

УДК 81'13; 811.161.1  
DOI: 10.17223/19986645/57/8

Е.В. Петрухина, О.В. Дедова

## ИНТЕРНЕТ КАК ИСТОЧНИК ЛИНГВИСТИЧЕСКОЙ ИНФОРМАЦИИ (ДЛЯ ИЗУЧЕНИЯ ДИНАМИКИ РУССКОГО СЛОВООБРАЗОВАНИЯ)<sup>1</sup>

*Рассматривается вопрос об использовании количественных показателей употребления новых производных слов, полученных при помощи браузеров Яндекс и Google, для оценки частотности и степени вхождения новообразований в словообразовательную и лексическую системы русского языка. Обосновывается методика, направленная на оптимизацию интернет-поиска как источника лингвистической информации. Исследование проводится на материале сетевого проекта М. Эпштейна «Дар слова» (2000–2010 гг.), что позволяет также изучить эффективность предлагаемого там способа пополнения лексической системы русского языка.*

*Ключевые слова: интернет-коммуникация, поисковые операторы, количественные данные, неологизмы, словотворчество, современный русский язык.*

### Введение

**Цель исследования.** Русский язык последние тридцать лет находится в русле общих для славянских языков процессов: усиления влияния английского языка, интернационализации и коллоквиализации лексики, активизации словотворчества [1–3]. Словообразование в славянских языках живо реагирует на изменения, происходящие в обществе, наряду с заимствованиями участвуя в номинации новых реалий. Развитие коммуникации в Интернете и совершенствование его поисковых систем позволяют «вывести» динамические инновационные процессы в русском языке в «наблюдаемую зону» и изучить их, опираясь на количественные данные интернет-поиска [4. С. 22]. В частности, возможности использования Интернета в лингвистических целях обсуждались на XI Международной конференции Комиссии по славянскому словообразованию при Международном комитете славистов «Новые явления в славянском словообразовании: система и функционирование», проходившей в Москве в 2009 г. Большой научный резонанс вызвал пионерский в этой области доклад Н.Д. Голева [5], послуживший отправной точкой для нашего исследования. Изучение вопроса о применении поисковых систем Интернета в лингвистических целях было про-

---

<sup>1</sup> Работа выполнена при поддержке гранта РФФИ 17-04-00532-ОГН\18 «Исконные и заимствованные форманты и модели в русском словообразовании на славянском фоне: семантические отношения, типы взаимодействия, стилистический потенциал».

должено в ряде публикаций Н.Д. Голева [4, 6], а также на других конференциях упомянутой Комиссии [7, 8]. Этот вопрос имеет большое значение для исследования динамических процессов в русском словообразовании – интернет-поиск позволяет объективировать наблюдения, связанные с активизацией некоторых словообразовательных моделей, с изменениями в семантике формантов (в том числе исконных и заимствованных) и их конкуренции.

В данной статье вопрос об Интернете как источнике лингвистической информации рассматривается на материале русского языка с учетом широкого спектра поисковых опций, предоставляемых современными браузерами. Нас интересует возможность использования количественных показателей употребления новых производных слов на сайтах Интернета, полученных при помощи браузеров Яндекс и Гугл (Google), для оценки частотности дериватов и степени их вхождения в словообразовательную и лексическую систему русского языка.

**Сопутствующие задачи.** Свое исследование мы проводим на материале словообразовательного сетевого проекта М. Эпштейна, предпринявшего в начале XXI в. попытку «творческого обновления» лексического состава русского языка при помощи индивидуального словотворчества и внедрения созданных слов в русскую речь посредством Интернета. Имеется в виду проект М. Эпштейна «Дар слова», действовавший более десяти лет (2000–2010 гг.) и посвященный, как указано на сайте проекта, «искусству создания новых слов и понятий, исследованию путей обновления лексики и грамматики русского языка, развитию корневой системы, расширению моделей словообразования» [9]. Наш выбор языкового материала связан прежде всего с тем, что на «словесных изобретениях» проекта «Дар слова» удобно проанализировать достоверность количественных показателей при применении браузеров в поиске выбранных слов ввиду реальности проверки этих данных вручную.

В рамках названного проекта новые слова создавались на основе авторских «расширительных моделей словообразования» как гнездовым способом от одного корня, так и по отдельным словообразовательным моделям (о методике словотворчества см. в [10, 11]). Для созданных слов подбирались значения и возможные контексты употребления, затем они размещались на сайте проекта и распространялись по интернет-рассылке. По мнению М. Эпштейна, «интернет делает возможным мгновенное распространение нового слова среди огромного количества читателей. Новообразование может быть подхвачено на лету, и его успешность легко проследить по растущему из года в год и даже из месяца в месяц числу употреблений» [12]. Созданные слова разгруппированы на сайте проекта «Дар слова» по 259 разделам в зависимости от времени создания, тематики, автора и т.д. [9]. Прошло более семи лет после завершения активной деятельности в рамках данного проекта – срок, с нашей точки зрения, достаточный для возможности анализа его словотворческой эффективности. Эту большую исследовательскую работу еще только предстоит выполнить – масштабный проект М. Эпштейна, по нашему мнению, заслужива-

ет подробного анализа. Здесь мы лишь ставим этот вопрос, рассматривая в связи с основной целью нашего исследования всего несколько дериватов из «проективного лексикона».

Применение интернет-технологий в лингвистических целях имеет, с нашей точки зрения, большое значение для изучения динамических изменений в русском языке. Нельзя не согласиться с мнением М. Эпштейна, что до создания Интернета трудно было определить истоки и сферу употребления новых слов [12]. Но его утверждение о том, что «с появлением Сети это делается простым нажатием клавиши в поисковой системе» [Там же] требует, по меньшей мере, проверки. Действительно ли все так просто?

Мы проводим исследование специфики и информативности количественных показателей интернет-поиска по выбранным словам с учетом того, что за последние несколько лет поисковые технологии усовершенствовались. Как известно, в настоящее время активно развивается WEB 3, или семантический WEB<sup>1</sup>, существенно расширяющий возможности поиска по ключевым словам и оптимизирующий его результаты (подробнее см. ниже). Поэтому в центре внимания в настоящей статье будут потенциальные возможности современного Интернета как источника достоверной лингвистической информации для изучения частотности производных лексем, прежде всего новообразований, в современных текстах, представленных в Сети. Данная проблема представляется актуальной, поскольку в настоящее время в отечественной лингвистике еще не выработаны методики сбора и систематизации лингвистического материала, полученного в результате интернет-поиска на основе использования браузеров, несмотря на то, что многие лингвисты обращаются к этому источнику данных.

### **Интернет-технологии с лингвистической точки зрения**

Развитие электронных коммуникативных технологий и Интернета оказывает самое непосредственное воздействие на национальные языки, в том числе и на русский. Изучение результатов этого воздействия прошло несколько этапов, что, видимо, отражает эволюцию самого явления. Если вначале в отечественной лингвистике основное внимание уделялось влиянию на русский язык сетевого общения, которое оценивалось неоднозначно [13] (предполагалась возможность потенциально негативного воздействия: так называемая «падонковская коммуникация», массовое нарушение равноуровневых норм в неформальных сообщениях и под.), то в настоящее время становится очевидным, что проблема функционирования национальных языков в Интернете более многопланова и многоаспектна. Активно развивается научное направление, получившее название «лингвистика

---

<sup>1</sup> Семантический Веб (Semantic Web) – термин, предложенный создателем Интернета Т. Бернерс-Ли (Tim Berners-Lee) для обозначения современной концепции развития поисковых технологий. В их основе – принцип автоматического семантического анализа документов с целью выполнения сложных поисковых задач пользователей.

Интернета». Термин впервые был употреблен Д. Кристалом [14]. С его точки зрения, эта актуальная область научных исследований имеет различные «концепции»: социолингвистическую, образовательную, стилистическую, практическую (последнее подразумевает документацию национальных языков, а также поддержание малых языков). Основным объектом исследования интернет-лингвистики является так называемый «язык Интернета». Данный термин стал способом совокупного обозначения многообразных сдвигов (речевых, текстовых, коммуникативных, семиотических), обусловленных распространением электронной сетевой коммуникации [15]. В последние годы растет количество работ, посвященных данной тематике (см., например, [16–18]). Но есть еще один очень важный аспект использования Интернета как источника лингвистического материала: при помощи поиска по ключевым словам исследователи могут выявлять и изучать факты, отражающие тенденции развития языка в целом.

Электронное общение предоставляет коммуникантам возможности обмена информацией, ранее не существовавшие в истории цивилизации. Интернет также не имеет и жанровых ограничений – здесь представлены практически все типы текстов и все функциональные стили. Следует учесть одно важное обстоятельство: инновационные характеристики электронного общения нивелировали существовавшие различия между письменной и устной речью, поскольку обмен письменными репликами стал возможен в реальном масштабе времени. Особенности речевого поведения в пределах межличностной интернет-коммуникации формируются практически теми же факторами, что и в ситуации устного диалогового общения, описанными, в частности, в [19]: спонтанность, темп, отсутствие строгой стилистической регламентированности и т.д. Это позволяет наблюдать явления, отражающие тенденции развития языка, причем время их узуальной адаптации может существенно сокращаться. Как следствие, Интернет стал уникальным источником того, что Л.В. Щерба называл «языковым материалом» (напомним, под этим понимается «совокупность всего говорящего и понимаемого в определенной конкретной обстановке в ту или другую эпоху жизни данной общественной группы» [20]).

Современные компьютерные технологии значительно упрощают и ускоряют обработку огромных массивов текстовой информации, и это их преимущество стало общедоступным благодаря Интернету. Отметим, что использование Интернета как источника лингвистической информации имеет как минимум два важных аспекта. Процедура поиска интересующих исследователя фактов может осуществляться через браузеры или же на основе использования так называемых национальных корпусов. Оба типа данных отражают реальное функционирование языковых единиц в текстах различных типов, но суть их отличается весьма существенно. Корпусы, являясь продуктом деятельности лингвистов, представляют собой универсальный и очень мощный источник информации. Система корпусной разметки, которая постоянно развивается и совершенствуется, имеет целью предоставление не только статистических данных о том или ином языко-

вом факте, но и справочной информации о нем. Так, в Национальном корпусе русского языка (НКРЯ (<http://ruscorpora.ru>)) в настоящее время используется пять типов разметки: метатекстовая, морфологическая (словоизменяемая), синтаксическая, акцентная и семантическая; планируется также внедрение словообразовательной разметки и упрощённой синтаксической разметки (<http://www.ruscorpora.ru/>). Аннотированный поиск в корпусах предоставляет научно достоверные данные, без которых уже трудно представить современное исследование в области грамматики, словообразования, лексики и других областях филологического знания. Поиск интересующих исследователя фактов на основе браузеров, напротив, стихийен и вариативен в своих результатах (см. об этом ниже). Но его главное исследовательское значение заключается в том, что в отличие от корпусных данных здесь нет предварительного отбора текстов. Именно здесь аккумулирован огромный массив только что созданных письменных текстов разного жанра, в том числе и воплотивших в себе специфику устного повседневного общения (так называемую «речевую пену дней»). Поэтому именно результаты интернет-поиска дают сведения о тех тенденциях и фактах, которые только начинают появляться в речи.

Огромные языковые ресурсы Интернета, мощность которых трудно переоценить, имеют особенно большое значение при изучении образования и употребления новых слов в разнообразных текстах, представленных в Сети. Языковой материал, полученный в результате интернет-поиска, позволяет также анализировать соотношение потенциальных моделей деривации и реальных лексем, образованных по этим моделям. Последние, хотя и не зафиксированы словарями русского языка, активно употребляются в речи. Не ставшие еще узальными лексемы часто остаются не отмеченными и в НКРЯ. Поэтому при исследовании динамики русского словообразования Корпус в ряде случаев оказывается недостаточно информативным.

Но при всей очевидности возможностей Интернета как источника лингвистической информации он таковым практически ещё не стал – не хватает исследований методов его использования для изучения русской речи. Видимо, причины этого коренятся не только в некотором недоверии к языковым данным, полученным в результате обращения к Интернету (см., например, [21]), но и в том, что до сих пор отсутствуют научно обоснованные методики использования поисковых запросов в качестве инструмента анализа современного русскоязычного узуса. Работы, выполненные на материале Интернета, посвящены главным образом специфическим явлениям непосредственно самого интернет-общения (интернет-жаргон, языковые особенности различных сетевых жанров и т.д., см., например, [22]).

### **3. Опыт применения интернет-поиска с лингвистическими целями в русистике**

В этой связи представляют большой интерес идеи Н.Д. Голева, высказанные им в уже упоминавшемся выше докладе и ряде статей [4–6], в которых автор ставит своей целью научно обосновать потенциал процедуры

интернет-поиска на основе браузеров для лингвистических исследований. В частности, на большом языковом материале Н.Д. Голев показал, что возможность поиска в Интернете по конкретному заданному слову, в том числе не по зафиксированному, а потенциально предполагаемому, позволяет перейти от «иллюстративного поиска лексических реализаций к системному, от эвристического описания к прогностическому» [5. С. 202]. При таком поиске реальна «возможность извлечения и описания непрерывных лексических, словообразовательных, лексико-словообразовательных, лексико-грамматических пространств (полей, парадигм, цепей, гнезд) и представление их в виде матриц» [Там же. С. 203]. Большое значение имеет также доступность «контекстов вхождения» языковых единиц, в том числе и новообразований, что позволяет проводить их полномасштабный семантический и стилистический анализ [там же]. Количественные данные поисковых систем Интернета анализировались также с точки зрения описания лексических реализаций словообразовательных типов, гнезд и парадигм. Было показано, что они «дают возможность более качественной оценки порождающей силы словообразовательной модели и мотивирующего потенциала мотиватора» [6. С. 233]. Главным критерием узуального статуса языкового факта, подвергаемого анализу, становится количественный показатель результатов поиска по соответствующему ключевому слову: чем больше количество контекстов вхождения, фиксируемых браузером, тем более актуально явление. По мнению Н.Д. Голева, «количественная характеристика, которую дает слову “квантитативный коэффициент”, извлеченный из поисковых систем Интернета, является одновременно информативной» [5. С. 198]. Действительно, в настоящее время употребление новых дериватов в речи трудно обнаружить и изучить в полном объеме без использования электронных средств коммуникации и компьютерных поисковых систем. И в этом смысле Интернет предстает как универсальный источник лингвистической информации, обладающий целым рядом преимуществ. Но некоторые вопросы использования Интернета в лингвистических интересах, затронутые в статье Н.Д. Голева, требуют более детального обсуждения, тем более что сам автор отмечает в ряде случаев «парадоксальность» «статистики Интернета»<sup>1</sup>.

Таким образом, количественные показатели интернет-поиска (и «квантитативная мощность» Интернета) требуют ряда уточнений. С одной стороны, порядок сопоставляемых цифр («квантитативный коэффициент») отражает некоторую объективную данность, а с другой – нельзя не учитывать целый ряд факторов, чье воздействие на предъявляемые поисковые

---

<sup>1</sup> «Например, иначе чем как парадоксальную не можем оценить статистику в паре *железобетон / железобетонный* (428 000 / 3 250 000): прилагательное здесь в 7,6 раза превосходит в количественном отношении существительное. Однако не исключаем, что этот парадокс имеет свои резоны, не замечаемые поверхностным взглядом» [6. С. 233].

результаты может быть весьма ощутимым. Рассмотрим наиболее существенные из них.

Интернет-технологии постоянно совершенствуются. Как мы уже упоминали, сейчас находит практическое воплощение концепция Веб 3.0 (Web 3.0), или так называемый семантический Веб (semantic Web). Суть данной концепции состоит в том, что она способствует профессиональной оптимизации контента и поиска, т.е. информационному совершенствованию Интернета. Оптимизация поиска идет за счет использования принципов семантических сетей и автоматического анализа текста (в отличие от изначального поиска на основе тезаурусов, когда учитывалось простое вхождение слова в текст – прежде всего в его название). Современные браузеры должны отранжировать (т.е. отсортировать) огромный объем данных в соответствии с поисковым запросом и предоставлять пользователю наиболее релевантную информацию (например, в настоящее время в Яндексе используется более 1500 факторов ранжирования). А. Сафронов, руководитель службы релевантности и лингвистики Яндекса, пишет по этому поводу: «Для того, чтобы представить себе, что такое ранжирование, хорошо подходит изображение нашей галактики Млечный путь. Потому что количество документов, которые проиндексировал Яндекс, и количество звезд в этой галактике – это числа приблизительно одного порядка. А задача ранжирования – показать десятку самых релевантных результатов» [23].

Одна из основных задач развития поисковых технологий – адаптировать предъявляемые результаты поиска к индивидуальным информационным потребностям конкретного пользователя. Как заявлено в «Миссии Яндекса» на официальном сайте компании, «качество поиска – это самый важный аспект для любой поисковой системы. Если она будет плохо искать, люди просто перестанут ей пользоваться» (<https://yandex.ru/company/rules/ranking/>), поэтому факторы ранжирования результатов поиска учитывают в том числе и предыдущие поисковые запросы, совершенные через данный аккаунт. Все это приводит к тому, что результаты поиска по конкретному ключевому слову или словосочетанию разнятся в зависимости от того, когда, кем и при помощи какого браузера осуществлены запросы.

Количественные показатели узального статуса того или иного явления могут быть уточнены в результате применения не одного, а нескольких браузеров, например Яндекса и Google, являющихся самыми используемыми поисковыми системами в России (47 и 48% соответственно от всей российской интернет-аудитории, по данным на май 2017 г. (<https://marketer.ua/rejting-poiskovyh-sistem-v-2017-statistika-stran-sng-i-mira/>)). Как известно, эти браузеры используют несовпадающие поисковые алгоритмы, поэтому результаты поиска могут разниться весьма существенно. В частности, Google, в отличие от Яндекса, осуществляет функцию дедупликации (метод сжатия массива данных, призванный исключить повторение копий в результатах поиска). Google изначально не предъявляет то, что в его интерфейсе называется «похожие результаты». Эти ссылки можно вызвать, нажав соответствующую кнопку, но количественный показатель результатов поиска,

предъявляемый по запросу, их не учитывает. Яндекс, в большей степени ориентированный на русскоязычный контент и отечественную аудиторию, отличается от Google своей гео independency, т.е. результаты поиска могут зависеть от того, где был осуществлен запрос. Как следствие, количественные показатели поиска по ключевым словам через Яндекс и Google практически никогда не совпадают.

Проиллюстрируем названные выше проблемы конкретными примерами. Проанализируем частотность и продуктивность образования глаголов от коммерческих названий программного продукта при помощи суффикса *-и-* типа *гуглить*. Данная модель последовательно используется в современном русском языке – глагольные дериваты образуются практически от всех наименований распространенных программ и приложений, включая названия браузеров. Проверив частотность некоторых потенциальных дериватов в Яндексе, мы получили следующие данные: *гуглить* – 112 000; *яндексить* – 6 000; *фотошопить* – 2 000 000; *инстаграмить* – 29 000; *рамблерить* – 359; *экселить* – 331 (по результатам на 07.08.2017, для уточнения результатов поиска был использован оператор «поиска по цитате» ["], об операторах поиска подробнее см. ниже). Полученные количественные результаты свидетельствуют об актуальности самой модели, от которой образуются приставочные дериваты типа *прогуглить*, *отфотошопить*, *заинстаграмить* и под. Для указанных выше глаголов поиск в Google дал несколько иные результаты: *гуглить* – 918 000; *яндексить* – 10 900; *фотошопить* – 1 260 000; *инстаграмить* – 22 600; *рамблерить* – 512; *экселить* – 160 (по результатам на 07.08.2017, для уточнения результатов поиска также был использован оператор «поиска по цитате»).

В следующем разделе мы рассматриваем более подробно данные использования названных выше браузеров при изучении неологизмов, имеющих в Рунете меньшую частотность, чтобы последовательно изучить всю информацию об их употреблении в текстах разного жанра. Такой анализ, помимо сформулированных выше целей, позволяет также решить ряд лингвистических проблем, связанных с образованием неологизмов и изучением когнитивно-дискурсивных условий вхождения их в русский язык. Как уже было отмечено, мы анализируем частотность в текстах Рунета ряда новообразований, связанных с интернет-проектом М. Эпштейна «Дар слова».

### Анализ результатов интернет-поиска

**Когнитивные и дискурсивные условия вхождения новых слов в русский язык.** Словообразовательная система участвует в пополнении лексики русского языка прежде всего за счет: 1) заполнения имеющихся в ней лакун (об этих процессах см. [24–26]); 2) расширения мотивационно-деривационных отношений ([27]); 3) действия законов аналогии в деривации [28]; 4) развития адаптирующей функции словообразовательных типов при освоении заимствований [2]; 5) окказионального словотворчества [29,



30]. Данные источники словообразования пополняют лексическую систему языка в определенных условиях: когда создание новых слов связано с познанием мира и осмыслением нового опыта в процессе коммуникативной деятельности человека. Этот процесс является когнитивным, выражающим (и формирующим) знания о мире, и одновременно дискурсивным, неотделимым от порождения речи в определенной коммуникативной ситуации [31. С. 391; 32]. Возможно ли вхождение новообразований в русский лексикон при нарушении одного из этих условий? Анализ количественных результатов интернет-поиска по лексемам, образованным в рамках проекта «Дар слова», позволяет в известной степени ответить и на этот вопрос.

Ниже мы приводим конкретные примеры из материалов практики, проведенной в 2016–2017 гг. на филологическом факультете МГУ под руководством авторов статьи<sup>1</sup>. В рамках практики мы изучали динамику русского словообразования и когнитивно-коммуникативные условия вхождения новых слов в русский язык, а также возможности интернет-поиска по ключевым словам в качестве лингвистического источника.

**Лингвистическая релевантность данных о частотности новообразований в Интернете.** Рассмотрим частотность конкретных новообразований, например потенциального каузативного глагола **общать**, образованного депостфиксацией. Он толкуется в проекте «Дар слова» следующим образом: «направлять и поддерживать процесс общения, вовлекать людей в общение и разговор, посредничать, предлагать темы, подбадривать собеседников, вызывать их интерес друг к другу» (<http://new.topos.ru/veer/24/dar29.html>). При обращении к поисковой системе Яндекс в начале декабря 2016 г. по данному неологизму без применения операторов поиска получено 4 млн вхождений. При проверке этих показателей 25.11.2017 Яндекс выдал 3 млн страниц. Для нашего исследования важным является вопрос о том, какая информация стоит за столь большими цифрами, полученными «простым нажатием кнопки»?

Анализ первых страниц сайтов, выбранных поисковиком Яндекс, показывает, что большая часть ссылок дается на словарь Даля, в котором употребляется омоним данного глагола: «Общать – общить что чему, приобщать, соединять, смешивать; считать вместе, заодно». Десятки ссылок на сайты автоматического подбора слов и форм можно считать пустыми, так как там не содержится никакой информации об употреблении данного глагола. Кроме того, на первых десяти страницах обнаружены ссылки на сайты с опечатками: с раздельным написанием *-ся* (*Но тем не менее, общать ся где-то надо*); с пропуском гласной *об(е)щать* и употреблением данной формы в значении ‘обещать’. На количественные результаты поиска влияет также наличие так называемых «зеркал» сайтов. Под зеркалом понимается полная или частичная копия одного сайта на другом. Изначально зер-

---

<sup>1</sup> В практике, которая проходила на филологическом факультете МГУ с 01.09.2016 по 20.12.2016, принимали участие студенты Н.С. Варивода, А.Д. Леоненко, М.Г. Шерварлы, Д.А. Ямилова.

кала сайтов появлялись в результате использования двух версий доменов, с «www» и без (например, [www.msu.ru](http://www.msu.ru) и [msu.ru](http://msu.ru)), сейчас они могут применяться в коммерческих целях, в целях безопасности и т.д. Естественно, на результаты поиска также влияют различные практики интернет-цитирования. Воздействие данных факторов, способное исказить реальное положение вещей, особенно существенно, когда анализу подвергается инновационное явление, недостаточно освоенное узусом. В нашем случае на результаты поиска, осуществленного через Яндекс, повлияли прямые ссылки на проект М. Эпштейна (<http://www.emory.edu/INTELNET/dar0>) и цитирование его материалов, а также зеркала этого сайта (например, <http://old.russ.ru/antolog/intelnet/dar0.html>).

Но самое главное, результаты поиска **без операторов** недифференцированно включают и возвратный глагол *общаться*, а также глагол *сообщать*, что кардинальным образом меняет количественные показатели употребления в Интернете невозвратного глагола *общать (кого-то)*<sup>1</sup>. Все сказанное вызывает большие сомнения в информативности полученных количественных результатов простого поиска по ключевым словам.

В связи с этим встает вопрос **об оптимизации поисковых запросов** в исследовательских целях на основе опций, предоставляемых современными браузерами. Эти системы уделяют много внимания лингвистическим аспектам совершенствования доступа к информации и предлагают поиск с учётом ряда морфологических, лексических и других критериев. Так, Яндекс по умолчанию осуществляет поиск в пределах заданной леммы, учитывая ее частеречную принадлежность и парадигму форм, т.е. при запросе [делать] будут предложены контексты, содержащие словоформы «делаю», «делаешь», «делать» и т.д., но не «деятель», «дело» и под. При этом формальный язык запросов включает целый **ряд операторов**, что позволяет конкретизировать поисковые задачи. Яндекс имеет две категории операторов, специфика которых в меню раздела «Помощь» обозначена как «Морфология и поисковый контент» и «Документные операторы» (<https://yandex.ru/support/search/query-language/qlanguage.html>). Операторы первого типа позволяют искать по форме слова (оператор ! [!делал]); по нескольким ключевым словам одновременно (оператор + [делал+дело]); осуществлять поиск документов, содержащих слова запроса в заданной последовательности и конкретных формах, так называемый «поиск по цитате» (оператор " ["сделал дело гуляй смело"]) и т.д.

