

Статистическая оценка функциональных свойств лексики по материалам интернета

Www statistical estimation of the functional properties of lexical items

Беликов В. И.

Институт русского языка им. В. В. Виноградова РАН, Москва

Ахметова М. В.

«Живая старина», Москва

В докладе рассматриваются возможности использовать статистику числа сайтов для объективной оценки устаревания лексики, ее территориального распространения и стилистического статуса. Индикатором функциональных различий лексических единиц служат расхождения в частоте их появления в текстовых массивах разных типов (классическая vs. сетевая литература, официальные тексты, блоги и т. п.), а также в сегментах Интернета с различной доменной или территориальной привязкой.

Принято считать, что норма, зафиксированная в толковых словарях, отражает немаркированный узус носителей литературного языка. Однако в действительности «кодификаторы ориентируются в первую очередь на собственный узус, во вторую — на узус своего круга, но лишь настолько, насколько этот узус пассивно знаком самим лексикографам <...> Словоупотребление “социально неблизких” высокообразованных слоев населения на узус лексикографов не влияет и никак не учитывается» [Беликов, в печати₁: 343–344]. То же касается жаргона и других типов ненормативной лексики. Остановимся на фактах, очевидных при объективном анализе реального узуса, но плохо соотносящихся со словарными описаниями лексики и фразеологии. Цель нашей статьи не в сборе очередной коллекции ошибочных интерпретаций, а в том, чтобы показать, что в настоящее время существуют легкодоступные и достаточно простые способы объективного выявления разнородных функциональных свойств лексики.

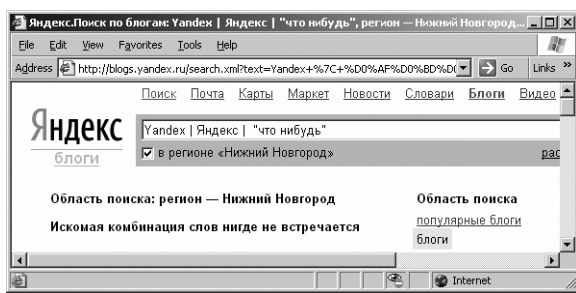
В силу ограниченности объема мы не будем типологизировать функциональные свойства лексики, лишь поясним, что речь идет о любых нетривиальных грамматических и стилистических особенностях слова или фразеологизма, упоминание которых оказалась бы полезным в словарной статье (оговоримся, что не всякая такая информация реально содержится в словарях). В рамках небольшой статьи удастся доказательно остановиться на интернет-верификации лишь некоторых типов функциональных свойств лексических единиц, так что текст невольно будет иметь несколько очерковый характер.

Охарактеризуем материал, **текстовые массивы Интернета**. Хорошим подспорьем для разного рода лингвистических разысканий стал Национальный корпус русского языка, но для детального исследования лексики его объема часто не хватает. