

Национальный корпус русского языка

Корпус и изучение истории слов

Светлана Олеговна Савчук, Институт русского языка им. В. В. Виноградова РАН
(Россия, Москва), savsvetlana@mail.ru

DOI:10.31857/S013161170004468-3

Аннотация: Национальный корпус русского языка имеет большой объем и включает тексты разных исторических эпох. Поэтому на материале корпуса можно проводить микродиахронические исследования, анализируя контексты употребления слов и фиксируя изменение их значений.

Большой массив текстов и лингвистическая разметка, которая обеспечивает быстрый поиск слов и форм, делает корпус незаменимым инструментом лексикографа, поскольку предоставляет в его распоряжение большое количество примеров и позволяет более точно описать оттенки значения слов и правила их употребления. Однако эта задача становится совсем непростой, если примеров слишком много. В этом случае на помощь приходят полезные инструменты корпуса, которые значительно облегчают анализ контекстов и извлечение нужной информации. В заметке на конкретных примерах показано использование двух режимов представления контекстов — обычный режим и режим KWIC (Key Word in Context). Подробно описывается также инструмент «Настройки», который позволяет вручную устанавливать порядок следования примеров, формат выдачи, количество документов на странице, количество примеров на странице выдачи, количество слов в контексте KWIC-выдачи, что значительно облегчает и ускоряет просмотр примеров.

Ключевые слова: Национальный корпус русского языка (НКРЯ), функциональные возможности, корпусной инструментарий, средства поиска

для цитирования: Савчук С. О. Корпус и изучение истории слов // Русская речь. 2019. № 2. С. 58–68. DOI:10.31857/S013161170004468-3.

Russian National Corpus

Using Corpus to Study History of Words

Svetlana O. Savchuk, Vinogradov Russian Language Institute (Russian Academy of Sciences)
(Russia, Moscow), savsverlana@mail.ru

ABSTRACT: The Russian National Corpus has a large volume and includes texts from various historical periods. Therefore, based on the corpus data one can perform microdiachronic studies by analyzing contexts of words and registering changes in their meanings. Thanks to its great volume and to linguistic markup that makes it possible to quickly find words and word forms, the corpus becomes an indispensable lexicographer's tool by providing access to numerous examples that help better describe the shades of the words' meanings and the rules for their use. Still, this task may become too complex when there are too many examples in the corpus. There are useful tools available in the corpus that can greatly help researchers to analyze contexts and extract the necessary information in such cases. The article illustrates (with specific examples) the use of two formats of context representation: the ordinary format and the KWIC format (Key Word in Context). The "Settings" menu is described in detail, including such settings as order of examples, output format, number of documents per page, number of examples per document, number of words in KWIC context. These settings can greatly facilitate the process of browsing through examples.

KEYWORDS: The Russian National Corpus, functional capabilities, corpus tools, search tools

FOR CITATION: Savchuk S. O. Using Corpus to Study History of Words. Russian Speech = Russkaya Rech'. 2019. No. 2. Pp. 58–68. DOI: 10.31857/S013161170004468-3.

Н

ациональный корпус русского языка имеет большой объем и включает тексты разных исторических эпох. Поэтому на материале корпуса можно проводить микродиахронические исследования, анализируя контексты употребления слов и фиксируя изменение их значений — расширение, сужение, появление новых значений у слова или исчезновение старых.

Большой массив текстов и лингвистическая разметка, обеспечивающая быстрый поиск слов и форм, делает корпус незаменимым инструментом лексикографа, поскольку предоставляет в его распоряжение большое количество примеров и позволяет более точно описать оттенки значения слов и правила их употребления. Однако эта задача становится совсем не-простой, если примеров слишком много — сотни или даже тысячи. В этом случае на помощь приходят полезные инструменты, которые значительно облегчают анализ контекстов и извлечение нужной информации. Некоторые из них мы покажем в этой заметке на конкретных примерах.