Документные операторы, операторы второго типа, дают возможность уточнить поисковый запрос с помощью данных, относящихся не к текстовому контенту, а к информации о страницах, например: искать в пределах

---

<sup>1</sup> По всей видимости, с неприменением операторов связана и «парадоксальная статистика» интернет-поиска в паре *железобетон / железобетонный* (428 000 / 3 250 000), упомянутая в сноске 2. При перепроверке данных с применением оператора [-], позволяющего исключить однокоренное слово из результатов поиска, мы получили следующие цифры: [*железобетонный* – *железобетон*] 486 000 (Google, 07.02.2018).

указанного сайта, хоста, домена, осуществлять поиск в файлах определенного типа, на конкретном языке и т.д. (<https://yandex.ru/support/search/query-language/search-operators.html>). Использование документных операторов при поиске по ключевым словам имеет социолингвистическую перспективу, поскольку это позволяет, например, реализовать «жанровые» критерии при отборе языкового материала (сайты СМИ, социальные сети и т.д.) или разграничить результаты поиска по близкородственным языкам. Отметим, что наиболее популярные поисковые операторы (выбор языка, сайта, времени запроса, типа файла, поиск по словоформе) представлены в фильтрах расширенного поиска Яндекса:

The image shows the Yandex search interface with the search term "гуллить" entered. Below the search bar, there are several filter categories: "ПОИСК", "КАРТИНКИ", "ВИДЕО", "КАРТЫ", "МАРКЕТ", "ПЕРЕВОДЧИК", "НОВОСТИ", and "ЕЩЁ". Underneath, there are three rows of filter boxes:

В Красногорске	<input type="checkbox"/>	На сайте	.ru	Точно как в запросе
Русский	Английский	Ещё ▾	Тип файла ▾	
За сутки	За 2 недели	За месяц	От	До

At the bottom right of the filter section, there is a button labeled "Очистить".

Возможности расширенного поиска предоставляются и другими браузерами, в частности Google (<https://support.google.com/websearch/answer/2466433?hl=ru>).

Посмотрим, как влияет использование поисковых операторов Яндекса на количественные показатели употребления в Интернете неузвального глагола *общать* (дата обращения: 25.11.2017).

В результате применения оператора «поиск по цитате» ["общать"], который задает поиск документов, содержащих слова запроса в заданной последовательности и форме, количественные показатели составили 14 тыс. результатов. Но и они не отражают реального числа употреблений именно этого глагола (см. ниже). Запрос на личные формы данного глагола дал практически отрицательные результаты. Несмотря на то, что поиск по форме ["общаю"] дал 4 тыс. контекстов вхождения, интересующие нас случаи употребления неузвального деривата единичны (например, *лично я общаю в основном анимешниц*). На результаты поиска повлияла спонтанная омонимия, возникающая в результате ошибок при написании форм других слов: прилагательного *общий* (\*Составить **общаю** характеристику класса); глагола *обещать* (*Я общаю, что дождусь*), раздельное написание постфикса *-ся* (*общаю сь со взрослыми*), а также употребление глагола *общаться* (с кем?) без постфикса *-ся* (*Я с душевнобольными не общаю*). Полученные результаты отражают одну очень существенную проблему, связанную с использованием интернет-поиска по ключевым словам как инструмента лингвистического исследования. Хотя современные браузеры, рабо-

тающие по принципам семантических сетей, становятся все более совершенными в области автоматического анализа текста, решить проблему омонимии они пока не в состоянии. В этом плане корпуса текстов, в частности НКРЯ, обладают огромным преимуществом, поскольку предоставляют данную возможность.

Максимально уточнить результаты поиска для глагола *общать* удалось в результате одновременного применения операторов [!] и [-] [!общать - общаться], что позволило исключить вхождение в выборку слова «общаться» при учете всех форм исследуемого потенциального невозвратного глагола. Мы получили результат 5 000 употреблений. При исключении и результатов поиска глаголов *сообщать* и *приобщать* [-сообщать - приобщать] список сократился до 4 тыс. Среди данной выборки, конечно, остались зеркала сайтов, ошибки, но реальное употребление данного неузального каузативного невозвратного глагола стало очевиднее, хотя результаты поиска нестабильны и могут меняться в зависимости от даты обращения.

Использование интернет-поиска как источника лингвистической информации затруднено также и тем, что в случае, если нас интересуют не только количественные показатели частотности деривата, но и их реальное употребление в тексте, мы сталкиваемся с достаточно сложной задачей. Необходимо «вручную» искать адекватные контексты в огромном объеме предъявляемых результатов поиска, причем они могут быть не на тех сайтах, которые находятся в так называемых «топовых» (т.е. верхних) позициях списка. Дело в том, что порядок предъявления поисковых результатов по конкретному запросу – одна из самых существенных сторон современного Интернета, и владельцы любого сайта прежде всего заинтересованы в повышении его релевантности (для этого, например, используются сео-технологии<sup>1</sup>). На релевантность результатов поиска влияют параметры ранжирования конкретного браузера, а также ряд других факторов. Так, при анализе глагола *общать* пример его использования как интересующего нас узального деривата был отмечен лишь во втором десятке поисковых результатов. Первые 10 ссылок были на упомянутую выше статью словаря В.И. Даля. В «топовых» результатах есть ссылки на проект «Дар слова» (вып. 29, 14 мая 2001). Кроме этого, ряд контекстов предполагает языковую игру: *Сейчас я кратенько, но по существу, расскажу как правильно меня **общать** и дружить* (на игровое использование глагола указывает употребление с ним другого окказионального каузативного глагола); *В динамике работу в команде можно рассматривать как отношения с девушкой. Девушке надо дарить подарки, ее надо «гулять», «общать», знакомить с друзьями. Одним словом, отношения должны развиваться* (глагол употреблен в кавычках в ряду других неузальных каузативных

---

<sup>1</sup> SEO (от англ. Search Engine Optimization, или SEO) – технология, позволяющая осуществить поисковую оптимизацию сайта, т.е. повысить его релевантность в результатах поиска по конкретному запросу.

глаголов, что подчеркивает осознание говорящими его необычности и стилистической маркированности). Эти и подобные примеры не являются свидетельством того, что *общать* как казуальный дериват получает распространение, поскольку суть языковой игры как раз и состоит в нарушении узусных норм. Данное явление популярно в межличностном интернет-общении, оно становится специфической формой самовыражения коммуникантов (так называемая креативность). Если говорить о бесспорных примерах употребления глагола *общать* в интересующем нас значении, то они встречаются не в первом десятке результатов и отмечены в основном на сайтах, поддерживающих неформальное межличностное общение (блоги, социальные сети, форумы): *Желающие меня общаться – машите лапой* (kyellinn.diary.ru>p176484663.htm). *В общем технически я уже дома. И меня даже можно общаться...* (lomelind.livejournal.com>487782.html). *Интернет не место для шуток юмора, тут люди думать и общаться друг друга собираются.*

Во второй сотне результатов поиска встретились производные данного глагола с приставкой *по-*: *Как вы оцениваете идею специально пообщать детей с болеющими ветрянкой?* (mama.ru). В целом же «ручная» проверка результатов поиска выявила следующую картину: в среднем на 15 позиций (эта цифра может незначительно варьироваться), выдаваемых одновременно на экране браузера при «листании» результатов поиска, приходится от 1 до 4 вхождений, адекватных нашей цели. Мы можем с большой вероятностью предположить, не рассматривая все остальные вхождения (которых осталось больше 3 тысяч), что картина будет сходная, т.е. из 15 одновременно выдаваемых ссылок в среднем от 1 до 4 могут представлять действительное употребление данного глагола.

Итак, наш анализ показал, что количественный результат простого запроса в системе Яндекс на потенциальный невозвратный глагол *общать* (3–4 млн) абсолютно неинформативен. Использование поисковых операторов позволило уточнить параметры запроса, снизив количественный результат до 4 тысяч. Частичная его проверка вручную показала, что реальных контекстов на употребление изучаемого глагола содержится не более 15–20% от этого количества (т.е.  $\approx 800$ ).

Каузативный глагол *общать* является потенциальным, он заполняет словообразовательную и грамматическую лакуны в русском языке. Этот невозвратный глагол использовался и до проекта М. Эпштейна, подтверждение чему мы нашли в НКРЯ, где имеется одно употребление данного глагола, зафиксированное задолго до упомянутого проекта: *А на судне, где экипаж с бору да с сосенки, перед ледовым плаванием следует людей сблизить и теснее перезнакомить, пообщать за праздничным столом* (Виктор Конецкий. Вчерашние заботы. 1979). Из всего сказанного следует вывод, что «внедрять» потенциальный глагол *общать* в русский язык нет необходимости – сама словообразовательная система подготовила для его образования место, которое легко заполняется при необходимости выразить данный каузативный смысл. Но можно предположить, что

включение данного неузального глагола в сетевой проект поддержало его употребление.

