

# О МОЩИ РУССКОГО ЯЗЫКА ПО ДАННЫМ СИНОНИМИЧЕСКОГО СЛОВАРЯ-СПРАВОЧНИКА КОМПЬЮТЕРНОЙ ОЦЕНОЧНОЙ СИСТЕМЫ ASIS®

УДК 81

**Виталий Николаевич Тришин**  
к.ф.-м.н., Председатель Совета директоров  
компании «ОКП», г. Москва  
<http://www.okp-okp.ru>  
Тел. (495) 222-22-58  
Эл. почта: [mail@trishin.ru](mailto:mail@trishin.ru)

В статье описывается электронный словарь-справочник синонимов русского языка системы ASIS® (более 500 тыс. слов и фразеологизмов, 190 тыс. синонимических связей). Программа может быть использована не только как словарь синонимов и близких по смыслу слов, но и в качестве орфографического и толкового словарей русского языка с целью проверки правильности написания и определения значений незнакомых слов. Словарь также может служить инструментом для филологических наблюдений и исследований языка благодаря развитой системе запросов по разным признакам слова (по значению, составу, синонимии и т.д.). Словарная база программы включает слова из словарей и справочников по всем отраслям знаний – от астрономии до японской живописи. При составлении словаря разработчиком были использованы изданные словари: орфографические, синонимические, фразеологические синонимов, толковые, иностранных слов и т.п. 19–21 столетий. Кроме того, для пополнения словаря активно использовались газеты, журналы и интернет-ресурсы.

Этот словарь практически показывает, что по количеству слов русский язык относится к наиболее развитым языкам мира, а по масштабу и плотности пространства синонимов ему, по мнению автора, нет равных.

**Ключевые слова:** «мощность» русского языка, мощь, количество слов в русском языке, синоним, синонимический, электронный словарь, ASIS®.

**Vitaly N. Trishin,**  
PhD in Physico-mathematical Science,  
Chairman of the board of directors of «OKP»  
company, Moscow  
<http://www.okp-okp.ru>  
Tel.: (495) 222-22-58  
E-mail: [mail@trishin.ru](mailto:mail@trishin.ru)

## THE MIGHT OF RUSSIAN LANGUAGE ACCORDING TO SYNONYMIC DICTIONARY BY COMPUTER EVALUATION SYSTEM ASIS®

The article describes electronic dictionary of synonyms in Russian language by ASIS® system (more than 500 000 words and collocations, 190 000 synonymic connections).

The program can be used not just as a dictionary of synonyms and close meaning words, but also as spelling dictionary and definition dictionary of Russian language in order to check the orthography and define the meaning of unknown words. The dictionary is also designed to be an instrument of philological surveys and studies of the language through the extensive query system on different characteristic of words (definition, composition, synonymy, etc.). Program's lexical base includes words from dictionaries and guides in all subject areas – from astronomy to Japanese painting. Over compilation of dictionary developer used published dictionaries: spelling, synonymic, definition dictionaries, dictionary of collocations, dictionary of foreign words and etc. of 19–21 cc. Newspapers, magazines and web-resources were active used as well for appending the dictionary.

This dictionary practically shows, that by the amount of words Russian language belongs with the most developed languages in the world, and by the scale and density of synonymic space, in the author's opinion, it has no equal.

**Keywords:** mightiness of Russian language, might, amount of words in Russian language, synonym, synonymic, electronic dictionary, ASIS®.

## 1. О русском языке, а также о назначении и особенностях словаря

*«В начале было Слово»*

*(Евангелие от Иоанна)*

*«Язык есть дом бытия. В жилище языка обитает человек. Мыслители и поэты – хранители этого жилища»*

*(М. Хайдеггер «Время и бытие»)*

В настоящее время диалектологами принято считать (см., например, работы Г.О. Винокура), что в русском языке имеются 3 диалекта (наречия), а именно: северный диалект (ладого-тихвинский, вологодский и костромской говоры) и южный диалект (курско-орловский и рязанский говоры), между которыми находится территория с населением использующим, как правило, среднерусский диалект (среднерусские говоры) – Псковская, Тверская, Московская, Владимирская, Ивановская, Нижегородская области. Отличия в речи населения различных регионов России уменьшились в связи с развитием в XX веке образования, СМИ, миграцией населения. В Сибири и на Дальнем Востоке нет относительно чёткого деления территории на диалектные зоны. Диалектные зоны сохранились по большей части в сельской местности. На диалектах, говорах написаны многие выдающиеся произведения русской литературы: например, поэма «Песнь о Великой Матери» Н.А. Клюева написана на северном диалекте, этот же диалект использовал и В.И. Белов; на южном диалекте писали многие свои произведения В.И. Даль, М.А. Шолохов. По-видимому, на исходном древнерусском языке говорили в Киевской Руси, который затем распространился на восток в связи с переносом столицы России из Киева (был столицей Руси в 882 – 1243 гг.) сначала во Владимир, а затем из Владимира в Москву (1389) (многие историки считают первой столицей России до Киева Великий Новгород (862 – 882), а также Ладогу). Западные же области Малороссии (Правобережье) попали, по большей части, на 4 с половиной века (разные территории по-разному) под власть Великого княжества Литовского (условно 1340 – 1569), Польши (условно 1569 – 1793), а также часть территории и в разные годы Турции и Австро-Венгрии. В 1793 в результате второго раздела Польши Правобережная Украина вошла в состав Российской Империи вместе с большей частью Польши, и сегодняшний украинский язык – это, так считают многие лингвисты-историки, ополяченный старорусский язык (паны говорили со своими крестьянами по-польски), см., например, книгу А.М. Волконского.

«Создававшийся преимущественно в Галиции, с использованием множества заимствований из польского, немецкого, других языков, украинский язык, который не следует путать с малорусским наречием, был понятен в основном галичанам, привыкшим жить бок о бок с поляками и немцами». [<http://telegrafua.com/country/12715/>]. Слова из малорусского и белорусского наречий помещал в свой словарь и Владимир Даль наряду со словами из губернских диалектов великорусского наречия. Эти три наречия и составляли, как тогда было признано, русский язык.

«До монгольского нашествия литературный язык, с незначительными диалектными различиями, был совершенно одинаков. Так же и быт был един от Киева до Ладоги» [В.В. Кожинов, см. [http://www.patriotica.ru/religion/kozhinov\\_idea/html](http://www.patriotica.ru/religion/kozhinov_idea/html)]. «В IX столетии Кириллу и Мефодию, когда они создавали славянскую азбуку, при их путешествиях не нужен был переводчик, потому ... все славяне говорили на одном языке» (Балязин В. Н. Неофициальная история России, 2007). И именно на Русском Севере сохранились былины Киевской Руси! После революции 1917 года малорусский и белорусский языки, считавшиеся в Российской Империи диалектами русского языка, приобрели статус официальных, государственных. И названия «Украина» и «украинцы», иногда встречавшиеся в литературных про-

изведениях до революции 1917 года, закрепились также за Малороссией и малороссами фактически только лишь при большевиках [<http://gold-nike.inspaceblog.net/?p=134>]. По мнению автора, истинной целью действий по «выдавливанию» русского языка и культуры сегодняшней «украинской элитой» да и «элитами» других стран бывшего СССР является ограничение доступа своих граждан к подлинной русской культуре, в частности, к великой русской поэзии Лермонтова, Тютчева, Фета, Анненского, Рубцова, Передреева, Соколова, Левитанского и многих других (более 100 поэтов мирового уровня, чего и близко нет ни в одной стране мира!) и к русской песне. (По словам Кожина «Русская песня – главная основа отечественной культуры». [<http://www.krounov.ru/news/2011/02/02/10783/>], «Антология русского лиризма. XX век», <http://studia-vasin.ru/vstupant.htm>]; «Без песен мы не люди, не русские люди» (А.Н. Васин, журнал «Родина», №3, 2005 год, <http://studia-vasin.ru/rodina.htm>).

А также для дальнейшего разделения единого в культурном плане народа и варваризации населения, и, тем самым, защиты своих экономических интересов и личного бизнеса. При этом и в России, и на Украине, и в других странах бывшего СССР пошлейшей масс-культуре на русском языке, непотребству, шансону-блатняку никакие препоны не ставятся. А русскую жизнь и в России, и за рубежом заставляют изучать уже второе столетие по произведениям Салтыкова-Щедрина, Достоевского и других «назначенных» классиков русской литературы [«Читаю Лермонтова» А.Н. Васина, [http://studia-vasin.ru/2012/Chitau\\_Lermontova.pdf](http://studia-vasin.ru/2012/Chitau_Lermontova.pdf)]. Отмечу, что через язык происходит и прямое управление людьми. Например, чтобы заставить одних людей относиться по отношению к другим определенным образом изобрели в 20–30-х годах 20-го века слова «кулацко-монархический», «кулацко-помещичий» и т.п.

Кстати, и в других языках (странах) имеются региональные особенности, например житель Нью-Йорка не всё может понять из английской речи жителя Техаса. Даже в одном городе могут быть несколько диалектов. (Так в Лондоне известны диалекты английского языка Cockney, Estuary English и др.). В Великобритании насчитывается

более 25 диалектов, а всего в английском языке имеется более 90 диалектов. [<http://www.homeenglish.ru/grdial.htm>]. Однако эти страны никто не разделяет, как бывший СССР.

Примечание. В данный словарь вошли и слова из местных диалектов, в том числе и многие из малорусских и белорусских слов, упоминаемые как русские в дореволюционных словарях русского языка (наряду с вологодскими, архангельскими и др.), но регион их употребления в словаре не указан. Интересующиеся данным вопросом могут обратиться к словарю В. И. Даля и другим.

Электронный синонимический словарь-справочник первоначально разрабатывался с 1993 года как часть компьютерной системы помощи оценщику и аудитору ASIS® (Appraiser Support Info System) для быстрого поиска с высокой точностью аналогов для оцениваемых машин и оборудования в базах данных системы, а также для правильной их кодификации по классификатору ОКОВ, см. статью В.Н.Тришина «Задача классификации основных средств согласно классификатору ОКОВ выбора амортизационной группы и оставшегося срока полезного использования», <http://www.trishin.ru/left/publishes/classification-task/>.