Как вы думаете, когда в русском языке появилось слово **виртуальный**? И в каком контексте оно чаще всего употребляется? Это, безусловно, заимствованное слово, но относится ли оно к новым заимствованиям? Если мы воспользуемся поиском Яндекса, то получим список подсказок, построенный на наиболее частотных запросах: *виртуальная школа, виртуальный помощник, виртуальная клавиатура, виртуальная реальность, виртуальная машина, виртуальный номер телефона* — все это употребления, отражающие реалии сегодняшнего дня, связанные с использованием интернета.

Какую информацию о слове можно извлечь из НКРЯ?

Построим запрос в основном корпусе НКРЯ. На странице лексико-грамматического поиска в поле «Слово» записываем ВИРТУАЛЬНЫЙ. В результате такого запроса получаем около двух тысяч примеров. Как сориентироваться в таком большом массиве?

Просматривать результаты выдачи удобно в формате KWIC (Key Word in Context, ‘ключевое слово в контекстном окружении’). Это наиболее распространенный формат представления конкордансов, при котором от каждого найденного примера показывается лишь одна строчка и иско-мое слово помещается в ее середину (рис. 1).

Режим KWIC устанавливается на верхней панели страницы выдачи. Рядом находится ссылка «Настройки», которая позволяет установить по-рядок следования примеров в режиме просмотра (рис. 2).

При нажатии на ссылку «Настройки» открывается диалоговое окно, кото-рое позволяет самостоятельно установить 1) режим упорядочивания при-меров, 2) формат выдачи (обычный или KWIC), 3) количество слов в кон-

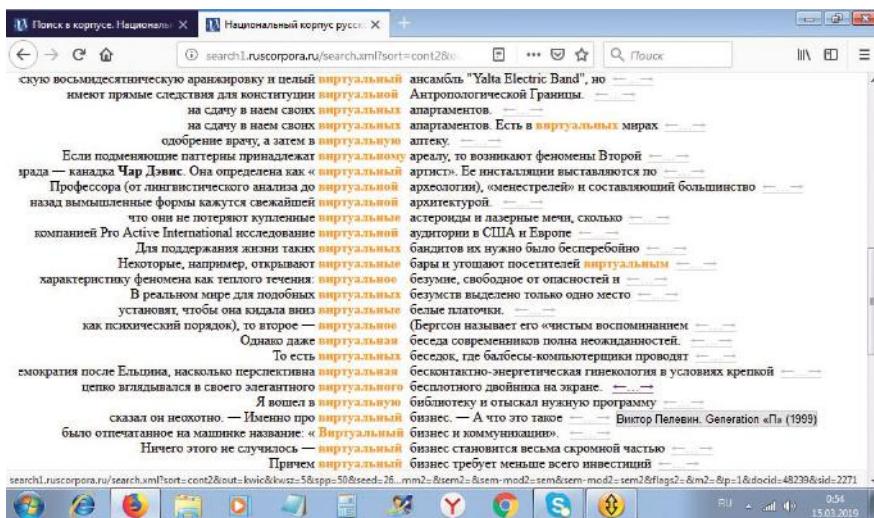


Рис. 1. Выдача примеров в формате KWIC | Fig. 1. Output of examples in KWIC format



Рис. 2. Выбор режима просмотра и сортировки найденных примеров | Fig. 2. Selection of the output format and sorting settings for the examples

тексте KWIC-выдачи (по умолчанию длина контекста — 5 слов), 4) количество примеров на странице (по умолчанию установлено 50 примеров).

По умолчанию примеры в Основном корпусе выдаются в следующем порядке: сначала все примеры из текстов со снятой омонимией, затем все примеры с неснятой омонимией, отсортированные, начиная с самых новых.

