбинациям); в) использовать установленные гипертекстовые ссылки (связи) между похожими (снятие омонимии и разделение многозначности) словарными статьями и/или их фрагментами, что дает возможность пользователю, найдя одну релевантную лексическую единицу, запросить похожие на неё лексемы; г) возможность реферирования и аннотирования, и редактирования словарных статей АСК РЯ (с фиксацией истории авторского комментирования и модификации) в интерактивном режиме в Интернете.

Академический словарный корпус русского языка, фактически являясь информационно-поисковой лексикографической системой, даст богатый материал для различного рода научных исследований не только в области русского языка, но и других современных языков, что в перспективе позволит на основе цифровых технологий при соответствующей классификации



и систематизации словарей объединить лексикографические материалы, обеспечить их оперативный ввод в научный оборот с целью оптимизации научных исследований в современной лексикографии.

Список использованных источников

- [1] Андрющенко В. М. Машинный фонд русского языка: идеи и суждения / Концепция и архитектура Машинного фонда русского языка. М.: Наука, 1986. С. 32.
- [2] Андрющенко В. М. Концепция и архитектура машинного фонда русского языка: дисс. д.ф.н. М., 1988. 360 с.
- [3] Андрющенко В. М. Концепция и архитектура Машинного фонда русского языка. М.: Наука, 1989. 196 с.
- [4] Андрющенко В. М. Машинный фонд русского языка: Интеграционный подход. М.: ВИНТИ, 1989. 80 с.
- [5] Булыгина Д. С., Лесников С. В. Алгоритм автоматизированного конструирования гипертекстового тезауруса (гизауруса) русского языка на основе оцифрованных словарей и справочников новых слов и значений для интерактивного лексикографического корпуса «Лексико-семантическая неология в русском языке начала XXI века» // Достижения вузовской науки 2019: сборник статей IX Международного научно-исследовательского конкурса (20 мая 2019 г. в г. Пенза). В 2 ч. Ч. 1. Пенза: МЦНС «Наука и Просвещение». 2019. С. 119-125.
- [6] Вторая Всесоюзная конференция по созданию машинного фонда русского языка: Доклады. М.: Б. и., 1987. 248 с.
- [7] Вторая Всесоюзная конференция по созданию машинного фонда русского языка: (Тез. докл.). М.: Б. и., 1987. 182 с.
- [8] Ершов А. П. Методологические предпосылки продуктивного диалога с ЭВМ на естественном языке // Вопросы философии. 1981. № 8. С. 109-119. URL: <https://elibrary.ru/item.asp?id=32667748> (дата обращения: 12.05.2019).
- [9] Загоровская О. В., Лесников С. В. Виды лексикографической информации в автоматическом словаре русских говоров Коми АССР и сопредельных областей // Машинный фонд русского языка: Предпроектные исследования. М.: Институт русского языка /ИРЯз/ АН СССР, 1988. С. 64-70.
- [10] Инструкция для составления «Словаря современного русского литературного языка» (в пятнадцати томах). М.-Л.: АН СССР, 1958. 87 с.
- [11] Караулов Ю. Н. Активная грамматика и ассоциативно-вербальная сеть. М.: ИРЯ РАН, 1999. 180 с.
- [12] Лесников С. В. Автоматизированная информационно-справочная система поиска лексико-семантических вариантов вокабул русского языка // Проблемы истории, филологии, культуры. 2009. № 2(24). С. 622–630. URL: <https://elibrary.ru/item.asp?id=16863201> (дата обращения: 12.05.2019).
- [13] Лесников С. В. Академические толковые словари русского языка как ядро академического словарного корпуса русского языка // Сборник научных статей по итогам работы Международного научного форума «Наука и инновации: современные концепции» (г. Москва, 5 апреля 2019 г.). Ч.1. М.: Инфинити, 2019. С. 38-47.
- [14] Лесников С. В. Академический словарный корпус /АСК/ русского языка // Русский язык: исторические судьбы и современность. МГУ, 2019. С. 213-214.
- [15] Лесников С. В. Академический словарный корпус (АСК) русского языка // Слово и словарь = Vocabulum et vocabularium: сборник научных статей / Нац. акад. Наук Беларуси, Центр исслед. белорус, культуры, языка и лит., Ин-т языкознания имени Якуба Коласа; редкол.: И.Л. Копылов (гл. ред.) Минск: Беларуская навука, 2019. Вып. 16. С. 111-114.
- [16] Лесников С. В. Анализ парадигматических отношений лингвистической терминосистемы // Памяти Анатолия Анатольевича Поликарпова. МГУ, 2015. С. 269-279.
- [17] Лесников С. В. Аналитический реферативно-аннотированный обзор оцифрованных словарей и справочников новых слов и значений для цифрового лексикографического корпуса «Лексико-семантическая неология в русском языке начала XXI века» // Сборник научных статей по итогам работы Международного научного форума «Наука и инновации: современные концепции» (г. Москва, 3 мая 2019 г.). Т.3. М.: Инфинити, 2019. С. 34-42.
- [18] Лесников С. В. Архитектура и суть информационно-поискового корпуса академических словарей русского языка // Научный обозреватель. 2019. № 3(99). С. 25-28.
- [19] Лесников С. В. Базовые блоки автоматизированной лексикографической системы // Вестник Челябинского гос. университета. Серия: филология, искусствоведение. 2011. Вып. 60. № 33(248). С. 200-202. URL: <https://elibrary.ru/item.asp?id=17799038> (дата обращения: 12.05.2019).
- [20] Лесников С. В. Базовые операторы языка поисковых запросов тезауруса метаязыка лингвистики // В мире научных открытий. 2012. № 7-2(31). С. 39-53. URL: <https://elibrary.ru/item.asp?id=17920312> (дата обращения: 12.05.2019).
- [21] Лесников С. В. Владислав Митрофанович Андрющенко - научный руководитель и консультант, главный конструктор Машинного фонда русского языка /МФРЯ/ // Тезисы всероссийской конференции «От языковых машинных фондов к лингвистическим корпусам: памяти В.М. Андрющенко». МГУ, ИРЯз РАН, 2018. С. 58-60.
- [22] Лесников С. В. Гипертекстовый информационно-поисковый тезаурус (гизаурус) «Метаязык науки» (структура; математическое, лингвистическое и программное обеспечения; разделы лингвистика, математика, экономика) // Русский язык: исторические судьбы и современность. МГУ, 2014. С. 268-269.
- [23] Лесников С. В. Гипертекстовый тезаурус метаязыка науки // Проблемы истории, филологии, культуры. 2011. № 3(33). С. 30-34. URL: <https://elibrary.ru/item.asp?id=17072389> (дата обращения: 12.05.2019).
- [24] Лесников С. В. К вопросу о содержании словарной статьи тезауруса метаязыка лингвистики // Обыденное метаязыковое сознание: онтологические и гносеологические аспекты. Ч.IV. Кемерово: КГУ, 2012. С. 190–203.
- [25] Лесников С. В. Конструирование гипертекстового генерального свода лексики русского языка: русский гизаурус / Международная научная конференция «Язык и культура». 14-17 сентября 2001. М.: ОЛЯ РАН, научн.