Интернет-поиск потенциального каузативного глагола **выспать**. Сопоставимые в плане итоговых цифр результаты были получены для другого узального каузативного глагола *выспать (кого-либо)* 'сделать так, чтобы некто выспался, способствовать засыпанию и отдыху во сне' (Дар слова. № 31. 04.06.2001). При простом поисковом запросе Яндекс выдал 54 тыс. результатов. При обсуждении узального статуса лексемы цифра может представляться значительной, но результаты недифференцированно включают глагол *выспаться*, а также спонтанную омонимию, возникающую в результате искаженного написания других слов (*выспать ся, выс(ы)пать*), а также ссылки на словарь Даля («ВЫСЫПАТЬ, выспать, или -ся, спать сколько хочется и проснуться, поспать вдоволь, сколько природа требует») и другие онлайн-словари (например, статья *выспать* создана в проекте «Викисловарь» (<https://ru.wiktionary.org/wiki>), но кроме самого ее названия, указывающего на глагол, в ней нет абсолютно никакой информации). Кроме того, в результатах поиска были представлены контексты, содержащие данный глагол в ином значении ('Спать в течение какого-л. – обычно длительного – времени' [33]): *Лишь в период младенчества еще можно быть уверенным, что **ребенок «выспит»** необходимое ему количество часов; Вот положено ему 15 часов, он их за сутки и должен **выспать***. Многочисленные ссылки на соответствующую статью словаря Т.В. Ефремовой также вошли в результаты поиска.

С помощью поисковых операторов были получены следующие результаты: «поиск по цитате» ["выспать"] – 9000; поиск по конкретной словоформе, исключая глагол *выспаться* [!выспать -выспаться] – 6000. Проверка вручную верхних 350 позиций последней подборки дала не более 30 употреблений глагола *выспать* в интересующем нас значении, что составляет 8,5 %. Представляет интерес то, что позицию объектного актанта в этом случае чаще всего занимает лексема *ребенок*, что уточняет неузальную семантику глагола (вносит смысловой компонент 'принудить ко сну'): *Решила **выспать** ребенка, так как спал он в 4 мес. в сутки 11-12 часов; Пытаюсь "выспать" ребенка по советам из статей на вашем сайте, но длинные сны нам все равно не даются, подскажите как быть*. Как и в случае, рассмотренном выше, сфера распространения этого глагола – сайты, поддерживающие неформальное межличностное общение, а возможное употребление кавычек указывает на неузальный статус данного каузативного глагола.

Неузальный глагол *выспать*, как и глагол *общать*, представляет собой потенциальный дериват, легко образуемый при необходимости в тексте. Так, единичные случаи употребления глагола *выспать* зафиксированы в НКРЯ в текстах, до начала проекта «Дар слова». Например: *Я вернулся вниз. Яков спал, уронив голову на пульт. Командор уже пытался отвезти его на турбазу, выкупать и **выспать**, но Яков уперся, как козел, и никуда не поехал* [Андрей Лазарчук. Все, способные держать оружие... (1995)].

Ситуативное и потенциальное словообразование – типологическая особенность русского языка, которую необходимо учитывать при исследовании актуальных словообразовательных процессов.

*Другие результаты интернет-поиска неузальных новообразований.*

Еще одна проблема при использовании количественных показателей результатов поиска по ключевым словам в лингвистических целях – это несовпадение количественных показателей результатов поиска, фиксируемых браузером, и числа реально предлагаемых ссылок. Первый показатель многократно превосходит число возможных переходов к конкретным контекстам вхождения. Мы столкнулись с этим при проверке малоупотребительных окказионализмов из проекта М. Эпштейна «Дар слова». Так, проверили окказионализм *временосец*, который М. Эпштейн толкует следующим образом: ‘человек, активно представляющий, знаменующий свое время, выразитель его понятий и ценностей’. В начале декабря 2016 г. поисковик Яндекс (без поисковых операторов) зафиксировал 73 результата для этого слова, но при этом было предложено лишь 17 ссылок. Из них оказалось, что 4 «пустые», связанные с ошибками в написании слов, 6 ссылок ведут к проекту М. Эпштейна и к его статьям ([http://www.ulita.net/gost\\_v4\\_b2.htm](http://www.ulita.net/gost_v4_b2.htm)); еще 4 – отсылают к публикации С.А. Кутолина «Словля поэзии», представленной на нескольких сайтах. У С.А. Кутолина в пределах одного абзаца встречается и *временосец*, и *временосец* ([http://www.bookol.ru/poeziya/poeziya\\_prochee/186070/str9.htm](http://www.bookol.ru/poeziya/poeziya_prochee/186070/str9.htm)), причем в другой своей публикации автор ссылается на словотворчество М. Эпштейна ([http://samlib.ru/k/kutolin\\_s\\_a/alchemy101.shtml](http://samlib.ru/k/kutolin_s_a/alchemy101.shtml)). Из трех последних ссылок одна представляет небольшое обсуждение творчества Эпштейна в ЖЖ (<http://tamara-borisova.livejournal.com/88166.html>), две другие демонстрируют употребление этого производного с неясным значением в блогах, посвященных играм (*Перед ней стоит Кронос, в своем настоящем облики величественного временосца, держа в массивной руке серп*) ([https://vk.com/topic-121921629\\_34208312](https://vk.com/topic-121921629_34208312)). Мы проверили употребительность данного окказионализма 27.06.2017, получив 84 вхождения, но реально было предложено лишь 18 ссылок. Из полученной информации можно сделать вывод, что данный созданный окказионализм является неупотребительным. На похожее слово *временосец* было получено без применения операторов 578 результатов, из которых доступны только 46 ссылок. При его поиске 27.06.2017 было получено без операторов 2 тыс. вхождений, при введении оператора [!] Яндекс показал 157 результатов, но открылось всего 17 ссылок.

Интересные результаты дает сопоставление количественных показателей поиска в Интернете близких по значению и форме потенциальных дериватов, например таких, как *российствовать* и *российничать*, а также других производных, образованных по сходной модели. Подобные дериваты М. Эпштейн предложил в словообразовательном типе «этнических глаголов», которые «восполняют... изъян в обозначении действий, характеризующих этнические и географические субъекты, с помощью двух суффиксов

-ствова- и -нича-): *российствовать*, *американствовать*, *французствовать*, *российничать*, *американничать* (<http://www.topos.ru/veer/46/dar38.html>). Определение М. Эпштейна: *российствовать* – «действовать по-российски, так, как присуще России и русским». *Российничать* – «действовать в подражание России и россиянам (имитировать, передразнивать, обезьянничать)» [Там же]. *Российствовать*: при интернет-поиске нашлось 343 результата и 150 вхождений (обращение 02.07.2017). Среди первых десяти вхождений – это ссылки на «Дар слова». Еще 4 ссылки на статью «*Российствующие молодчики*» (<http://www.kommersant.ru/doc/624060>), где *российствовать* – значит ‘пропагандировать националистические идеи’. Также 3 ссылки относятся к двум стихотворениям разных поэтов, в которых слово употреблено в похожем значении ([http://world.lib.ru/k/kutolin\\_s/alchemia56.shtml](http://world.lib.ru/k/kutolin_s/alchemia56.shtml) и <http://www.proza.ru/2013/03/31/1080>). Среди форм данного глагола преобладает причастная форма *российствующий* с отрицательными коннотациями. Одна из ссылок указывает на автореферат А.Н. Сокальской «Словотворчество как компонент научного идиостиля Г.Д. Гачева» (2007), в котором приводятся авторские новообразования Г.Д. Гачева, в частности глаголы с суффиксом *-ствова-*, образованные от названия стран *азербайджанствовать*, *российствовать* (образованные задолго до начала проекта М. Эпштейна). В произведениях Г.Д. Гачева также можно найти *американствовать*, например: *Минуты роковые истории вершились, пока я тут американствовать привыкал* (Г.Д. Гачев. Как я преподавал в Америке, 1997). Созданное же в рамках проекта слово *российничать* показало крайне низкую частотность – всего 4 результата (вхождение 02.07.2017), из них 2 ошибочных (тексты написаны с ошибками, соответствующих дериватов не нашлось), 2 относятся к проекту «Дар слова».