В зависимости от решаемых задач пользователь может изменить порядок выдачи примеров, выбрав соответствующую строку в диалоговом окне настройки. Возможны следующие варианты упорядочения:

- по **правому контексту**, то есть по слову, ближайшему к искомому;
- по **правому контексту**, с учетом формы исходного слова, — в отличие от предыдущей настройки, в алфавитном порядке упорядочиваются словоформы искомого слова;

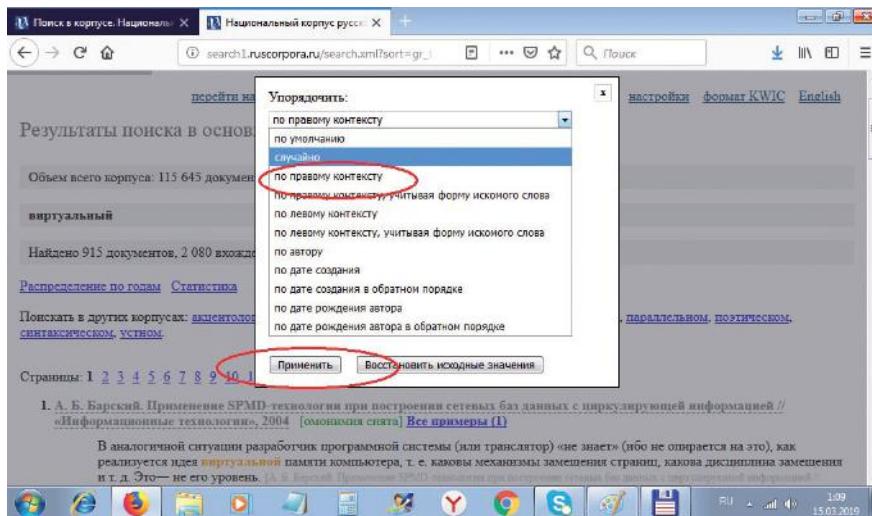


Рис. 3. Выбор настроек для установления порядка следования примеров

Fig. 3. Selection of the order settings for the examples

- **по левому контексту**, то есть по слову, которое предшествует исключому;
- **по левому контексту**, с учетом формы исходного слова, — то же, что в предыдущей настройке, но в алфавитном порядке упорядочиваются словоформы исключенного слова;
- **по автору**, то есть контексты упорядочиваются в соответствии с прямым алфавитным порядком фамилий их авторов;
- **по дате создания**, при такой настройке контексты отсортированы, начиная с самых ранних;
- **по дате создания** в обратном порядке, при такой настройке контексты отсортированы, начиная с самых новых;
- **по дате рождения автора** — контексты отсортированы в соответствии с датой рождения автора, начиная с самой ранней;
- **по дате рождения автора** в обратном порядке — то же самое, но начиная с самой поздней;
- **случайно** — этот режим выбирается, когда пользователю надо получить статистику или сделать какие-то иные выводы на основе случайной выборки.

Для того чтобы быстро просмотреть контексты употребления слова *виртуальный*, используем настройку «по правому контексту» (рис. 3) и получим отображение примеров в виде, представленном на рис. 1. Та-

кой порядок делает наглядной сочетаемость прилагательного *виртуальный*, поскольку рядом оказываются примеры, в которых прилагательное сочетается с формами одного и того же существительного: *виртуальный автор*, *виртуального автора*, *виртуальные авторы*, ... *виртуальный бизнес*, *виртуального бизнеса*, *виртуальному бизнесу* и т. д. При этом сведения о тексте, из которого взят пример, можно получить во всплывающем окне, если навести курсор мыши на две стрелки в конце строки: ←...→ (рис. 1). При нажатии на эти стрелки в отдельном окне откроется расширенный контекст употребления слова (чаще всего он равен абзацу).

Рождество должно помочь мировому виртуальному бизнесу. Конец года может принести для мирового Интернета большую сенсацию — лидер мировой интернет-торговли компания Amazon.com заявила, что, возможно, получит прибыль в четвертом квартале уходящего года. Интрига заключается в том, что этот виртуальный магазин никогда не был прибыльным. Если событие случится, то вся мировая интернет-торговля будет реабилитирована (Известия. 25.11.2001).