- журнал «Вопросы филологии», Инст. иностранных языков, 2001. С. 280-281.
- [26] Лесников С. В. Конструирование гипертекстового свода лексики народных говоров русского языка // Актуальные проблемы русской диалектологии. Материалы международной конференции 26-28 октября 2018 г. М.: ИРЯз им. В. В. Виноградова РАН, 2018. С. 148-149.
- [27] Лесников С. В. Конструирование информационно-поискового свода академических словарей русского языка (Свод АСРЯ) // Лексический атлас русских народных говоров (Материалы и исследования) 2018. СПб.: ИЛИ РАН, 2018. С. 226-257.
- [28] Лесников С. В. Конструирование словаря терминов метаязыка СМИ с помощью методики вычисления веса базисных терминов // Слова ў кантэксте часу: матэрыялы IV Міжнар. навук.-практ. канф., прысвеч. 90-годдзю з дня нараджэння д-ра філал. навук праф. А.І. Наркевіча, Мінск, 14–15 сак. 2019 г. Мінск: Беларус. дзярж. ун-т, 2019. С. 66-69. URL: <https://elibrary.ru/item.asp?id=37143924> (дата обращения: 12.05.2019).
- [29] Лесников С. В. Конструирование словника словаря терминов метаязыка лингвистики с помощью методики вычисления веса базисных терминов метаязыка лингвистики // Социально-когнитивное функционирование языка. Кемерово: КГУ, 2017. С.155-170.
- [30] Лесников С. В. Моделирование тезауруса метаязыка лингвистики на базе гипертекстовых фреймов // Вестник Вятского государственного гуманитарного университета. 2011. № 3(2). С. 51–54. URL: <https://elibrary.ru/item.asp?id=17567363> (дата обращения: 12.05.2019).
- [31] Лесников С. В. Основные латинские терминологические элементы и термины метаязыка лингвистики // Научные ведомости БелГУ. Сер. Гуманитарные науки. 2011. № 12(107). Вып.10. С. 37–45. URL: <https://elibrary.ru/item.asp?id=17298245> (дата обращения: 12.05.2019).
- [32] Лесников С. В. Предпосылки конструирования и базовые первоисточники академического словарного корпуса русского языка // Сборник научных статей по итогам работы Международного научного форума «Наука и инновации - современные концепции» (г. Москва, 22 марта 2019 г.). Т.2. М.: Изд-во Инфинити, 2019. С. 76-83.
- [33] Лесников С. В. Словарь русских словарей. М.: Азбуковник, 2002. 334 с.
- [34] Лесников С. В. Типология русских словарей лингвистической терминологии // Мир науки, культуры, образования. 2011. № 6(31), Часть 2. С. 6–10. URL: <https://elibrary.ru/item.asp?id=18155364> (дата обращения: 12.05.2019).
- [35] Лесников С. В. Формирование терминологического фонда русского языка // Наука сегодня: вызовы и решения: материалы международной научно-практической конференции, г. Вологда, 30 января 2019 г. Вологда: ООО «Маркер», 2019. С. 98-102.
- [36] Лесников С. В. Фрагмент словаря базовых терминов метаязыка лингвистики // Лексический атлас русских народных говоров (Материалы и исследования) 2017. СПб., 2017. С. 335-360.
- [37] Лесников С. В. Фреймовое конструирование тезауруса метаязыка лингвистики // Вестник Поморского университета. Сер. Гуманитарные и социальные науки. 2011. № 4. С. 84-88. URL: <https://elibrary.ru/item.asp?id=16996432> (дата обращения: 12.05.2019).
- [38] Лесников С. В., Загоровская О. В. Формальная грамматика словарной статьи автоматического словаря русских говоров Коми АССР и сопредельных областей /АСРГ-КА/ // Вторая Всесоюзная конференция по созданию Машинного фонда русского языка: Материалы конференции. М.: ИРЯз АН СССР, 1988. С. 107-119.
- [39] Лесников С. В., Латкин С. А. Программный модуль «Оптимальная расстановка» объектно-ориентированного пакета прикладных программ // Тезисы десятой Коми республиканской молодежной научной конференции. Сыктывкар: Коми филиал АН СССР, Коми Обком ВЛКСМ, Коми Областной Совет НТО, 1987. С. 134-135.
- [40] Материалы III Всесоюзной конференции по созданию Машинного фонда русского языка / Под ред. С. Ф. Гилязова, Ю. Н. Караулова. МГУ, 1990. 146 с.
- [41] Андрущенко В. М. Машинный фонд русского языка: идеи и суждения. М.: Наука, 1986. 239 с.
- [42] Андрущенко В. М. Машинный фонд русского языка: предпроектные исследования. М.: Б. и., 1988. 294 с.
- [43] Проект словаря современного русского литературного языка. М.-Л.: АН СССР, 1938. 98 с.
- [44] Третья Всесоюзная конференция по созданию машинного фонда русского языка: Тез. докл. Ч.1. М.: Б.и., 1989. 207 с. Ч.2. М.: Б.и., 1989. 158 с.
- [45] Щерба Л. В. Опыт общей теории лексикографии // Изв. АН СССР. Отд. лит-ры и языка. 1940. № 3. (Переизд.: Щерба Л.В. Опыт общей теории лексикографии // Языковая система и речевая деятельность. Л., 1974).
- [46] Bush V. As We May Think (Life Magazine 9-10-1945). 1945. Pp. 112-124.
- [47] Chomsky N. Tree models for the description of language. IRE Trans of Inform. Theory. IT-2. 1956. Pp. 113-124.
- [48] Chomsky N. On the certain formal properties of grammars. Information and Control. 1959. Pp. 137-167.
- [49] Chomsky N. On the notion "rule of grammar" / Structure of language and its mathematical aspects. Providence (Rhode Island), 1961. Pp. 6-24.
- [50] Chomsky N. Formal properties of grammars // Handbook of mathematical psychology / R. D. Luce, R. R. Bush, E. Galanter (eds). New York, NY: John Wiley & Sons, 1963. Pp. 323-418.
- [51] Chomsky N., Miller G.A. Finite state languages. Information and Control. 1958. Pp. 91-112.
- [52] Chomsky N., Schützenberger M. P. The algebraic theory of context-free languages // Computer programming and formal systems / P. Braffort, D. Hirschberg (eds). Amsterdam, 1963. Pp. 118-161.
- [53] Conklin J. Hypertext: An Introduction and Survey // Computer. 1987. Vol. 20, No. 9. Pp. 17-41. DOI: 10.1109/MC.1987.1663693
- [54] Minsky M. Semantic Information Processing. MIT Press, Cambridge, MA, 1968.
- [55] Nelson T. Computer Lib / Dream machines. Sausalito, CA: Mindful Press, 1974.
- [56] Nelson T. Literary machines. Sausalito, CA: Mindful Press, 1993.