В процессе развития словарь приобрел самостоятельное значение и в настоящее время распространяется отдельно от системы ASIS® с сайта компании «ОКП». В месяц его скачивают с сайта более 1000 человек практически со всех стран мира. В помещенной на вышеупомянутом сайте версии 6.9 от 24.09.2013 (язык программирования Visual Fox Pro) словаря-справочника содержится 508 тыс. слов и выражений (490 тыс., если убрать дублирование слов с буквами «е» и «ё», см. ниже), из которых 390 тыс. снабжены синонимами, квазисинонимами, фразеологическими синонимами. Всего синонимов (поясняющих слов, синонимических связей) около 1998 тысяч (это сумма по всем словам относящихся к ним синонимических связей). Отметим также, что сокращения (а их в русском языке более 130 тысяч, см. сайт <http://www.sokr.ru/>) в данный словарь не входят, кроме слоговых сокращений типа комсомол, минфин, спецназ; не входят в словарь и топонимы для не крупных объектов (села, фамилии, речки и т.п.) и т.п.

В более современной программной оболочке словаря версии 8.3 (от 15.04.2013), написанной на языке С++, имеется более мощная система запросов, в ней 489 тыс. слов. Однако в ней имеется около 300 слов, исправленных в новой версии 6.9, и в настоящее время эта версия существенно перерабатывается.

Словарная база включает слова не только технического содержания, но и слова из словарей и справочников по всем отраслям знаний: от астрономии до японской живописи. При составлении словаря использованы изданные словари: орфографические, синонимические, фразеологических синонимов, толковые, церковно-славянских слов, иностранных слов, арг (кроме заведомо ненормативной лексики) и т.п. 19–21 столетий. Кроме того, для пополнения словаря активно использовались газеты, журналы, Интернет, автоподсказчики слов по их началу из Yandex и Google (с проверкой по другим источникам), так как даже в самых крупных печатных словарях отсутствуют десятки тысяч широко распространенных в быту и прессе современных слов, например, госучастие, доследственный, эсэмэска, моногород, спецлаборатория, промпроизводство, медсправка, педколледж, зоопарк-махер, внеоборотный, внеофисный, дошкольный, межкорпоративный, самораспаковывающийся, кибертерроризм, погранслужба, госсимволика, автоэксперт и т.д. В данном словаре имеется 272 слов, начинающихся с «нано», а в новом Большом академическом словаре таких слов имеется всего 13. Основной пласт слов академических словарей занимают, как правило, слова и отрывки с соответствующими словами из произведений русских писателей 19–20 веков А.И. Герцена, Н.П. Огарева, А.Ф. Писемского, В.Г. Белинского, Н.Г. Чернышевского, А.М. Горького и многих других, которые считались классиками в работах советских литературоведов. Слов из современного обиходного языка в них, как правило, немного. В словаре-справочнике представлены и устаревшие и редкие слова из академических словарей. Этот словарь является попыткой создания объединённого словаря русского языка, о необходимости создания которого писал русский литературовед и лингвист-русист, академик АН СССР В.В. Виноградов: «Между тем потребность в историческом словаре русского лите-

ратурного языка была осознана очень давно – почти при самом зарождении русской лексикографической науки. Мысль об историческом словаре русского литературного языка не вполне была чужда патриарху русского «корнесловия» акад. А. С. Шишкову. «Все находимые в священных писаниях, в летописях, в законах, в преданиях, в народных сказках и песнях самые старинные, хотя бы неупотребительные, но чистые славянские слова должно внести в Словарь» – учил он.» (Виноградов В.В. Избранные труды. Лексикология и лексикография. – М., 1977.). Что касается роста числа слов в крупнейших словарях, то «знаменательны такие соотношения цифр: в «Словаре Академии Российской» (1806–1822) содержалось 51388 слов; в «Словаре церковнославянского и русского языка» (1847) уже было помещено 114749 слов; «Толковый словарь» В. И. Даля выходил за пределы 200 000 слов» (В.В. Виноградов).

Предлагаемый электронный словарь многофункционален.

Он может быть использован как орфографический словарь русского языка для проверки правильности написания слов, особенно при составлении документов с использованием сложных и редко встречающихся специальных терминов. Следует сказать, что многих слов из данного словаря нет и в самых больших орфографических словарях, а есть такие слова, написание которых в словарях, справочниках, энциклопедиях отличается. Например, в орфографическом словаре В. В. Лопатина имеются слова «кросскурс», «кортико-висцеральный» и «кристально честный», а в орфографическом словаре Б. З. Букчиной, Л. П. Какалуцкой – они же пишутся как «кросскурс», «кортиковисцеральный» и «кристально-честный». Отметим также, что в орфографическом словаре В.В. Лопатина с началом «кортико» имеется 5 слов, а в данном словаре их 45. Еще пример – слово «крупнодроблённый». В словарях В. В. Лопатина и С.И. Ожегова его нет, в словаре С.Н. Корчемкиной, С.К. Кашкиной оно оформлено как «крупнодроблённый», а в словаре Б.З. Букчиной, Л.П. Какалуцкой как «крупно дробленый». В Русском орфографическом словаре РАН под редакцией В.В. Лопатина и О. Е. Ивановой имеются слова «пухоперовой», «пушно-заготовительный», «пушно-меховой», а в Большом

академическом словаре русского языка эти же слова пишутся как «пухо-перовой», «пушнозаготовительный», «пушно-меховой». Единицы измерений «авто-час», «нормо-час», «кормо-час», «станко-час», «пече-час», «полёто-час» и др. в разных орфографических словарях, справочниках, энциклопедиях пишутся то слитно, то через дефис. В данном словаре они все пишутся через дефис. Зато слово «политчас», не являющееся единицей измерений, пишется слитно. Слова с окончанием «кислый» (а их в данном словаре 278) пишутся в словарях и справочниках по химии то слитно, то через дефис. В данном словаре такие слова все пишутся слитно (кроме «нехимических» терпко-кислый, солёно-кислый и др.). И даже в орфографических словарях разных лет выпуска под ред. В.В. Лопатина есть слова по-разному оформленные, например, слова плоско-вогнутый и плоско-выпуклый (плосковогнутый и плосковыпуклый), фликер (фликкер) и фликерный (фликкерный) и др. Имеются расхождения в словарях в написании существительных с окончанием -ие (таких слов в словаре 14838), например, бдение, безлуние, безмолвие и с окончанием -ье (таких слов в словаре 1943), например, ауканье, безлюдье, безлесье. Приходилось в этих и подобных случаях принимать «волевое» решение на основании даты выпуска словаря.