Просмотр контекстов показывает, что активность слова *виртуальный* связана прежде всего с развитием информационных технологий и интернета. Самые частотные сочетания: *виртуальная реальность* (134), *виртуальный мир* (89), *виртуальное пространство* (69), *виртуальное предприятие* (19), *виртуальная экономика* (20), *виртуальный магазин* (22), *виртуальное общение* (19), *виртуальный образ* (17), *виртуальная жизнь* (13), *виртуальная игра* (8) и под.

Эти новые употребления слова зафиксированы в словаре «Новые слова и значения» (далее — НСЗ), см. [Буцева (ред.) 2009].

ВИРТУАЛЬНЫЙ, ая, ое. 1. Имитируемый с помощью компьютерной техники, интернет-технологий. Виртуальное пространство (чего). То же, что веб-пространство (см.); интернет-пространство (см.). <...> Виртуальная реальность, действительность. Среда, ситуация, имитируемая с помощью компьютерных устройств (звуком, зрительными образами, тактильными ощущениями и др.), интернет-пространство (см.). <...> Виртуальная библиотека. Совокупность текстовой, музыкальной, графической информации на электронных носителях, интернет-сайтах (см.), доступных с помощью компьютерной техники. <...>

2. Такой, которого нет, не может быть фактически, физически. Президент-сомнамбула... ведет страну в виртуальное «светлое демократическое будущее» и как виртуальный Главком руководит призрачной виртуальной армией. СР 31.12.98. [Буцева (ред.) 2009: 309–312].

По данным корпуса, рост частотности этого слова наблюдается с конца 1990-х годов, что подтверждается и данными словаря НСЗ. Уточним эти наблюдения, используя функцию «Настройки», только на этот раз упоря-

дочим примеры по дате создания. И тут нас ждет неожиданность: оказывается, вопреки нашим предположениям, слово это было известно еще в XIX веке.

*Задача человечества состоит не в чем другом, как в проявлении, в разные времена и разными племенами, всех тех сторон, всех тех особенностей направления, которые лежат **виртуально** (в возможности, *in potentia*) в идее человечества* (Н. Данилевский. Россия и Европа. 1869).

В XX в. слово использовалось как термин в физике, психологии, философии.

*Другой участник событий при этом — физический вакуум с его **виртуальными** частицами* (Техника — молодежи. 1977); *Виртуальная* пара может поглотить фотон и перейти в возбужденное состояние, после испускания фотона она переходит в прежнее невозбужденное состояние (Техника — молодежи. 1990); *Например, процессы зрительного восприятия как бы записаны в особенностях устройства зрительной системы человека, но только в **виртуальной** форме — как их возможность* (А. Леонтьев. Деятельность. Сознание. Личность. 1974); *Она будет «данным», только если примет **виртуальную** форму моего пространства преобразования (или будет, как выражается Декарт в юношеской латинской записи, «сенсибилией, годной для олимпийских познаний»)* (М. Мамардашили. Картезианские размышления. 1981–1993).

Это первоначальное терминологическое значение зафиксировано в словарях — Большой энциклопедии С. Н. Южакова [Южаков (ред.) 1901], словаре Д. Н. Ушакова [Ушаков (ред.) 1935], Большом академическом словаре (БАС), словарях иностранных слов:

ВИРТУАЛЬНЫЙ, -ая, -ое; -лен, -льна, -о. 1. Способный к действию; возможный. □ *Матем.* Виртуальное перемещение — перемещение материальной точки или системы, которое ей в данный момент доступно при существующих связях. 2. □ *Филос.* Не существующий непосредственно и сам по себе, но способный возникнуть при наличии известных условий [БАС: 387]. Согласно БАСу, слово восходит к новолатинскому слову *vīrtuālis* — сильный, способный, мощный.

В конце XX в. к имеющимся значениям добавилось еще одно специальное значение, относящееся к сфере информатики и представленное в таких сочетаниях, как *виртуальная память*, *машина*, *виртуальное пространство*, *хранилище* и под.