На основе результатов проведенного поиска в Яндексе новообразований из лексикона проекта «Дар слова» М. Эпштейна можно сделать вполне обоснованный вывод, что предложенный в рамках проекта «Дар слова» в 2002 г. производный глагол *российничать* не вошел в русскую речь. А глаголы *российствовать*, *американствовать*, представляющие реализацию потенциальной и продуктивной модели глагольной деривации, были зафиксированы в русской речи и до проекта М. Эпштейна.

Что же касается сотен слов, образованных в рамках проекта «Дар слова» по индивидуальным моделям, то выборочная проверка их частотности в Рунете показала, что большая часть таких слов представлена единичными текстами, в основном связанными с проектом М. Эпштейна. Например, следующие производные от корня *-люб-* имеют такие индексы частотности в Яндексе (поиск проводился с использованием оператора [""] 25.11.2017 г.): *любовь* (48, открылось 8 ссылок, 3 связаны с проектом, 2 ошибочные), точно такие же показатели у новообразования *налюбовь*; *улюбье* (78, открылось 7 ссылок, 5 из них связаны с проектом М. Эпштейна); *вприлюбку* (146, открылось 6 ссылок, 5 из них относятся к проекту); *дальнолюбие* (76, открылось 4 ссылки, все связаны с упомянутым проектом), *предлюбье* (70, открылось 15 ссылок, 2 на проект М. Эпштейна, 10 на книгу



футуристов «Дохлая луна», а именно на творение В. Хлебникова «Любхо», текст которого во многом перекликается с производными от корня *-люб-*, образованными в рамках проекта «Дар слова») (<http://www.emory.edu/INTELNET/dar0.html>).

### Выводы

1. Использование интернет-поиска как источника лингвистического материала при всей своей привлекательности для лингвиста не является столь легкой задачей, как это иногда представляется (как «простое нажатие кнопки»). Развитие информационных технологий связано с оптимизацией их применения для пользователей, но данную задачу нельзя считать полностью решенной. Поэтому, как было отмечено, более предпочтительным пока является использование в лингвистических научных целях НКРЯ. Корпусные данные отличаются от сетевого представления электронной информации двумя основными особенностями: сбалансированностью состава текстов и наличием лингвистической разметки – дополнительной информации о текстах и их единицах. Это делает НКРЯ мощным инструментом лингвистического анализа. Однако при исследовании отдельных инновационных лингвистических явлений, например неодериватов, данные НКРЯ не всегда показательны: многие актуальные неологизмы в Корпусе оказываются незафиксированными. Так, из всех рассмотренных выше неологических глаголов, образованных от коммерческих названий программного продукта при помощи суффикса *-и-*, в НКРЯ зафиксирован только *зуглить* (5 вхождений). Что же касается потенциальных новообразований, связанных с проектом М. Эпштейна, то они, за небольшим исключением (*пообщать*, *выспать*), вообще не попали в Корпус.

2. Наше исследование продемонстрировало, что простой поиск по ключевым словам не может быть однозначно признан в качестве метода получения адекватных количественных показателей для исследуемого лингвистического материала – количественные данные являются заведомо более значительными, чем реальное число употреблений слова в конкретном значении с реализацией соответствующей синтагматики. Как было показано, оптимизировать поиск можно с помощью *поисковых операторов*. Однако при оценке количественных показателей надо иметь в виду, что и применение операторов не может снять проблему полисемии и омонимии (как языковой, так и спонтанной), а также влияния зеркалов сайтов, онлайн-словарей (включая непрофессиональные), сайтов по автоматизированному составлению слов из определенных букв (например, <http://getword.ru/>) или поиску слов (<http://poiskslov.com/>) и др. С учетом всего сказанного можно было бы поставить вопрос о *коэффициенте достоверности количественных результатов интернет-поиска*, и это требует дополнительных исследований.

3. Несмотря на рассмотренные погрешности описанных процедур, сама возможность поиска в Интернете неузualmente неодериватов представляется актуальной и информативной, поскольку на разнообразных по жанру

интернет-сайтах, вовлеченных браузерами в поисковый процесс, впервые в истории коммуникации масштабно фиксируется новейший «речевой материал», по которому можно проследить историю возникновения и распространения неологизмов. Соответственно, результаты сбора языкового материала при помощи Яндекса и других браузеров в ходе изучения словообразовательных процессов, несмотря на все технические издержки, имеют большую лингвистическую ценность. Она повышается, если сопоставляются конкурирующие дериваты. Достоверные результаты дает поиск новообразований с целью изучения их вхождения в язык, так как можно проверить значительную часть ссылок на тексты с данными дериватами и исследовать контекстуальные условия их употребления. Отметим, что и до массового распространения интернет-технологий исследователи искали возможность получить сведения о реальном узуальном статусе новообразования. С этой целью, например, использовали экспериментальные данные, полученные от информантов (см. [34]). Интернет-технологии оптимизируют изучение условий и этапов вхождения новых слов в лексическую систему.

4. Возможности применения оптимизированного поиска недериватов были проанализированы на материале ряда новообразований из проекта М. Эпштейна. Наше исследование показало невостребованность ряда сконструированных в ходе проекта производных слов, в частности с корнем *-люб-*, и тщетность попыток внедрить эти и подобные им слова в русскую речь. Выявленный небольшой неологический эффект оригинального и многолетнего проекта в области русского словотворчества обусловлен, с нашей точки зрения, нарушением когнитивно-дискурсивных условий порождения нового деривата и его вхождения в язык. Кроме того, поиск в Интернете с учетом всех возможностей его оптимизации позволил установить, что некоторые неузуальные слова из анализируемого словотворческого проекта были созданы и употреблены задолго до его начала (например, глаголы *общать* и *выспать*, *российствовать* и *американствовать*). Подобные дериваты создаются по продуктивным моделям и в соответствующих дискурсивных условиях легко заполняют имеющиеся в русской словообразовательной системе лакуны, но при этом, как контекстно и ситуативно ориентированные производные, не записываются в словари. Мы также пришли к выводу, что результаты многолетней творческой работы по созданию новых русских слов в рамках проекта «Дар слова» заслуживает более подробного и масштабного лингвистического анализа.

5. Наличие ряда технических факторов, снижающих научную достоверность поисковых результатов, делает насущной задачей создание методик использования процедуры интернет-поиска в лингвистических целях. Это возможно сделать с учетом всего арсенала опций, предоставляемых современными поисковыми технологиям. Комплексное использование доступных в настоящее время средств (разные браузеры, операторы формального языка поисковых запросов, а также данные статистики этих запросов) позволяет существенно конкретизировать результаты поиска, что значитель-

но повышает достоверность Интернета как лингвистического источника. Не приходится сомневаться, что при дальнейшем развитии поисковых технологий перечень опций, способствующих оптимизации лингвистической релевантности Интернета, будет возрастать. В любом случае лингвист, исследующий актуальные явления русскоязычного узуса, не может не учитывать их отражение в сетевой коммуникации.

### Литература

1. *Крысин Л.П.* Об интернационализации фонда словообразовательных морфем // Современное русское языкознание и лингводидактика. Вып. 2: сборник научных трудов, посвященный 85-летию со дня рождения академика РАО Н.М. Шанского. М., 2007. С. 69–72.
2. *Новые явления в славянском словообразовании: система и функционирование: доклады XI Международной научной конференции Комиссии по славянскому словообразованию при Международном комитете славистов / под ред. Е.В. Петрухиной.* М., 2010.
3. *Новые тенденции в русском языке начала XXI века / под ред. Л.В. Рацибурской.* 4-е изд., стер. М., 2016.
4. *Голев Н.Д.* Лексическое функционирование словообразования и лексико-словообразовательная системность русского языка в свете статистики Интернета // Вестник Томского государственного университета. Филология. 2011. № 1. С. 22–31.
5. *Голев Н.Д.* Поисковые системы Интернета как лингвистический источник (на примере решения некоторых теоретических и прикладных вопросов русского словообразования) // Новые явления в славянском словообразовании: система и функционирование : доклады XI Международной научной конференции Комиссии по славянскому словообразованию при Международном комитете славистов / под ред. Е.В. Петрухиной. М., 2010. С. 424–443.
6. *Голев Н.Д.* Лексическая реализация как функциональная характеристика словообразовательной системы русского языка и количественные параметры ее описания // Осемьдесят : сборник научных статей к 80-летию И.С. Улуканова. М., 2015. С. 225–236.
7. *Słowotwórstwo słowiańskie: system i tekst. Prace Komisji Słowotwórczej przy Międzynarodowym Komitecie Słowistów / red. J. Sierociuk.* Poznańskie Towarzystwo Przyjaciół Nauk, seria 13. Poznań, 2012. С. 287–297.
8. *Словообразование и Интернет / ред. Б. Тошович.* Graz : Institut für Slawistik der Karl-Franzens-Universität Graz, 2016.
9. *Эпитейн М.* Дар слова: Проективный лексикон. 2000–2010. URL: <http://www.emory.edu/INTELNET/dar0.html>
10. *Эпитейн М.* Слово как произведение. О жанре однослова // Новый мир. 2000. № 9. URL: [http://magazines.russ.ru/novyi\\_mi/2000/9/epsh.html](http://magazines.russ.ru/novyi_mi/2000/9/epsh.html)
11. *Эпитейн М.* О будущем языка // Знамя. 2000. № 9.
12. *Эпитейн М.* Русский язык в свете творческой филологии разыскания // Знамя. 2006. № 1. URL: <http://magazines.russ.ru/znamia/2006/1/ep13.html>
13. *Иванов Л.Ю.* Язык интернета: заметки лингвиста. 2000. URL: <http://www.faq-www.ru/lingv.htm>
14. *Crystal D.* Language and the Internet (2nd ed.). Cambridge : Cambridge University Press, 2006.
15. *Дедова О.В.* О языке Интернета // Вестник Московского университета. Сер. 9. Филология. 2010. № 3. С. 25–38.
16. *Современный русский язык в интернете / ред. Я.Э. Ахапкина, Е.В. Рахилина.* М. : Языки славянской культуры, 2014.