В настоящее время происходит сверка данного словаря-справочника с 4-ым изданием (2013 г.) Русского орфографического словаря РАН. И даже в этом словаре есть неточности. Например, в словарях по химии нет слова «эстрадиол-пропионат», зато есть слово «эстрадиол-дипропионат». А в словаре РАН – наоборот.

Этот словарь может служить в качестве подсказчика слов по их началу в интернет-поисковике существенно более мощного, чем автоподсказчики слов Yandex и Google. Возможно создать также автоподсказку по дополнительным параметрам, например, выдавать слова с заданным не только началом, но и окончанием.

Как словарь синонимов и поясняющих слов он может использоваться при поиске близких по смыслу слов и выполнять функции толкового словаря при определении значений незнакомых слов. Он может быть реализован в виде Интернет онлайн-версии со всеми поисковыми возможностями, реали-

зованными в компьютерных версиях 6.\* (на языке Visual Fox Pro) и 8.\* [на языке C#, см. ниже]. Для сравнения, в синонимическом словаре Н. А. Абрамова, размещенном на сайте Yandex в online, имеется 19 тысяч слов. Данный словарь вместе с алгоритмом поиска может быть встроен нами и в другие компьютерные и Интернет системы и служить инструментом для филологических наблюдений и исследований языка благодаря развитой системе запросов по разным признакам слова (по значению, составу, шаблону, синонимии и т.д.), что невозможно ни в online-словарях, ни тем более в книжных словарях.

Сочинители стихов имеют возможность пользоваться словарём как машиной для подбора рифм (шутка).

Специалисты по всевозможным отраслям знаний могут им пользоваться как справочником для поиска нужной информации. В словарях и справочниках в виде книг слова упорядочены по алфавиту, и, если вы не помните точно, как оформлено нужное слово, то вам приходится просматривать сплошь весь словарь или справочник. В электронном словаре можно искать нужное слово по обобщенному понятию, например запрос (в версии 6.9) по слову «рыба» выдаст в алфавитном порядке список из 800 названий различных рыб, по слову «растение» – 4468, «песня» – 163, певец 98, «гриб» – 356, «столица» – 274 названия существующих и бывших столиц, «минерал» – 5651, «взрывчатка» – 234, «фермент» – 305, «богиня» – 347, милиционер – 67, «быстро» – 305, «умный» – 117, «глупый» – 229, «хороший» – 289, «плохой» – 309, «избивший» (избить) – 260, «трезвый» – 35, «выпивший» (выпить) – 354, «пьяный» – 338 (американский президент Бенджамин Франклин собрал около 200 синонимов слова «пьяный» (англ. Drunk)), и т.д.

В качестве примера «мощности», многозначности русского языка укажем на такие данные из этого словаря версии 6.9:

в русском языке, согласно словарю, имеют синонимические связи (синонимы, квазисинонимы, фразеологизмы) более 390119 слов и словосочетаний, более 10 синонимов имеют 42692 слов и словосочетаний,

более 20 синонимов имеют 17252 слов и словосочетаний,

более 30 синонимов имеют 8526 слов и словосочетаний,

более 40 синонимов имеют 4700 слов и словосочетаний,

более 60 синонимов имеют 1720 слов и словосочетаний,

более 100 синонимов имеют 430 слов и словосочетаний,

более 200 синонимов имеют 89 слов и словосочетаний,

более 400 синонимов имеют 25 слов,

более 600 синонимов имеют 15 слов,

более 1000 синонимов имеют 9 слов.

Такое большое количество синонимов в данном словаре обусловлено двумя причинами:

1. В различных словарях синонимов словарные статьи к одному и тому же слову значительно отличаются, и, по сути, они дополняют друг друга. А поскольку в электронный словарь включены слова с синонимами практически из всех известных словарей, то в данном словаре пользователь получает их объединенный вариант.

2. Механизм перекрестных ссылок, реализованный в словаре, резко увеличивает число синонимов к словам. Поясним на примере. В словарях арго одним из синонимов к слову «оттянувшийся» обычно указывают слово «отдохнувший», но отдельной словарной статьи со словом «отдохнувший» в словарях арго нет. Более того, ни в одном синонимическом словаре среди синонимов к слову «отдохнувший» мы не нашли слова «оттянувшийся». Другие примеры: «схлестнувшийся» – повздоровивший, «сдавший» – «выдавший», «сдвинутый» – «ненормальный», «содрванный» – «украденный», «убойный» – «замечательный» и т.д. Электронный словарь устанавливает все возможные соответствия: синонимами в словаре могут быть не только слова и словосочетания современного общеупотребительного языка, но и устаревшие слова, просторечные, жаргонные, областные, слова профессиональной речи и т.п.