Поступила 12.05.2019; принята к публикации 21.06.2019;
опубликована онлайн 25.07.2019.



Об авторах:

Лесников Сергей Владимирович, ведущий специалист (программист) отдела лексикографии современного русского языка, группа Большого академического словаря, Институт лингвистических исследований Российской академии наук (199053, Россия, г. Санкт-Петербург, Тучков пер., д. 9), кандидат филологических наук, доцент,
ORCID: <http://orcid.org/0000-0001-5816-0996>, serg@lsw.ru

Лесников Александр Владимирович, инженер Альманаха «ГОВОР», механико-математический факультет, Московский государственный университет имени М.В. Ломоносова (119991, Россия, г. Москва, ГСП-1, Ленинские горы, д. 1), ORCID: <http://orcid.org/0000-0002-6125-3007>, helpt@mail.ru

Лесников Глеб Сергеевич, модератор Альманаха «ГОВОР», Северный государственный медицинский университет (163000, Россия, г. Архангельск, пр. Троицкий, д. 51), ORCID: <http://orcid.org/0000-0001-6738-0835>, gleb@lsw.ru

Фарина Алена Михайловна, редактор Альманаха «ГОВОР», Сыктывкарский государственный университет имени Питирима Сорокина (167001, Россия, Республика Коми, г. Сыктывкар, Октябрьский пр., д. 55), ORCID: <http://orcid.org/0000-0003-1167-0045>

Все авторы прочитали и одобрили окончательный вариант рукописи.

References

- [1] Andryushchenko V.M. Machine Fund of the Russian language: ideas and judgments. Concept and architecture of the Machine Fund of the Russian language. Nauka, Moscow, 1986; 32. (In Russ.)
- [2] Andryushchenko V.M. The Concept and architecture of the machine Fund of the Russian language: Diss. Dr. Sci. (Philology). Nauka, Moscow, 1988. 360 pp. (In Russ.)
- [3] Andryushchenko V.M. *Kontseptsiya i arkhitektura Mashinogo fonda russkogo yazyka* [Concept and Architecture of the Computer Fund of Russian Language]. Moscow, Nauka Publ., 1989, 196 pp. (In Russ.)
- [4] Andryushchenko V.M. Machine Fund of the Russian language: Integration approach. M.: VINITI, 1989. 80 pp. (In Russ.)
- [5] Bulygina D.S., Lesnikov S.V. Algoritm avtomatizirovannogo konstruirovaniya gipertekstovogo tezaurusa (gizaurusa) russkogo yazyka na osnove otsifrovannykh slovarey i spravochnikov novykh slov i znacheniy dlya interaktivnogo leksikograficheskogo korpusa "Leksiko-semanticheskaya neologiya v russkom yazyke nachala XXI veka" [Algorithm for automated construction of a hypertext thesaurus (hysaurus) of the Russian language based on digitized dictionaries and reference books of new words and meanings for the interactive lexicographic corpus "Lexical and semantic neology in Russian language of the beginning of the 21st century"]. *Sbornik statey 9 Mezhdunarodnogo nauchno-issledovatel'skogo konkursa "Dostizheniya vuzovskoy nauki 2019"*: v 2 ch. [Proceedings of 9th International Scientific and Research Competition "Achievements of the University Science 2019": in 2 pts]. Penza, International Center for Scientific Cooperation "Nauka i Prosveshcheniye" Publ. 2019; 1:19-125. (In Russ.)
- [6] The second All-Union conference on the creation of the machine Fund of the Russian language: Reports. M. B. I., 1987. 248 pp. (In Russ.)
- [7] The second All-Union conference on the creation of the machine Fund of the Russian language: (TEZ. Doc.). M. : B. I., 1987. 182 pp. (In Russ.)
- [8] Ershov A.P. Methodological prerequisites of productive dialogue with computers in natural language. *Voprosy Filosofii*. 1981; 8:109-119. (In Russ.)
- [9] Zagorovskaya O.V., Lesnikov S.V. Vidy leksikograficheskoy informatsii v avtomaticheskome slovare russkikh govorov Komi ASSR i sopredel'nykh oblastey [Types of lexicographic information in the automatic dictionary of Russian patois of the Komi Autonomous Soviet Socialist Republic and adjacent regions]. *Mashinnyy fond russkogo yazyka: Predproyektnyye issledovaniya* [Computer Fund of Russian Language: Pre-design Research]. Moscow, Russian Language Institute of the Academy of Sciences of the USSR Publ. 1988; 64-70. (In Russ.)
- [10] Instructions for compiling the "Dictionary of modern Russian literary language" (in fifteen volumes). M.-L.: USSR Academy of Sciences, 1958. 87 pp. (In Russ.)
- [11] Karaulov Y.N. *Aktivnaya grammatika i assotsiativno-verbal'naya set'* [Active Grammar and Associative-Verbal Network]. Moscow, Russian Language Institute of RAS Publ., 1999. 180 pp. (In Russ.)
- [12] Lesnikov S.V. Computer-based Russian word retrieval reference system. *Problemy istorii, filologii, kul'tury* = Journal of Historical, Philological and Cultural Studies. 2009; 2(24):622- 630. Available at: <https://elibrary.ru/item.asp?id=16863201> (accessed 12.05.2019). (In Russ., abstract in Eng.)
- [13] Lesnikov S.V. Akademicheskiye tolkovyye slovari russkogo yazyka kak yadro akademicheskogo slovarnogo korpusa russkogo yazyka [Academic explanatory dictionaries of the Russian language as the core of the academic vocabulary of the Russian language]. *Sbornik nauchnykh statey po itogam raboty Mezhdunarodnogo nauchnogo foruma "Nauka i innovatsii: sovremennyye kontseptsii"* [Proceedings of the International Scientific Forum "Science and Innovations: Contemporary Concepts"]. Moscow, Infinity Publ. 2019; 1:38-47. (In Russ.)
- [14] Lesnikov S.V. Akademicheskiy slovarnyy korpus (ASK) russkogo yazyka [Academic vocabulary corpus (AVC) of the Russian language]. *Materialy 6 Mezhdunarodnogo kongressa issledovateley russkogo yazyka "Russkiy yazyk: istoricheskiye sud'by i sovremennost"* [Proceedings of the 6th International Congress of Russian Language Researchers "Russian Language: Historical Fate and Modernity"]. Moscow, Lomonosov Moscow State University Publ. 2019; 213-214. (In Russ.)
- [15] Lesnikov S.V. Akademicheskiy slovarnyy korpus (ASK) russkogo yazyka [Academic vocabulary corpus (AVC) of the Russian language]. *Slovo i slovar'* = Vocabulum et vocabularium [Word and Dictionary = Vocabulum et vocabularium]. 2019; 16:111-114. (In Russ.)
- [16] Lesnikov S.V. Analiz paradigmaticeskikh otnosheniy lingvisticheskoy terminosistemy [Analysis of the paradigmatic relations of the linguistic term system]. *Pamyati Anatoliya Anatol'yevicha Polikarpova* [In Memory of Anatolya Anatol'yevicha Polikarpova]