*Кроме соответствия исходных кодов блок подкачки также проверяет идентичность разбиения на страницы **виртуальной** памяти для обоих регионов* (Информационные технологии. 29.09.2003); *Недаром в научном мире, инженерной практике, в среде разработчиков и пользователей средств информатики все чаще употребляются такие понятия, как*

интеллектуальный потенциал, интеллектуальная среда, **виртуальное** информационно-интеллектуальное пространство, мировой мозг, ноосфера, psi-фактор и т. п. (Информационные технологии. 18.08.2003); Этим способом публикуются, редактируются и удаляются **виртуальные** объекты, относящиеся как к контенту, так и к управлению функциональными возможностями и дизайном сайта (Информационные технологии. 21.07.2003).

В отличие от общенаучного термина *виртуальный*, который восходит к латыни, источником компьютерного термина стал английский язык, в котором в эпоху компьютеризации сформировались сочетания *virtual memory*, *virtual reality*, заимствованные в 1990-е годы в русский и другие языки [Комлев 1995; Маринова 2008, 354]. В начале XXI в. слово *виртуальный* стало выходить за пределы профессиональной речи [Геккина 2000; Маринова 2008, 355], что было отмечено в словарях новых слов, в частности, в упомянутом выше НСЗ, а также в словаре Г. Н. Скляревской [Скляревская (ред.) 2006] («иллюзорный, умозрительный, не существующий в реальности»).

Однако анализ данных НКРЯ обнаруживает, что непрофессиональное значение слова *виртуальный*, синонимичное словам *ирреальный*, *сюрреальный*, *вымышенный*, *искусственный*, *воображаемый*, *умозрительный*, встречается в текстах и до начала XXI в. Он является продуктом эры новой *виртуальной* политической реальности, рожденной средствами масовой информации, в которой важно не столько быть, сколько казаться (Коммерсантъ-Daily. 23.01.1996); Я согласен жить куда менее интересно, чем тогда, лишь бы ЖИЗНЬ БЫЛА, а не ее *виртуальный* коммунистический муляж со всеми этими «стройками», «гэбухами», «починами», «самиздатом», пьянством, бравадой, страхом и стыдом (Е. Попов. Подлинная история «Зеленых музыкантов». 1997); Становящееся затрапанным слово «*виртуальный*» точно характеризует нашу реальную нереальность (А. Вознесенский. На виртуальном ветру. 1998). У В. Аксенова встречаем даже вариант, восходящий к языку-источнику: Теперь в тишине долины Шэнандоа этот десятимесячный «русский фильм», словно «*виртуал рэлити*», включался в сознание Палмер абсурдно перемешанными кусками, то по ночам на подушке, то за рулем «Тойоты», то в супермаркете, то во время бега, то перед телевизором... (В. Аксенов. Негатив положительного героя. 1996).

Таким образом, в случае со словом *виртуальный* в современном значении мы имеем дело не с семантическим развитием философского термина, а с вторичным заимствованием, которое вошло в русский язык с распространением интернета и освоением большого пласта лексики, с ней связанного (как, например, слово *программа*, *вирус* и т. п.).

Рассмотрим другой пример — слово **забугорный**. Как вы можете охарактеризовать значение и употребление этого слова? С чем оно ассоциируется? Кажется, что с жаргонным выражением «за бугром», то есть ‘за границей’, и появилось это значение в период 1980–1990 гг. Проверим свое предположение по словарям и корпусу.

Толковый словарь В. И. Даля (1863–1866) дает прямое значение: **ЗАБУГОРНЫЙ**, находящийся позади бугра.

Новое переносное значение зафиксировано в Толковом словаре Т. Ф. Ефремовой [Ефремова 2000].

ЗАБУГОРНЫЙ. Забугорный прил. разг. Заграничный, иностранный; зарубежный.

НСЗ приводит производные существительные: **ЗАБУГОРЬЕ**, -я, ср. со-бир. Зарубежные страны; зарубежье (*разг.-шутл. или пренебр.*). **ЗАБУГОРНИК**, -а, м. Иностранец (обычно из стран Запада, Америки) (*разг., пренебр.*) [Буцева (ред.) 2009: 587–588].

