

Елена Вячеславовна Маринова

Россия, Нижегородский государственный университет им. Н. И. Лобачевского

Гносеологические возможности интернет-поисковых материалов в лингвистических исследованиях

Ключевые слова: интернет-поисковые материалы, эмпирическая база лингвистического исследования, окказиональное vs. узуальное, регулярность, частотность, вариантность слова

Key words: Internet search materials, empirical base of the linguistic research, occasional vs. regular, regularity, frequency, word variation

Abstract

In the article the question of Internet search materials as the empirical basis of linguistic research is considered. The author demonstrates the possibilities that this source provides for the study of occasional use of words, for defining regularity and frequency of the language fact and for the description of word variation.

В последние десять лет в исследованиях современных языковых процессов все чаще стали использоваться так называемые интернет-поисковые материалы. В научной речи появилась даже аббревиатура ИПМ, сокращенно обозначающая этот неисчерпаемый документальный источник. По нашим наблюдениям, он начинает конкурировать или, по крайней мере, использоваться наравне, с такими современными источниками лингвистических исследований, как корпуса (Национальный корпус русского языка, например) и электронные базы данных («Интегрум»). Однако использование поисковых систем Интернета («Яндекс», «Гугл» и т. п.) в качестве эмпирической базы для лингвистических исследований имеет свои особенности – и по сравнению с корпусами, и по сравнению с электронными базами. И это вполне объяснимо, ведь поисковые системы, согласно первоначальному замыслу, предназначены для поиска **неструктурированной документальной** информации на сайтах Всемирной паутины, но не для фиксации языковых фактов, речевых произведений, как в корпусе. Что же касается электронной базы «Интегрум»,

так же, как поисковики, предназначенной для работы с информацией (исначально «Интегрум» разрабатывался для бизнесменов [Никипорец-Танигава (ред.) 2006]), то в ней, поскольку это все-таки база, информация хранится, резервируется, т. е. наряду с добавлением новых текстовых источников информации, старые не исчезают, в то время как в поисковых системах, при постоянном обновлении информации, не предполагается ее хранение, хотя некоторая статическая коллекция документов имеется.

Итак, зачем, в каких целях лингвисты используют поисковики? Почему обращаются к ним? Чего недостает в других современных источниках лингвистической информации?

На наш взгляд, обращение к ИПМ помогает решить многие исследовательские задачи. Обратимся к некоторым из них.

1) ИПМ позволяют подтвердить или опровергнуть субъективные представления исследователя об **окказиональности** или, напротив, узуальности какого-либо языкового факта.

В последнее время сама категория окказиональности требует, по-видимому, пересмотра и переосмысления, поскольку отдельные окказионализмы, вопреки традиционному представлению об окказиональном слове как слове, «живущем» только в индивидуально-авторском контексте, воспроизводятся подобно узуальным словам.

Так, мы проследили условия употребления слова *аптечество*, которое встретилось нам в рекламном тексте (в новейших словарях русского языка оно не зафиксировано; отсутствует оно и в словаре с самым большим на сегодняшний день – ок. 200 тыс. единиц – словником [см. Лопатин 2013]). По своему назначению это слово эргоним (коммерческое название сети аптек), а по происхождению – **окказионализм**, в том смысле, что слово придумано для конкретного случая (*ad hoc*), с нарушением, как и подобает настоящему, полноценному окказионализму, узуальной (языковой) словообразовательной модели: собирательное существительное образовано от неодушевленного существительного, а по модели – производящее слово должно быть одушевленным (ср. *студенчество*, *купечество* и т. п.). Но это только по происхождению. Если же судить о статусе этого слова по его воспроизводимости в других контекстах, то в его окказиональности уже можно усомниться. Слово *аптечество* используется в СМИ, на форумах, встречается даже в поэтических текстах (шутливого содержания). Причем, как показывает большинство выбранных поисковиком сайтов, слово используется как апеллатив в значении ‘фармацевтика; собирательное к *аптека*’, часто с аллюзией на прецедентные тексты, в которых обыгрывается созвучие слова со словом *отечество* (*Аптечество в опасности!* и т. п.).

Эти и другие случаи употребления окказионального по происхождению слова свидетельствуют о том, что наши представления о степени окказиональности/узуальности слова нуждаются в уточнении, тем более что в новой

речевой реальности формируется особый тип окказиональных по происхождению слов (мы бы назвали их **массовыми окказионализмами**), более или менее свободно воспроизводящихся в узусе.