- ly Anatolyevich Polikarpov]. Moscow, Lomonosov Moscow State University Publ. 2015; 269-279. (In Russ.)
- [17] Lesnikov S.V. Analiticheskiy referativno-annotirovannyi obzor otsifrovannykh slovarey i spravochnikov novykh slov i znacheniy dlya tsifrovogo leksikograficheskogo korpusa «Leksiko-semanticheskaya neologiya v russkom yazyke nachala XXI veka [Analytical abstract annotated review of digitized dictionaries and reference books of new words and meanings for the digital lexicographic corpus "Lexical and semantic neology in Russian language in the beginning of the 21st century]. *Sbornik nauchnykh statey po itogam raboty Mezhdunarodnogo nauchnogo foruma "Nauka i innovatsii: sovremennyye kontseptsii"* [Proceedings of the International Scientific Forum "Science and Innovations: Contemporary Concepts"]. Moscow, Infinity Publ. 2019; 3:34-42. (In Russ.)
- [18] Lesnikov S.V. Arkhitektura i suť informatsionno-poiskovogo korpusa akademicheskikh slovarey russkogo yazyka [Architecture and essence of the information-retrieval corpus of academic dictionaries of the Russian language]. *Nauchnyy obozrevatel' = Scientific Reviewer*. 2019; 3(99):25-28. (In Russ.)
- [19] Lesnikov S.V. Basic blocks of an automated lexicographic system. *Vestnik Chelyabinskogo gosudarstvennogo universiteta = Bulletin of Chelyabinsk State University*. 2011; 33(248):200-202. Available at: <https://elibrary.ru/item.asp?id=17799038> (accessed 12.05.2019). (In Russ.)
- [20] Lesnikov S.V. The basic operators of the search queries thesaurus metalanguage of linguistics. *V mire nauchnykh otkrytiy = Siberian Journal of Life Sciences and Agriculture*. 2012; 7-2(31):39-53. Available at: <https://elibrary.ru/item.asp?id=17920312> (accessed 12.05.2019). (In Russ., abstract in Eng.)
- [21] Lesnikov S.V. Vladislav Mitrofanovich Andryushchenko – nauchnyy rukovoditel' i konsul'tant, glavnyy konstruktor Mashinnogo fonda russkogo yazyka (MFRYA) [Vladislav Mitrofanovich Andryushchenko – scientific advisor and consultant, chief designer of the Computer Fund of Russian language (CFRL)]. *Tezisy vserossiyskoy konferentsii "Ot yazykovykh mashinnykh fondov k lingvisticheskim korpusam: pamyati V.M. Andryushchenko"* [Proceedings of the All-Russian Conference "From Linguistic Computer Funds to Linguistic Corpus: to Memory of V.M. Andryushchenko"]. Moscow, Lomonosov Moscow State University Publ., Russian Language Institute of RAS Publ. 2018; 58-60. (In Russ.)
- [22] Lesnikov S.V. Hypertext information retrieval thesaurus (hesaurus) "a meta-language of science" (structure; mathematical, linguistic and software; topics linguistics, mathematics, Economics). Russian language: its historical destiny and present state. Moscow State University. 2014. 268-269. (In Russ.)
- [23] Lesnikov S.V. Hypertext thesaurus of science metalanguage. *Problemy istorii, filologii, kul'tury = Journal of Historical, Philological and Cultural Studies*. 2011; 3(33):C. 30-34. Available at: <https://elibrary.ru/item.asp?id=17072389> (accessed 12.05.2019). (In Russ., abstract in Eng.)
- [24] Lesnikov S.V. K voprosu o sodержanii slovarnoy stat'i tezaurusa metazyazyka lingvistiki [To the issue of dictionary entry content of the metalanguage thesaurus of linguistics]. *Obydennoye metazyazykovoye soznaniye: ontologicheskiye i gnoseologicheskiye aspekty* [Ordinary Metalinguistic Consciousness: Ontological and Epistemological Aspects]. Kemerovo, Kemerovo State University Publ. 2012; 4:190-203. (In Russ.)