17. Тошович Б. Интернет-стилистика. М. : ФЛИНТА : Наука, 2015.
18. *Словарь языка интернета* / под ред. М.А. Кронгауза. М. : АСТ-Пресс, 2016.
19. *Русская разговорная речь* / под ред. Е.А. Земской. М. : Наука, 1973.
20. *Щерба Л.В.* Языковая система и речевая деятельность. Л., 1974.
21. *Кошкарёва Н.Б.* Тундровый, тундряной, тундренный или \*тундренный, или Как Интернет способствует появлению и закреплению исключений // Интернет как источник лингвистической информации : сб. науч. ст. Бийск, 2014.
22. *Лутовинова О.В.* Лингвокультурологические характеристики виртуального курса. Волгоград : Перемена, 2009.
23. *Сафронов А.* Поиск Яндексa: Как найти лучшие ответы. URL: <https://www.searchengines.ru/poisk-yandeksa.html>
24. *Земская Е.А.* Словообразование как деятельность. М., 1992.
25. *Земская Е.А.* Активные процессы современного словопроизводства // Русский язык конца XX столетия (1985–1995). М., 1996. С. 90–142.
26. *Улуханов И.С.* Единицы словообразовательной системы русского языка и их лексическая реализация. М., 1996.
27. *Улуханов И.С.* Мотивация в словообразовательной системе русского языка. М., 2005.
28. *Кубрякова Е.С.* Роль аналогии в порождении новых производных слов // Новые явления в славянском словообразовании: система и функционирование : доклады XI Международной научной конференции Комиссии по славянскому словообразованию при Международном комитете славистов / под ред. Е.В. Петрухиной. М., 2010. С. 14–25.
29. *Попова Т.В.* Русская неология и неография : учеб. пособие. Екатеринбург : УГТУ-УПИ, 2005.
30. *Пахомова М.А.* Окказиональные слова и словари окказионализмов // Вестник МГГУ им. М.А. Шолохова. Сер.: Филологические науки. 2013. № 3.
31. *Кубрякова Е.С.* Язык и знание: На пути получения знаний о языке: части речи с когнитивной точки зрения. Роль языка в познании мира. М. : Языки славянской культуры, 2004.
32. *Петрухина Е.В.* Образование новых слов в русском языке: теоретические аспекты и когнитивно-дискурсивный анализ // *Slavische Wortbildung im Vergleich: Theoretische und pragmatische Aspekte.* Swetlana Mengel (Hrsg.), серия Reihe Slavica varia Hagensi. Münster, 2014. С. 417–435.
33. *Ефремова Т.Ф.* Новый словарь русского языка. Толково-словообразовательный. М. : Рус. яз., 2000.
34. *Плотникова Л.И.* Словотворчество как феномен языковой личности (порождение, функционирование, узуализация нового слова) : дис. ... д-ра филол. наук. Белгород, 2004. URL: <http://www.disscat.com/content/slovotvorchestvo-kak-fenomen-zykovoi-lichnosti-porozhdenie-funktsionirovanie-uzualizatsiya>

### **The Internet as a Source of Linguistic Information (for Studying the Dynamics of Russian Word Formation)**

*Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology.* 2019. 57. 137–159. DOI: 10.17223/19986645/57/8

*Elena V. Petrukhina, Olga V. Dedova,* Lomonosov Moscow State University (Moscow, Russian Federation). E-mail: elena.petrukhina@gmail.com / ov.dedova@gmail.com

**Keywords:** Internet communication, search operators, quantitative data, neologisms, word-making, modern Russian.

The article deals with the analysis of the internet search option as a source of linguistic information which linguists frequently apply nowadays. The authors verified the validity of quantitative indicators of using the derivatives on the Internet that were received from Yandex and Google browsers.

The research material is the web-based project “Dar Slova” by M. Epstein (2000–2010). The project was aimed at Russian lexicon “creative updating” with the help of individual word formation and derivative integration into the Russian discourse through the Internet. Since this initiative has had a considerable impact on the Russian language, its effectiveness has been analyzed by the authors as well.

According to the research, the quantitative indicators received from simple keyword internet search are knowingly more significant than the real number of particular word use. In this context, the search results received from Yandex and Google differ significantly (the reasons for this have been examined). The special operators that can increase the search accuracy are considered in the article. Taking these factors into consideration makes the analysis more relevant. Many other problems have been revealed during the analysis, for instance, the cases connected with the degree of internet search accuracy, in particular the mismatch of search quantitative indicators fixed by the browser and the number of available links (they could be several times less). The authors concluded that, in this regard, it is preferable to use the Russian National Corpus (RNC). However, when it concerns linguistic innovations like neoderivatives the RNC does not always provide robust data: in certain cases some new non-conventional derivatives are not fixed in the RNC as well as most of the words considered. This fact shows that one of the major linguistic research methodological objectives is keyword search optimization.

The sample review of derivatives from “Dar Slova” showed that some of them, formed along the project, are not requested at all (for example, the words with the root -lub-). In our opinion, this partly “neological” effect of such an extraordinary and long-term project with regard to Russian word formation depends on the violation of cognitive and discourse terms of new word integration into the lexicon. Nevertheless, the results of this multi-year work concerning new word formation in terms of the “Dar Slova” project deserve a detailed and extensive linguistic analysis. Moreover, internet search under condition of its optimization has revealed that some of the non-conventional words from the project were created on the basis of productive models and used in appropriate discourse conditions so far in advance of the project start. Their usage helped to fill the gaps in the word formation system.

The overall conclusions of the research are the following. The internet search option for non-conventional new derivatives scanning is a very important and relevant opportunity because, on a wide variety of different genre websites involved into the search process by the browsers, it helps to fix the latest “language material” which allows to follow the history of the neologisms origin and distribution. The integrated use of different functions available today (such as various browsers, operators of query languages, search statistics data) can make search results considerably more specified.

### References

1. Krysin, L.P. (2007) Ob internatsionalizatsii fonda slovoobrazovatel'nykh morfem [On the internationalization of the fund of word-formation morphemes]. In: *Sovremennoe russkoe yazykoznanie i lingvodidaktika* [Modern Russian linguistics and linguodidactics]. 2. pp. 69–72.
2. Petrukhina, E.V. (ed.) (2010) *Novye yavleniya v slavyanskom slovoobrazovanii: sistema i funkcionirovanie: Doklady XI Mezhdunarodnoy nauchnoy konferentsii Komissii po slavyanskomu slovoobrazovaniyu pri Mezhdunarodnom komitete slavistov* [New phenomena in Slavic word formation: system and functioning: Reports of the XI International Scientific Conference of the Commission on Slavic Word Formation at the International Committee of Slavists]. Moscow: Faculty of Philology, Moscow State University.
3. Ratsiburska, L.V. (ed.) (2016) *Novye tendentsii v russkom yazyke nachala XXI veka* [New trends in the Russian language of the beginning of the 21st century]. 4th ed. Moscow: Flinta.
4. Golev, N.D. (2011) Lexical functioning of word-building and the lexical-word-building system of the Russian language in the light of Internet statistics. *Vestnik Tomskogo*

*gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 1. pp. 22–31. (In Russian).