2) С помощью ИПМ можно уточнить данные о регулярности или, напротив, нерегулярности какого-либо языкового феномена.

Например, исследуя класс аналитических прилагательных иноязычного происхождения в современном русском языке, мы обратили внимание на тот факт, что одни аналиты регулярны в речи, т. е. употребляются в сочетании более чем с одним именем существительным (*аптека онлайн*, *библиотека онлайн*, *заказ онлайн* и т. п.). Другие же, напротив, представляют собой **уникальные** единицы, т. е. употребляющиеся в сочетании только с одним существительным: *реалити*- 'показывающее отношения между реальными людьми' (*реалити-шоу*); *ток*- 'разговорное' (*ток-шоу*); *хард*- 'тяжелый' (*хард-рок*) и др.

В то же время граница между регулярными и уникальными иноязычными определителями очень подвижная, и уникальность аналита может быть переходящим явлением. Например, прилагательное *фри* долгое время употреблялось только в сочетании с существительным *картофель* (или его разговорным вариантом *картошка*), однако в XXI в., судя по данным ИПМ, слово заметно расширило свою сочетаемость (*шарики фри*, *судак фри*, «*Лето-фри*» – о блюде, *кальмары-фри*, *лук «фри»*, *креветки фри*, *говядина фри*, *пирожки фри*) и изменило позицию по отношению к определяемому существительному – ранее только в постпозиции (*фри-выпечка*, *фри-меню* и др.).

Аналогичные наблюдения были получены и в отношении слова *смайт* и *инстант*, ранее употреблявшихся только в сочетании с одним существительным – *смайт-карта*, *кофе инстант* [подробнее см.: Маринова 2008: 248–250].

Отметим, что регулярность того или иного языкового факта можно уточнить, используя еще один параметр употребительности слова – его **частотность**. Этот параметр также можно исследовать с помощью ИПМ. М. А. Кронгауз, например, в книге *Самоучитель албанского* ссылается на данные двух специальных сервисов поисковых машин («Гугла» – books.google.com/ngrams и «Яндекса» – «Пульс блогосферы»), выстраивающих графики частоты слова в заданный отрезок времени [Кронгауз 2013].

Наблюдая за частотностью слова, исследователь может обнаружить еще один параметр слова – полифункциональность (способность употребляться в роли разных частей речи). В последнее время класс полифункциональных слов расширяется за счет новых заимствований. Данные ИПМ помогли нам сделать предположение о том, что почти каждое новое иноязычное слово (существительное) способно выступать в роли аналитического прилагательного в составе очень распространенной в современной русской речи модели типа *бизнес-встреча*, *пиар-ход*, *интернет-кафе*. Вот несколько примеров:

фэнтези-игра, фрисби-фестиваль, бэби-доктор, попкорн-автомат, боулинг-клуб, ноу-хау центр, спрей-уход, файл-сервер, промоушн-бизнес. И с постпозицией: *обои хай-фай, индустрия фастфуд* и т. п. Отсюда следует, что в оценке частотности слова необходимо учитывать и его морфологическую функцию.

3) С помощью ИПМ можно подтвердить или опровергнуть гипотезу о существовании так называемого потенциального слова. В отличие от окказионального слова, факт употребления которого все же известен исследователю, в случае с потенциальным словом исследователь имеет дело с предположением о возможном слове, т. е. слове, образованном по имеющейся в языке модели, но «вживую» еще не зафиксированном на момент исследования.

Приведем пример. В наших работах по адаптации иноязычных неологизмов много внимания уделялось группе англицизмов с нетрадиционным для русского языка исходом на *-ин, -жн* (до конца XX в. имен существительных, оканчивающихся такими звуковыми сочетаниями, не было, о чем свидетельствуют данные словарей [напр.: Зализняк 1987; Обратный... 1974]). Это существительные типа *ресепин, промоушн, фьюжн*. По нашим первоначальным данным (2007 г.), дериватов такие существительные не образуют из-за непривычности исхода (по этой же причине задерживается нередко и словоизменение этих слов: ср. *девушка на ресепин* и *девушка на ресепине*).