- [25] Lesnikov S.V. Gipertekstovyy informatsionno-poiskovyy tezaurus (gizaurus) "Metazyazyk nauki" (struktura; matematicheskoye, lingvisticheskoye i programmnoye obespecheniya; razdely lingvistika, matematika, ekonomika) [Hypertext information retrieval thesaurus (hysaurus) "Metalanguage of science" (structure; mathematical, linguistic and software provision; sections of linguistics, mathematics, economics)]. *Materialy 5 Mezhdunarodnogo kongressa issledovateley russkogo yazyka "Russkiy yazyk: istoricheskiye sud'by i sovremennost'"* [Proceedings of the 5th International Congress of Russian Language Researchers "Russian Language: Historical Fate and Modernity"]. Moscow, Lomonosov Moscow State University Publ. 2014; 268-269. (In Russ.)
- [26] Lesnikov S.V. Konstruirovaniye gipertekstovogo svoda leksiki narodnykh govorov russkogo yazyka [Constructing a hypertext corpus of folk patois vocabulary of the Russian language]. *Materialy mezhdunarodnoy konferentsii "Aktual'nyye problemy russkoy dialektologii"* [Proceedings of the International Conference "Current Issues of Russian Defectology"]. Moscow, V.V. Vinogradov Russian Language Institute of RAS Publ. 2018; 148-149. (In Russ.)
- [27] Lesnikov S.V. Konstruirovaniye informatsionno-poiskovogo svoda akademicheskikh slovarey russkogo yazyka (Svod ASRYA) [Constructing an information retrieval corpus of academic dictionaries of the Russian language (ADRL corpus)]. *Leksicheskiy atlas russkikh narodnykh govorov (Materialy i issledovaniya)* [Lexical Atlas of Russian National Patois (Materials and Studies)]. St. Petersburg, Institute of Linguistic Studies of the Russian Academy of Sciences Publ. 2018; 226-257. (In Russ.)
- [28] Lesnikov S.V. Konstruirovaniye slovyara terminov metazyazyka SMI s pomoshch'yu metodiki vychisleniya vesa bazisnykh terminov [Constructing a terms dictionary of mass media metalanguage using the method of calculating the weight of basic terms]. *Materyaly 4 Mizhnar. navuk-prakt. kanf, prysvyech. 90-hoddzyu z dnya narodzhennya d-ra filal. navuk praf. A.I. Narkyevicha "Slova w kantekstsye chasu"* [Proceedings of the 4th International Scientific and Practical Conference Dedicated to the 90th Anniversary of Doctor of Philology, Professor A.I. Narkevich "Word in a Context of Time"]. Minsk, Belarusian State University Publ., 2019, pp. 66-69. Available at: <https://elibrary.ru/item.asp?id=37143924> (accessed 12.05.2019). (In Russ.)
- [29] Lesnikov S.V. Konstruirovaniye slovnika slovyara terminov metazyazyka lingvistiki s pomoshch'yu metodiki vychisleniya vesa bazisnykh terminov metazyazyka lingvistiki [Constructing a glossary of terms vocabulary of linguistics metalanguage using the method of calculating the weight of the basic terms of linguistics metalanguage]. *Sotsial'no-kognitivnoye funktsionirovaniye yazyka* [Social and Cognitive Functioning of the Language]. Kemerovo, Kemerovo State University Publ. 2017; 155-170. (In Russ.)
- [30] Lesnikov S.V. Modeling the metalanguage thesaurus of linguistics-based hypertext frames. *Vestnik Vyatskogo gosudarstvennogo gumanitarnogo universiteta = Herald of Vyatka State University*. 2011; 3(2):51-54. Available at: <https://elibrary.ru/item.asp?id=17567363> (accessed 12.05.2019).