Предлагаемый словарь имеет несколько особенностей.

1. Отсутствие деления омонимов, что приводит к тому, что в одной группе синонимов к слову, скажем, «стан» оказываются слова «лагерь», «блюминг» и «талия». Для развитого носителя языка это не представляет сложности, но для иностранца, желающего углубиться в недра русского языка, отсутствие деления омони-

мов будет, конечно, представлять большие сложности. Для того же «стана» правильно было бы иметь несколько словарных статей, например:

Стан 1 ⇔ блюминг, трубопрокатник, листопрокатник, ...

Стан 2 ⇔ становище, стойбище, станица, бивуак, лагерь, ...

Стан 3 ⇔ торс, талия, поясница, ...

В будущих версиях словаря, возможно, такое разделение будет предусмотрено.

2. Отсутствие деления синонимических и родо-видовых связей. Родо-видовые отношения – это, вообще говоря, предмет тезаурусов. Синонимическое отношение – это отношение эквивалентности, которое разбивает все слова на классы эквивалентности. А родо-видовое отношение – отношение частичного порядка. В идеале должна быть вкладка для синонимов, для родительских терминов и для дочерних. Такое разделение также планируется произвести в программе словаря после завершения работы над его пополнением.

3. Отнесение слова в словаре к какой-либо лексико-грамматической категории (группы частей речи) чисто условно, что объясняется прикладным характером словаря. Слова в версии 6.9 распределены на четыре группы:

1. Существительные, числительные – 226180.

2. Прилагательные, причастия, глаголы, деепричастия – 254415.

3. Наречия, сравнительные степени, сращения – 26341.

4. Частицы, предлоги, междометия и др. – 1082.

При этом в справке и запросе названия групп «обрезаны», а сокращения, первоначально присутствующие в ранних версиях словаря, удалены, как правило, из словарной базы.

Если слово может выступать в роли разных частей речи (омоформы), то оно отнесено к той, в форме которой наиболее часто встречается. Фразеологические и синтаксические синонимы, не определяемые вообще морфологически (как словосочетания), приписаны к части речи того слова, к которому эти фразеологические и синтаксические синонимы привязаны. В словаре версии 8.3 имеется 5 групп частей речи (в отдельные группы выделены прилагательные и фразеологизмы, в которые ошибочно попали некоторое количество обычных слов, а 3-я и 4-ая группы из версии 6.\* объединены), ко-

торые на экран для слов не выводятся, а даны только в запросе и справке по запросу. Часть речи используется для ускорения поиска нужных слов и лингвистических исследований. Замечу, что часть речи не приводится и во многих книжных словарях.

4. В словах специально-технического лексикона буква ё заменена на букву е, в обрабатываемых системой ASIS® материалах буква ё заменена также на букву е, поэтому в таких словах эта замена в словаре-справочнике сделана. В последних версиях словаря-справочника слова с буквой ё имеются, а единственным синонимом к ним служат такие же слова с заменой буквы ё на букву е. Всего слов и выражений с буквой ё в словаре имеется 18405, в основном из словаря Т. Ф. Ефремовой, академических словарей (последним отдавалось предпочтение в случае разночтений), других источников; в частности, использовались запросы из данной же словарной базы по фрагментам слов с имеющейся буквой ё с последующим анализом найденных слов. Это значительно прибило число слов с буквой ё из специальных словарей, например технических, где буква ё, к сожалению, обычно игнорируется.

5. Во всех словах вместо прописных букв используются строчные буквы (в следующей версии словаря это планируется исправить).

6. Нет знаков ударений в словах.

7. Отсутствуют пометы: устарелое, просторечное, областное и т.п.

8. Частично уточнена орфография касательно «бесов», поврежденная в 1918 году, а именно, наряду со словами бесчестный, бестолковый, бессовестный и т.п. введены в словарь и слова безчестный, безтолковый, безсовестный и т.п. [книга «Жизнь происходит», изд-во «Студия», Москва, 2009]. Хорошо хоть этим бесам не удался планируемый ими перевод русского языка с кириллицы на латиницу!

9. Глаголы представлены не в одной нейтральной форме, а в нескольких наиболее часто используемых причастных формах, как они представлены и в Русском орфографическом словаре РАН под редакцией В. В. Лопатина и О. Е. Ивановой.

Однако устарелые, просторечные, жаргонные глаголы – только в одной.

Перечисленные особенности возникли из специфики и целевого назначения первоначального словаря, заключающихся в распознавании

близких по смыслу именных словосочетаний в прикладной компьютерной системе ASIS®. На использование словаря в практической работе с текстами и документами отмеченные особенности не влияют.

Понятие «синоним» используется в расширенном смысле. Нас интересует в первую очередь, являются ли взаимозаменяемыми конкретные слова в именных словосочетаниях (с сохранением хотя бы частично их смысла) или нет. Иначе говоря, улучшит ли установление синонимической связи между двумя конкретными словами качество распознавания смысла именных словосочетаний или нет при компьютерной обработке словосочетаний. А является ли данное слово точным синонимом другого слова или только смежно ему по смыслу (то, о чем часто спорят филологи) нас интересует во вторую очередь.