По прошествии трех лет мы вновь обратились к этим словам и проверили наличие производных. Гипотетически мы предположили существование относительных прилагательных от каждого слова анализируемой группы. Например, от слова *экишн*, с учетом его фонетического варианта *экишен*, возможны прилагательные с суффиксом *-ск-* (*экишнский* и *экишенский*) или *-овск-* (*экишовский* и *экишеновский*), т. е. четыре потенциальных прилагательных. Последовательно, одно за другим, «вбивая» каждое гипотетическое прилагательное в строку запроса поисковика и не очень надеясь на положительный результат, мы обнаружили все четыре варианта (!). Оказалось, что у каждого из слов на *-ин* дериваты есть, именно той структуры, какая предполагалась (частотность каждого из прилагательных приблизительно одинакова). См.: *ресепинский, ресепиенский (стол), ресепиновский, ресепишеновский (диванчик), промоушнский, промоушенский (центр), промоушновская, промоушеновская (компания), сейшнский, сейшенский, сейшиновский, сейшеновский (вариант песни), фьюжнский, фьюженский, фьюжновский, фьюженовский (форум)* и под.

4) ИПМ в значительной степени дополняют представление о вариантности слова. Такие данные особенно актуальны в отношении новых слов. Неологизмы могут употребляться в речи в разных модификациях, разновидностях, не нарушающих их лексического тождества. В поисковых документах отражены те типы вариантов слова, различия которых проявляются либо только в письменной речи: это орфографические и графические варианты слова, либо и в устной, и в письменной: это фонематические варианты слова

(типа *ланч* – *ленч*) и грамматические, различающиеся отдельными грамматическими признаками, например родом (*евро* – м. и ср. р.).

Так, в ИПМ широко представлены **орфографические варианты**, различающиеся наличием/отсутствием двойных согласных, слитным, отдельным или дефисным написанием и др. ИПМ позволяют увидеть конкуренцию вариантов и даже конфликт между **кодификацией** орфографической нормы в словаре и **узусом** (на сегодняшний день это самый острый конфликт в системе литературного языка). Это касается, например, варьирования слов *блоггер* и *блогер* (написания слова *блоггер* с двойной согласной повсеместны, но такие написания противоречат рекомендациям *Русского орфографического словаря* [Лопатин 2013]); аналогично: *масс-медиа* и *массмедиа* (последний вариант рекомендован в [Лопатин 2013]) и т. п.

Графические варианты слова различаются отдельными чертами письменного облика, не связанными с правилами орфографии. Различия таких вариантов зависят, например, от выбора алфавита, на котором пишется слово, – кириллица или латиница. Первоначальная стадия освоения иноязычного слова – употребление его в качестве **вкрапления**, т. е. в оригинальной графике. По мнению некоторых кодификаторов, отдельные слова не имеют кириллического варианта, вследствие этого они подаются в толковых словарях, обычно в приложении к основному корпусу, в написании латиницей. Так произошло со словом *user-friendly* ‘дружественный к пользователю’. Этот англицизм, в его оригинальной графике, зафиксирован в толковом словаре русского языка (парадокс!) [Скляревская 2006], однако ИПМ показывают, что кириллические варианты широко используются, см.: *юзерфрендли интерфейс*; *юзер-френдли программа* (возможно, переход на кириллицу произошел позже, после словарной фиксации слова).

Отражается в текстах ИПМ и **грамматическая вариантность** слов. На сегодняшний день наиболее частотны варианты несклоняемых имен существительных, различающиеся родом. Наблюдения показывают, что многие несклоняемые неодушевленные существительные, относящиеся к новой иноязычной лексике, употребляются в разных родовых значениях, причем слово употребляется либо как существительное, относящееся к **двум грамматическим родам** (*биеннале*, *джакузи*, *сити*), либо – как существительное, относящееся к **трем грамматическим родам**. К последним, например, относится слово *хэндс-фри* ‘устройство, предназначенное для «громкой связи» по мобильному телефону’, см.: *Подскажите плиз, работает ли какой хэндс фри с панасоником?* – из речи посетителя форума; *Есть штатная хэндс-фри* – из объявления; *Хэндс-фри новое* – объявление о продаже; *Спешите приобрести замечательное хэндс-фри*. Можно отметить также пример рефлексии говорящего по поводу рода этого техницизма, см.: *Кто знает: можно ли в Москве найти такой проводок и подойдет ли эта (этот) хэндс-фри к gsm аппарату?* (из речи посетителя форума). Как существительное всех

трех родов употребляются и слова *барбекю* (*Вошедшее в моду барбекю; Романтический барбекю; Вес каменной барбекю*), *фрисби* (*Купил новую фрисби; Модный фрисби; Запустил свое фрисби*) и др. В 90-е гг. XX в. подобные родовые варианты отмечались и у слова *евро* (*евро падала* [Буцева 2009], *евро вырос, одно евро*).