- (In Russ., abstract in Eng.)
- [31] Lesnikov S.V. Key terms and Latin term elements of linguistics metalanguage. *Nauchnyye vedomosti Belgorodskogo gosudarstvennogo universiteta. Seriya: Gumanitarnyye nauki* = Belgorod State University Scientific Bulletin. Humanities. 2011; 12(107):37-45. Available at: <https://elibrary.ru/item.asp?id=17298245> (accessed 12.05.2019). (In Russ., abstract in Eng.)
- [32] Lesnikov S.V. Predposylki konstruirovaniya i bazovyye pervoistochniki akademicheskogo slovnogo korpusa russkogo yazyka [Prerequisites for the construction and basic sources of the academic vocabulary of the Russian language]. *Sbornik nauchnykh statey po itogam raboty Mezhdunarodnogo foruma "Nauka i innovatsii: sovremennyye kontseptsii"* [Proceedings of the International Scientific Forum "Science and Innovations: Contemporary Concepts"]. Moscow, Infinity Publ. 2019; 2:76-83. (In Russ.)
- [33] Lesnikov S.V. *Slovar' russkikh slovarey* [Dictionary of the Russian Dictionaries]. Moscow, Azbukovnik Publ., 2002. 334 pp. (In Russ.)
- [34] Lesnikov S.V. Typology of Russian dictionaries of linguistic terminology. *Mir nauki, kul'tury, obrazovaniya* = The World of Science, Culture and Education. 2011; 6-2(31):6-10. Available at: <https://elibrary.ru/item.asp?id=18155364> (accessed 12.05.2019). (In Russ., abstract in Eng.)
- [35] Lesnikov S.V. Formirovaniye terminologicheskogo fonda russkogo yazyka [Terms fund of the Russian language development]. *Materialy mezhdunarodnoy nauchno-prakticheskoy konferentsii "Nauka segodnya: vyzovy i resheniya"* [Proceedings of the International Scientific and Practical Conference "Science Today: Challenges and Solutions"]. Vologda, LLC "Marker" Publ. 2019; 98-102. (In Russ.)
- [36] Lesnikov S.V. Fragment slovary bazovykh terminov meytazyka lingvistiki [Fragment of the dictionary of basic terms of linguistics metalanguage]. *Leksicheskiy atlas russkikh narodnykh govorov* (Materialy i issledovaniya) [Lexical Atlas of Russian National Patois (Materials and Studies)]. St. Petersburg, 2017; 335-360. (In Russ.)
- [37] Lesnikov S.V. Frame construction of thesaurus of linguistics metalanguage. *Vestnik Pomorskogo universiteta. Seriya: Gumanitarnyye i sotsial'nyye nauki* = Vestnik of Northern (Arctic) Federal University. Series "Humanitarian and Social Sciences". 2011; 4:84-88. Available at: <https://elibrary.ru/item.asp?id=16996432> (accessed 12.05.2019). (In Russ., abstract in Eng.)
- [38] Lesnikov S.V., Zagorovskaya O.V. Formal'naya grammatika slovarnoy stat'i avtomaticheskogo slovary russkikh govorov Komi ASSR i sopredel'nykh oblastey (ASRGKA) [Formal grammar of dictionary entry of automatic dictionary of Russian patois of the Komi Autonomous Soviet Socialist Republic and adjacent regions (ADRPKA)]. *Materialy 2 Vsesoyuznoy konferentsii po sozdaniyu Mashinnogo fonda russkogo yazyka* [Proceedings of the 2nd All-Russian Conference on Development of Computer Fund of Russian language]. Moscow, Russian Language Institute of the Academy of Sciences of the USSR Publ. 1988; 107-119. (In Russ.)