**Замечание.** Если устранить дублирование слов с буквой «ё», глаголов в различных формах, а также фразеологизмы, то уникальных слов в словаре останется не менее 415 тысяч, а вовсе не 200 тысяч, как ошибочно указано на популярном филологическом портале Philology.ru, (см. <http://www.philology.ru/links.htm>).

Отмечу, что в самом полном словаре английского языка Уэбстера имеется 425 тыс. слов, число же синонимических связей в английском языке неизвестно.

## 2. Устройство словаря

Экран компьютера разделен на две части (панели), размеры которых можно регулировать с помощью мыши.

На экране слева помещены слова в порядке алфавита вместе с указанием части речи и числа синонимов. Справа в порядке алфавита перечислены синонимы к тому слову на левой панели, на котором стоит курсор. К словам-синонимам на правой панели также указаны части речи и число синонимов.

Любой синоним (поясняющее слово) к конкретному слову имеет, в свою очередь, также синонимы (поясняющие слова). Для их просмотра достаточно в поле «наименование» правой панели сделать двойной клик левой клавишей мыши на выбранном синониме. При этом выбранное слово переходит из правой панели в левую.

Для удаления синонимов (толкований) нужно выделить слово в списке

синонимов (толкований), а затем нажать кнопку со знаком минус из меню.

## 3. Поиск слов

Простейший поиск слов выполняется по совпадению начального фрагмента слова с поисковым выражением. Для получения статистической информации и анализа структуры словарного фонда русского языка в систему добавлена возможность получения выборок по запросам.

В данном электронном словаре можно осуществлять практически любые запросы к словарной базе. В нём можно искать искомое слово по маске слова (включающей известные буквы и условные символы, обозначающие произвольную букву, произвольное число произвольных букв и т.п. – версии 8.3), а также по его синониму или обобщённому понятию.

Для каждого слова отображаются синонимы всех его смысловых значений и слова, связанные с ним родо-видовым отношением как с определяющим, так и с обобщающим понятием.

Для лиц, нуждающихся только в простых запросах, предлагается «Построитель запросов» (версия 8.3) – аналог системы запросов версии 6.9. В качестве параметров «Построителя запросов» используются:

- начало слова,
- фрагмент слова,
- окончание слова,
- часть речи,
- количество синонимических связей слова,
- количество символов в слове.

При сложных запросах можно осуществлять поиск по комбинации параметров.

Онлайн-словари делать такие запросы не позволяют.

Отобранный список выдается на экран, упорядоченным по алфавиту, по числу синонимов и т.д.

Например. Запрос по фрагменту «вита» (слово «вита» в переводе с латинского означает «жизнь») даст 282 слова, а запрос с флексией «метрия» (от греческого слова «метрео» – «измеряю») выдаст список из названий 424 дисциплин, в которых что-то измеряют, а запрос с флексией «логия» (от греческого слова «логос» – слово, учение) выдаст список из названий 1294 научных и квазинаучных дисциплин. Замечу, что в «Советском энциклопедическом словаре» (4-е изд., 1990)

содержится только около 200 статей о науках, направлениях и квазинауках, включающих в название элемент «-логия», которые перечислены и в Википедии, см. <http://ru.wikipedia.org/wiki/Логия>.

## 4. Алгоритм поиска близких по смыслу наименований объектов в системе ASIS®

Наименования объектов представляют собой именные словосочетания, содержащие главное слово и несколько уточняющих слов. Например, «станок настольно-сверлильный 2М-112», «вертикально-фрезерный станок», «холодильная установка 24 МКТ 280». Предполагается, что в наименовании всегда есть одно главное слово, которое находится в именительном падеже в единственном или множественном числе. Алгоритм поиска близких по смыслу наименований начинается с выделения главного слова и всех его синонимов. Затем выделяются уточняющие слова и составляются списки их синонимов. Поиск близких по смыслу наименований производится в базе наименований и начинается с выделения всех наименований, имеющих главное слово, совпадающее с главным словом входного наименования или его синонимом. Для каждого из отобранных наименований рассчитывается значение критерия близости. Критерий является суммой экспериментально подобранных весовых коэффициентов для главного и уточняющих слов. Например, главное слово имеет вес 1, синоним главного слова 0.9, уточняющее слово 0.8, синоним уточняющего слова 0.7, а, допустим, наименование марки изделия – 2.5. Все наименования-кандидаты упорядочиваются в порядке убывания значения критерия и в таком виде предлагаются пользователю для окончательного выбора.

В системе ASIS® поиск близких по смыслу наименований производится в базах аналогов машин и оборудования, а также классификатора ОКОФ.

Для ускорения поиска предварительно выполняется разбор всех наименований в каждой базе аналогов. Результаты этого разбора хранятся в специальных файлах и содержат не сами слова, а их коды в виде ссылок на записи словаря слов системы ASIS®. Аналогичный файл результатов разбора составляется и для наименований классов классификатора ОКОФ.