По нашим данным, среди новых заимствований, пополнивших класс несклоняемых неодушевленных существительных (всего – более 80 ед.), **около 50% слов не относятся к существительным среднего рода (как должно быть по правилу) или же варьируются**. Род имен существительных все чаще формируется по принципу смысловой аналогии, а именно «перенимается» род слова-эквивалента или близкого по значению слова русского языка (*биеннале* ж. р. и *выставка* ж. р.; *джакузи* ж. р. и *ванна* ж. р.; *гран-при* м. р. и *приз* м. р.; *евро* м. р. и *доллар, рубль* м. р.; и др.). Этот же принцип – принцип аналогии – действует и в оформлении рода **всех**, по данным ИПМ, обозначений разновидностей кофейных напитков – *американо, эспрессо, капучино, мокко, мокачино, латте, ристретто, глясе* и др.: несмотря на то, что они представляют собой несклоняемые существительные на гласный, они употребляются с определениями, стоящими в форме мужского рода (по аналогии с родом слова-гиперонима *кофе*). Например, с определениями *горячий, классический, лучший* и т. п. (определения вводились в строку запроса гипотетически). Разумеется, на поиски таких примеров употребления (прил. + сущ.) в устной речи потребовалось бы несколько лет. Таким образом, благодаря ИПМ сокращается время с момента выдвижения гипотезы до ее верификации.

Еще один тип грамматической вариантности можно проследить в ИПМ – это варьирование имени существительного по признаку склоняемости/несклоняемости. Ср., например: *Производители фастфуда* и *Индустрия фастфуда*; *Интервью с суперстар* и *Фильм голливудского суперстара*; *Время для прам-тайм* или для *прам-тайма*; *По данным Гринпис* и *Представители Гринписа*; *Прелести ол-инклюзив* и *Никакого ол-инклюзива*; *Победить в праймериз* и *Победить в праймеризе*; *Искусство фэн-шуй* и *Энциклопедия фэн-шуйя*; *Школа бонсай* и *Школа бонсяя* и др.

Наблюдения за этим типом вариантности имен существительных очень важны, так как в фактах отсутствия склонения у существительных на согласный, традиционно склоняемых, можно усмотреть влияние тенденции к аналитизму на этом, когда-то сильном участке языка. В то же время в большинстве случаев среди новых иноязычных существительных на согласный, варьирующихся с точки зрения словоизменения, отмечается устранение вариантности в пользу склоняемого варианта (так, выражения типа *подключиться к Интернету* вытеснили бытовавшее ранее *подключиться к Интернет*). Кстати, динамику в языковых изменениях также можно проследить с помощью ИПМ.

Таковы познавательные возможности ИПМ в лингвистических исследованиях. Следует добавить, что для большей степени верифицируемости этих

данных необходимо учитывать жанровую специфику электронного текста и различать исконные электронные жанры, имитирующие живую разговорную речь (блог, чат, пост) vs. жанры традиционной письменной речи, существующие на электронном носителе (тексты ЭСМИ; деловая переписка по имейлу). Потенциальные слова, например, могут встречаться в некодифицированной речи и отсутствовать в речи кодифицированной.

Безусловно, возможности ИПМ ограничены поисками фактов лишь некоторых языковых уровней и участков языка (лексика, словообразование, отчасти грамматика, орфография). Поэтому другие, как новые, так и традиционные, источники лингвистических исследований также остаются актуальными.

Литература

- Буцева Т. Н. (ред.), 2009, *Новые слова и значения: Словарь-справочник по материалам прессы и литературы 90-х гг.*, т. I: А–К, Санкт-Петербург: Дмитрий Буланин.
- Зализняк А. А., 1987, *Грамматический словарь русского языка*, 3-е изд., Москва: Русский язык.
- Кронгауз М. А., 2013, *Самоучитель албанского*, Москва: АСТ CORPUS.
- Лопатин В. В. (ред.), 2013, *Русский орфографический словарь*, Москва: АСТ-ПРЕСС КНИГА.
- Маринова Е. В., 2008, *Иноязычные слова в русской речи конца XX – начала XXI в.: проблемы освоения и функционирования*, Москва: ЭЛПИС.
- Никипорец-Танигава Г. (ред.), 2006, *Integrum: точные методы и гуманитарные науки*, Москва: Летний сад.
- Обратный... 1974: *Обратный словарь русского языка*, Москва: Советская энциклопедия.
- Склярская Г. Н. (ред.), 2006, *Толковый словарь русского языка начала XXI в. Актуальная лексика*, Москва: Эксмо.