Примеры из НКРЯ (1984–2000-е гг.) иллюстрируют это значение: *Сначала мы вместе пришли к радиоприемнику, жадно ловя, как в еще недавние времена глушилок и психушек, забугорные последние известия* (Е. Евтушенко. «Волчий паспорт». 1999); *Поэтому я без энтузиазма выслушал информацию по забугорному радио о наших муках* (В. Попов. Будни гарема. 1994); *Не удосужились забугорные психологи ответить на вопрос: «Кому на Руси жить хорошо?»* (И. Охлобыстин. Мозг Ихтиандра // «Столица». 15.04.1997).

Специалист по современному жаргону как будто бы подтверждает нашу гипотезу о происхождении слова: «Уже прочно вошли в общенародный язык и обросли различными языковыми связями арготизмы первой волны “перестройки” (конца 1980-х – начала 1990-х годов) — такие слова, как отмывание денег (то есть их легализация), бугор — ‘заграница’, свинтить, свалить за бугор — ‘ухать за границу’, забугорный — ‘заграничный’ и пр.» (М. Грачёв. Интервенция криминального языка // «Наука и жизнь». 2009).

Однако вопреки авторитетному мнению, наша гипотеза оказывается ошибочной. Если упорядочить найденные примеры по дате создания, то НКРЯ отошлет нас к произведению В. В. Крестовского «Петербургские трущобы» (1864), в котором встречается выражение *Палестины забугорные*.

Так я тебе, милый человек, про матушку нашу рассказывать стану. ...Широки, брат, эти Палестины забугорные!.. Реки у нас широкие, — Волга супротив наших — тыфу! (В. Крестовский. Петербургские трущобы. Книга о сытых и голодных. Ч. 4. 1864). Жизнь — все та же, что и прежде, с тою только вечно повторяющейся разницей, что на место некоторых старых «жильцов», угодивших либо на волю, либо на Владимирку, в Палестины забу-

горные прибывают день ото дня «жильцы» новые, с тою же по большей части перспективой воли — «с подозрением», да длинной Владимирки и Уральских бугров. (В. Крестовский. Петербургские трущобы. Книга о сытых и голодных. Ч. 4. 1864).

Согласно словарю Н. Смирнова, выражение *Палестину забугорные* означает ‘Сибирь зауральская’ [Смирнов 1899]. Как видим, прилагательное *забугорный* имеет архетическое происхождение, но в конце XX века его, по-видимому, не изобрели, а «вспомнили». Если на жаргоне 1990-х *за бугром* означает ‘за границей’, то на жаргоне XIX века *за буграми* значило ‘за Уральскими горами, в Сибири’.

Источники

Национальный корпус русского языка [Электронный ресурс]. URL: <http://ruscorpora.ru>

Studiorum. Образовательный портал Национального корпуса русского языка [Электронный ресурс]. URL: <https://studiorum-ruscorpora.ru>

Литература

БАС – Словарь современного русского литературного языка. В 17 томах. Т. 2. М.–Л.: Изд-во Академии наук СССР, 1951.

Буцева Т. Н. (ред.). Новые слова и значения. Словарь справочник по материалам прессы и литературы 90-х годов ХХ века в двух томах. Т. 1 (А – К) / Сост. Т. Н. Буцева, Е. А. Левашов, Ю. Ф. Денисенко, Н. Г. Стулова, Н. А. Козулина, С. Л. Гонобоблева; Отв. ред. Т. Н. Буцева. Ин-т лингвистических исследований РАН. СПб.: Изд-во «Дмитрий Буландин», 2009. 870 с.

Геккина Е. Н. К истории прилагательного *виртуальный* в русском языке // Филология на рубеже тысячелетий: Материалы Междунар. научн. конф. Вып. 2. Ростов н/Д., 2000. С. 21–23.

Ефремова Т. Ф. Новый словарь русского языка. Толково-словообразовательный. М.: Русский язык, 2000. 2310 с.