- [39] Lesnikov S.V., Latkin S.A. Program module "Optimal arrangement" of object-oriented package of applications. *Tezisy desyatoy Komi respublikanskoj molodezhnoj nauchnoj konferentsii* [Theses of the tenth Komi Republican youth scientific conference]. Syktyvkar: Komi branch of the USSR Academy of Sciences, Komi regional Komsomol Committee, Komi Regional Council NTO. 1987; 134-135. (In Russ.)
- [40] Materials of III All-Union Conference on the establishment of the Machine Fund of Russian language. S.F. Gilyazova, Yu.N. Karaulov (eds). Moscow State University, 1990. 146 pp. (In Russ.)
- [41] Andryushchenko V.M. Mashinnyy fond russkogo yazyka: idei i suzheniya [Computer Fund of Russian Language: Ideas and Judgments]. *Kontseptsiya i arkhitektura Mashinnogo fonda russkogo yazyka* [Concept and Architecture of the Computer Fund of Russian Language]. Moscow, Nauka Publ. 1986; 26-44. (In Russ.)
- [42] Machine Fund of the Russian language: pre-project studies. M.: B. I., 1988. 294 pp. (In Russ.)
- [43] Project dictionary of modern Russian literary language. M.-L.: USSR Academy of Sciences, 1938. 98 pp. (In Russ.)
- [44] The third All-Union conference on the creation of the machine Fund of the Russian language: TEZ. Doc. Part 1. Moscow: B. I., 1989. 207 pp. V. 2. Moscow: B. I., 1989. 158 pp. (In Russ.)
- [45] Shcherba L.V. Opyt obshchey teorii leksikografii [Experience of the general lexicography theory]. *Izvestiya AN SSSR. Otdelenie literatury i yazyka* [News of Academy of Sciences of the USSR. Language and Literature Department]. 1940, no. 3. (In Russ.) (Revised ed.: Shcherba L.V. Opyt obshchey teorii leksikografii [Experience of the general lexicography theory]. *Yazykovaya sistema i rechevaya deyatel'nost'* [Linguistic System and Speech Activity]. Leningrad, 1974. (In Russ.))
- [46] Bush V. As We May Think (Life Magazine 9-10-1945). 1945; 112-124. (In Eng.)
- [47] Chomsky N. Tree models for the description of language. *IRE Trans of Inform. Theory*. IT-2. 1956; 113-124. (In Eng.)
- [48] Chomsky N. On the certain formal properties of grammars. *Information and Control*. 1959; 137-167. (In Eng.)
- [49] Chomsky N. On the notion "rule of grammar". *Structure of language and its mathematical aspects*. Providence (Rhode Island). 1961; 6-24. (In Eng.)
- [50] Chomsky N. Formal properties of grammars. In: Luce R.D., Bush R.R., Galanter E. (eds). *Handbook of mathematical psychology*. New York, NY: John Wiley & Sons. 1963; 323-418. (In Eng.)
- [51] Chomsky N., Miller G.A. Finite state languages. *Information and Control*. 1958; 91-112. (In Eng.)
- [52] Chomsky N., Schützenberger M. P. The algebraic theory of context-free languages. In: Braffort P., Hirschberg D. (eds). *Computer programming and formal systems*. Amsterdam. 1963; 118-161. (In Eng.)
- [53] Conklin J. Hypertext: An Introduction and Survey. *Computer*. 1987; 20(9):17-41. (In Eng.) DOI: 10.1109/MC.1987.1663693
- [54] Minsky M. *Semantic Information Processing*. MIT Press, Cambridge, MA, 1968. (In Eng.)
- [55] Nelson T. *Computer Lib / Dream machines*. Sausalito, CA: Mindful Press, 1974. (In Eng.)
- [56] Nelson T. *Literary machines*. Sausalito, CA: Mindful Press, 1993. (In Eng.)