## 5. В заключение о мощности русского языка

Интенсивную работу над словарем-справочником планируется закончить примерно к концу 2014 года. По прогнозам автора словаря в нём будет около 545 тыс. слов и выражений, из которых примерно 415 тысяч будут снабжены синонимами, квази-синонимами, поясняющими словами (сходными по смыслу выражениями). Всего синонимических связей будет около 1950 тыс. (за вычетом задвоенных с буквой «ё» будет около 525 тыс. слов и словосочетаний, и 1925 тыс. связей). По-видимому, цифры 455 тыс. уникальных слов (без дублирования форм глаголов и слов с буквой Ё и учёта фразеологизмов), 1630 тыс. синонимических связей примерно и характеризуют «мощность» словарного состава русского языка, не считая сокращений и многочисленных топонимов для некрупных объектов, однако я не нашёл данных о работах по реальному сравнению «мощности» и «разветвлённости» различных языков. Именно выявление этих цифр было одной из главных задач создания данного словаря. Приводящиеся в популяризаторских книгах данные о суммарном количестве слов в различных языках, как правило, документально не подтверждаются их авторами (см., например, книгу «Рекорды Гинесса»). Заметим, что в замечательной книге Льва Успенского «Слово о словах» (1971 г., 5-е издание) отмечается, что, «Сотрудники словарного отдела (Института языкознания в Ленинграде) полагают, что ими зарегистрировано около 400–500 тысяч разных русских слов. Вероятно, в коллекции есть пропуски (она и сейчас непрерывно пополняется), но, в общем, цифра эта близка к исчерпывающей».

Кроме того, поэт Е. А. Евтушенко, сравнивая русский и английский языки, высказал мнение, что «в английском нет столько рифм, сколько у нас и приходится рифмованный стих в связи с израсходованностью рифм на английский переводить без рифмы», см. <http://www.echo.msk.ru/programs/beseda/43457/>.

По мнению писателя Саши Соколова: «Английский не так пластичен, не обладает свободой русского – Тургенев был совершенно прав. Русский инверсивен, фраза строится без всяких там... Всё можно переиначить, начать с любого слова. А в английском нельзя

слишком вольничать. Ещё русский народ каким-то образом – может, в силу лености своей? – избавился от частиц и артиклей. Лишнего в языке нет. Насколько загадочней получается язык, когда убирается артикль! Ты говоришь не об одном предмете, а о целом их ряде. Я говорил и с австрийскими, и с немецкими филологами о русском языке. Они такие аккуратисты, их раздражает и коробит в русском неопределённость, размытость описаний, диалогов. А мне это очень легло на душу. У народа, который никогда не стремился к внешней свободе, есть внутренняя, даже какая-то анархия. Поэтому, как устроен язык, можно судить о народе, о степени его изобретательности», см. [http://www.chaskor.ru/article/sasha\\_sokolov\\_sily\\_eshche\\_est\\_grebu\\_ya\\_horoshu\\_11971](http://www.chaskor.ru/article/sasha_sokolov_sily_eshche_est_grebu_ya_horoshu_11971).

Таким образом, с помощью языка человек познаёт мир, и от языка зависит, как именно человек познаёт, видит мир и как он мыслит. Именно язык народа во многом определяет менталитет народа, его национальные ценности и особенности наряду с религией, территорией обитания и т.д.

Обновление версий словаря на сайте <http://www.trishin.ru> в ближайший год планируется производить один раз в полгода, в дальнейшем обновление словаря будет происходить раз в несколько лет.

В следующей версии электронного словаря версии 8.4 для облегчения изучения работы программы пользователями-иностранцами (осваивающих русский язык) планируется дополнительно переписать интерфейс программы и встроенное руководство пользователя на 7–9 иностранных языках (английском, французском, немецком, испанском и др.), с тем чтобы в зависимости от установленного при загрузке компьютера языка, этот же язык выбирался для интерфейса и встроенного руководства словаря-справочника. Скриншот готового словаря на 4-е языка приведён в Приложении.

Ссылки на синонимы в словаре начиная с версии 6.3 закодированы в качестве защиты от недобросовестных пользователей, а таких в Интернете немало, разместивших его на своих сайтах без указания автора. Например, если поискать в Google ошибочное словосочетание «без вознаграждения» (вместо «без вознаграждения») – будет 710 ссылок (на дату 24.09.2013). Эта ошибка была допущена в размещённой на нашем сайте до 2010 года незашиф-

рованной версии словаря. Не случайно ни количество слов, ни количество синонимов в этих словарях не указано. Цель большинства этих «лексикографов» – увеличить число посетителей своих сайтов, на которых рекламируются разные товары и услуги, вплоть до порнографии.

И этот последний открытый словарь 2010 года версии 6.2 на 383 тыс. слов стал слишком (в нём, к сожалению, много опечаток), по моему мнению, популярен в Интернете по сравнению с новыми версиями этого же словаря да и другими словарями. Например, запрос на дату 24.09.2013 в [Google.com](http://www.google.com) «Словарь»+«Тришин»+«2010» даёт 2,04 млн. ссылок на словарь. Для сравнения укажем данные по другим, неоднократно переиздаваемым словарям русского языка (кроме многотомного Большого академического словаря), упоминаемым в статье Википедии как якобы крупнейшие словари русского языка ([http://ru.wikipedia.org/wiki/Словари\\_русского\\_языка](http://ru.wikipedia.org/wiki/Словари_русского_языка)) по им соответствующим запросам:

- 1) «Словарь» + «Даль» – 5,49 млн. ссылок (около 200000 слов).
- 2) «Словарь» + «Лопатин» – 1,67 млн. ссылок (около 200000 слов).
- 3) «Сводный» + «словарь» + «лексико» – 1,27 млн. ссылок (около 170000 слов).
- 4) «Словарь» + «Ефремова» – 3,96 млн. ссылок (около 160000 слов).
- 5) «Словарь» + «Тихонов» – 0,274 млн. ссылок (около 145000 слов).
- 6) «Словарь» + «Кузнецов» – 2,21 млн. ссылок (около 130000 слов).
- 7) «Большой» + «академический» + «словарь» – 3,29 млн. ссылок (около 120000 слов).
- 8) «Словарь» + «Ушаков» – 4,14 млн. ссылок (около 90000 слов).
- 9) «Словарь» + «Ожегов» – 2,87 млн. ссылок (около 50000 слов).