Комлев Н. Г. Словарь новых иностранных слов. М.: Изд-во МГУ, 1995. 144 с.

Маринова Е. В. Иноязычные слова в русской речи конца ХХ – начала ХXI в.: проблемы освоения и функционирования. М.: Языки славянских культур, 2008. 712 с.

Скляревская Г. Н. (ред.). Толковый словарь русского языка ХХI в. Актуальная лексика / Под ред. Г. Н. Скляревской. М.: Эксмо, 2006. 1136 с.

Смирнов Н. Слова и выражения воровского языка, выбранные из романа Вс. Крестовского «Петербургские трущобы». СПб., 1899.

- Ушаков Д. Н. (ред.). Толковый словарь русского языка. В 4-х томах. Т. 1. М.: Государственный институт «Советская энциклопедия», 1935.
- Южаков С. Н. (ред.). Большая энциклопедия. Словарь общедоступных сведений по всем отраслям знания / Под ред. С. Н. Южакова. Т. 5. СПб.: Просвещение, 1901. 794 с.

References

- BAS – *Slovar' sovremennoi russkogo literaturnogo yazyka. V 17 tomakh* [Dictionary of modern Russian literary language]. Vol. 2. Moscow–Leningrad, Izd-vo Akademii nauk SSSR, 1951.
- Butseva T. N. (ed.). *Novyye slova i znacheniya. Slovar'-spravochnik po materialam pressy i literatury 90-kh godov XX veka v 2 t.* [New words and meanings. Dictionary of reference materials of the press and literature of the 90s of the twentieth century in two volumes]. Vol. 1 (A – K). In-t lingvisticheskikh issledovanii RAN. St. Petersburg, Dmitrii Bulanin Publ., 2009. 870 p.
- Gekkina E. N. [To the history of the adjective *virtual'nyi* ('virtual') in Russian / Philology at the turn of the millennium: Materials of the Intern. scientific conf.]. *Filologiya na rubezhe tysyacheletii: Materialy Mezhdunar. nauchn. konf. Vyp. 2.* Rostov na Donu, 2000. Pp. 21–23. (In Russ.)
- Komlev N. G. *Slovar' novykh inostrannykh slov* [A Dictionary of New Foreign Words]. Moscow, 1995. 144 p.
- Marinova E. V. *Inoyazychnye slova v russkoi rechi kontsa XX – nachala XXI v.: problemy osvoeniya i funktsionirovaniya* [Foreign words in Russian at the end of the 20th – beginning of the 21st century: problems of mastering and functioning]. Moscow, Yazyki slavyanskikh kul'tur Publ., 2008. 712 p.
- Sklyarevskaya G. N. (ed.) *Tolkovyj slovar' russkogo yazyka kontsa XXI veka. Aktual'naya leksika* [Explanatory Dictionary of Russian language of the late XX century. Actual vocabulary]. St. Petersburg, Folio-Press, Institute of Linguistic Studies, 1998. 704 p.
- Smirnov N. *Slova i vyrazheniya vorovskogo yazyka, vybrannye iz romana Vs. Krestovskogo «Peterburgskie trushchoby»* [Words and expressions of the thieves' language, selected from the novel "Petersburg Slums"]. St. Petersburg, 1899.
- Ushakov D. N. (ed.). *Tolkovyj slovar' russkogo yazyka. V 4-kh tomakh.* [Explanatory dictionary of the Russian language in 4 vols.]. Vol. 1. Moscow, Gosudarstvennyi institut «Sovetskaya entsiklopediya», 1935.
- Yefremova T. F. *Novyi slovar' russkogo yazyka* [New Dictionary of the Russian Language. Interpretive word-building]. Moscow, Russkii yazyk Publ., 2310 p.
- Yuzhakov S. N. (red.). *Bol'shaya entsiklopediya. Slovar' obshchedostupnykh svedenii po vsem otrasmam znanija* [Great Encyclopedia. Dictionary of publicly available information on all branches of knowledge]. Vol. 5. St. Petersburg, Prosveshchenie Publ., 1901. 794 p.