5. Golev, N.D. (2010) Poiskovye sistemy Interneta kak lingvisticheskiy istochnik (na primere resheniya nekotorykh teoreticheskikh i prikladnykh voprosov russkogo slovoobrazovaniya) [Internet search systems as a linguistic source (on the example of solving some theoretical and applied issues of Russian word formation)]. In: Petrukina, E.V. (ed.) *Novye yavleniya v slavyanskom slovoobrazovanii: sistema i funktsionirovanie: Doklady XI Mezhdunarodnoy nauchnoy konferentsii Komissii po slavyanskomu slovoobrazovaniyu pri Mezhdunarodnom komitete slavistov* [New phenomena in Slavic word formation: system and functioning: Reports of the XI International Scientific Conference of the Commission on Slavic Word Formation at the International Committee of Slavists]. Moscow: Faculty of Philology, Moscow State University. pp. 424–443.

6. Golev, N.D. (2015) Leksicheskaya realizatsiya kak funktsional'naya kharakteristika slovoobrazovatel'noy sistemy russkogo yazyka i kolichestvennye parametry ee opisaniya [Lexical representation as a functional characteristic of the derivational system of the Russian language and the quantitative parameters of its description]. In: Malygina, M.A. (ed.) *Osm' des'at'". Sbornik nauchnykh statey k 80-letiyu I.S. Ulukhanova* [Eighty. Collection of articles on the 80th anniversary of I.S. Ulukhanov]. Moscow: Azbukovnik.

7. Sierociuk, J. (ed.) (2012) *Słowotwórstwo słowiańskie: system i tekst. Prace Komisji Słowotwórczej przy Międzynarodowym Komitecie Slawistów* [Slavonic Word formation: system and text. Works of the Word-formation Commission at the International Committee of Slavists]. Poznań: Poznańskie Towarzystwo Przyjaciół Nauk. pp. 287–297.

8. Toshovich, B. (ed.) (2016) *Slovoobrazovanie i Internet* [Word formation and the Internet]. Graz: Institut für Slawistik der Karl-Franzens-Universität Graz.

9. Epshteyn, M. (2000–2010) *Dar slova. Proektivnyy leksikon* [Gift of word. Projective Lexicon]. [Online] Available from: <http://www.emory.edu/INTELNET/dar0.html>.

10. Epshteyn, M. (2000) Slovo kak proizvedenie. O zhanre odnosloviya [The word as a work. On the genre of single-wordness]. *Novyy mir*. 9. [Online] Available from: [http://magazines.russ.ru/novyi\\_mi/2000/9/epsh.html](http://magazines.russ.ru/novyi_mi/2000/9/epsh.html).

11. Epshteyn, M. (2000) O budushchem yazyka [On the future of the language]. *Znaniya*. 9.

12. Epshteyn, M. (2006) Russkiy yazyk v svete tvorcheskoy filologii razyskaniya [Russian in the light of creative philology of research]. *Znaniya*. 1. [Online] Available from: <http://magazines.russ.ru/znaniya/2006/1/ep13.html>.

13. Ivanov, L.Yu. (2000) *Yazyk interneta: zametki lingvista* [Internet language: notes of a linguist]. [Online] Available from: <http://www.faq-www.ru/lingv.htm>.

14. Crystal, D. (2006) *Language and the Internet*. 2nd ed. Cambridge: Cambridge University Press.

15. Dedova, O.V. (2010) The language of the Internet. *Vestnik Moskovskogo universiteta. Ser. 9, Filologiya – Moscow State University Bulletin. Series 9. Philology*. 3. pp. 25–38. (In Russian).

16. Akhupkin, Ya.E. & Rakhilin, E.V. (eds) (2014) *Sovremennyy russkiy yazyk v internete* [Modern Russian language on the Internet]. Moscow: Yazyki slavyanskoj kul'tury.

17. Toshovich, B. (2015) *Internet-stilistika* [Internet stylistics]. Moscow: FLINTA: Nauka.

18. Krongauz, M.A. (ed.) (2016) *Slovar' yazyka interneta* [Dictionary of the Internet language]. Moscow: AST-Press.

19. Zemskaya, E.A. (ed.) (1973) *Russkaya razgovornaya rech'* [Russian colloquial speech]. Moscow: Nauka.

20. Shcherba, L.V. (1974) *Yazykovaya sistema i rechevaya deyatel'nost'* [Language system and speech activity]. Leningrad: Nauka.

21. Koshkareva, N.B. (2014) Tundrovyy, tundryanóy, tundrennyy ili \*tundrennyy, ili kak Internet sposobstvuet poyavleniyu i zakrepleniyu isklyucheniy [Tundrovyy, tundryanóy, tundrennyy or \*tundrennyy, or How the Internet contributes to the emergence and

consolidation of exceptions]. In: *Internet kak istochnik lingvisticheskoy informatsii* [The Internet as a source of linguistic information]. Biysk: AGAO.

22. Lutovinova, O.V. (2009) *Lingvokul'turologicheskie kharakteristiki virtual'nogo diskursa* [Linguocultural features of virtual discourse]. Volgograd: VGPU "Peremena".

23. Safronov, A. (2016) *Poisk Yandeksa: Kak nayti luchshie otvety* [Yandex Search: How to find the best answers]. [Online] Available from: <https://www.searchengines.ru/poisk-yandeksa.html>.

24. Zemskaya, E.A. (1992) *Slovoobrazovanie kak deyatel'nost'* [Word formation as an activity]. Moscow: KomKniga.

25. Zemskaya, E.A. (1996) Aktivnye protsessy sovremennogo slovooproizvodstva [Active processes of modern word production]. In: Vorontsova, V.L. et al. *Russkiy yazyk kontsa XX stoletiya (1985–1995)* [Russian language of the end of the twentieth century (1985–1995)]. Moscow: Yazyki russkoy kul'tury.

26. Ulukhanov, I.S. (1996) *Edinitsy slovoobrazovatel'noy sistemy russkogo yazyka i ikh leksicheskaya realizatsiya* [Units of the word-formation system of the Russian language and their lexical representation]. Moscow: LKI.

27. Ulukhanov, I.S. (2005) *Motivatsiya v slovoobrazovatel'noy sisteme russkogo yazyka* [Motivation in the Russian derivation system]. Moscow: Azbukovnik.

28. Kubryakova, E.S. (2010) Rol' analogii v porozhdenii novykh proizvodnykh slov [The role of analogy in generating new derivative words]. In: Petrukhina, E.V. (ed.) *Novye yavleniya v slavyanskom slovoobrazovanii: sistema i funktsionirovanie: Doklady XI Mezhdunarodnoy nauchnoy konferentsii Komissii po slavyanskomu slovoobrazovaniyu pri Mezhdunarodnom komitete slavistov* [New phenomena in Slavic word formation: system and functioning: Reports of the XI International Scientific Conference of the Commission on Slavic Word Formation at the International Committee of Slavists]. Moscow: Faculty of Philology, Moscow State University. pp. 14–25.

29. Popova, T.V. (2005) *Russkaya neologiya i neografiya* [Russian Neology and Neography]. Yekaterinburg: UGTU–UPI.

30. Pakhomova, M.A. (2013) Occasional Words and Occasionalism Dictionaries. *Vestnik MGGU im. M.A. Sholokhova. Seriya "Filologicheskie nauki"*. 3. (In Russian).

31. Kubryakova, E.S. (2004) *Yazyk i znanie. Na puti polucheniya znaniy o yazyke: chasti rechi s kognitivnoy tochki zreniya. Rol' yazyka v poznanii mira* [Language and Knowledge. On the way to gaining knowledge about the language: parts of speech from a cognitive point of view. The role of language in cognizing the world]. Moscow: Yazyki slavyanskoy kul'tury.

32. Petrukhina, E.V. (2014) *Obrazovanie novykh slov v russkom yazyke: teoreticheskie aspekty i kognitivno-diskursivnyy analiz* [The formation of new words in the Russian language: theoretical aspects and cognitive-discursive analysis]. In: Mengel, S. (ed.) *Slavische Wortbildung im Vergleich: Theoretische und pragmatische Aspekte* [Slavic word formation in comparison: Theoretical and pragmatic aspects]. Reihe Slavica varia Halensi. 12. Münster: LIT Münster. pp. 417–435.

33. Efremova, T.F. (2000) *Novyy slovar' russkogo yazyka. Tolkovo-slovoobrazovatel'nyy* [The New Dictionary of the Russian Language. Explanation. Word-Building]. Moscow: Russkiy yazyk.

34. Plotnikova, L.I. (2004) *Slovtvorchestvo kak fenomen yazykovoy lichnosti (porozhdenie, funktsionirovanie, uzualizatsiya novogo slova)* [Word-making as a phenomenon of linguistic personality (generation, functioning, usualization of a new word)]. Philology Dr. Diss. Belgorod. [Online] Available from: <http://www.dissercat.com/content/slovtvorchestvo-kak-fenomen-yazykovoi-lichnosti-porozhdenie-funktsionirovanie-uzualizatsiya>.