Данный словарь-справочник в этой статье не указан, а в статье Википедии о словарях синонимов почему-то указана очень старая версия данного словаря на 163 тыс. слов кем-то анонимно размещённая в Интернете в 2003 году вместо современной, см. [http://ru.wikisource.org/wiki/Словарь\\_синонимов](http://ru.wikisource.org/wiki/Словарь_синонимов)

Надеюсь, что эти недоразумения в Википедии будут исправлены.

Автор словаря награжден за его создание несколькими общественными наградами.

## 6. Выводы

1. Наличие такого словаря практически доказывает, что русский язык является одним из самых мощных языков мира.

2. На сегодняшний день данный словарь, реализованный в виде компьютерной базы данных, является самым большим словарём русского языка как по количеству слов, так и по количеству синонимических связей.

3. Он пополнялся вручную около 20 лет его автором на основе как изданных словарей, справочников и энциклопедий, так и других источников (газет, журналов, Интернет и др.).

4. Этот компьютерный словарь позволяет производить практически любые запросы к словарному составу и осуществлять различные лингвистические эксперименты.

## Литература

1. Голубицкий С.М. «Побочный продукт», компьютерный еженедельник «Компьютерра», № 26–27, 2004, <http://offline.computerra.ru/2004/550/34752/>.

2. Голубицкий С.М. «Морфинг»,

компьютерный еженедельник «Компьютерра», № 29–30, 2004.

3. Костырко С.П. «WWW-обозрение», журнал «Новый мир», №9, 2005.

4. Тришин В.Н. «Электронный словарь-справочник синонимов русского языка системы ASIS» в книге «Владимир Даль в счастливом доме на Пресне», М.: изд-во «Academia», 2010, стр. 158–165.

5. Волконский А.М. Историческая правда и украинофильская пропаганда, 1920 (переиздание 2005), <http://www.edrus.org/content/view/51/56/>.

6. Линниченко И.А. «Труды подготовительной по национальным делам комиссии. Малорусский отдел».

7. Флоринский Т.Д. «Лекция по славянскому языкознанию», ч. 1, Киев, 1895, ч. 2, СПб, 1896.

8. Ляпунов Б.М. «Единство русского языка в его наречиях». – Одесса, 1919.

9. Будилович А.С. «О единстве русского народа». – СПб., 1907.

## References

1. Golubickij S.M. Lateral discharge. // *Kompyuternyj ezhenedelnik «Kompyu-*

*terra*», № 26–27, 2004, <http://offline.computerra.ru/2004/550/34752/>.

2. Golubickij S.M. Morphing. // *Kompyuternyj ezhenedelnik «Kompyuterra»*, № 29–30, 2004.

3. Kostyrko S.P. WWW-review. // «Novyj mir», №9, 2005.

4. Trishin V.N. Electronic dictionary of synonyms in Russian language by ASIS system.// *Vladimir Dal v schastlivom dome na Presne*, M.: izd-vo «Academia», 2010, str. 158–165.

5. Volkonskij A.M. Historical truth and ukrainophil propaganda, 1920 (re-edition 2005), <http://www.edrus.org/content/view/51/56/>.

6. Leennichenko I.A. Tractates of preparatory commission for national issues. Little Russian Department. – Odessa: tip. Izd. «Russkaya kultura», 1919. – Vyp. 1: Sbornik statej po malorusskomu voprosu. – 119 str.

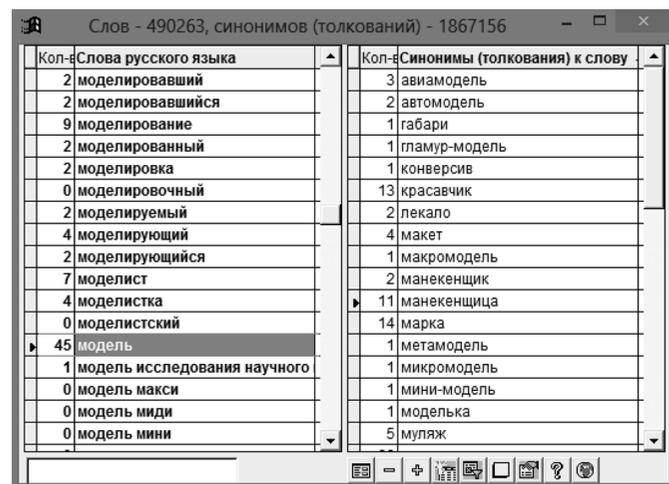
7. Florinskij T.D. Lecture on Slavonic linguistics, part 1, Kiev, 1895, part 2 SPB, 1896.

8. Lyapunov B.M. Unity of the Russian language is in its adverbs. – Odessa, 1919.

9. Budilovich A.S. About the unity of Russian nation. – SPb., 1907.

*Приложение.*

А) Скриншот словаря-справочника версии 6.8 на Fox Pro



Б) Скриншот словаря-справочника версии 8.4 на С# с интерфейсом и помощью на 4-х языках