Submitted 12.05.2019; revised 21.06.2019;
published online 25.07.2019.



About the authors:

Sergey V. Lesnikov, Leading Specialist (IT Specialist) of Contemporary Russian Language Lexicography Department, Great Academic Dictionary Group, Institute for Linguistic Studies of the Russian Academy of Sciences (9 Tuchkov Ln., St. Petersburg 199053, Russia), Ph.D. (Philology), Associate Professor, ORCID: <http://orcid.org/0000-0001-5816-0996>, serg@lsw.ru

Alexander V. Lesnikov, engineer of the Almanah "GOVOR", Faculty of Mechanics and Mathematics, Lomonosov Moscow State University (1, Leninskie gory, Moscow 119991, Russia), ORCID: <http://orcid.org/0000-0002-6125-3007>, helpt@mail.ru

Gleb S. Lesnikov, moderator of the Almanah "GOVOR", Northern State Medical University (51 Troitskiy Ave., 163000, Arkhangelsk, Russia), ORCID: <http://orcid.org/0000-0001-6738-0835>, gleb@lsw.ru

Alena M. Farina, editor of the Almanah "GOVOR", Pitirim Sorokin Syktyvkar State University (55 Oktyabrsky Ave., Syktyvkar 167001, Komi Republic, Russia), ORCID: <http://orcid.org/0000-0003-1167-0045>

All authors have read and approved the final manuscript.

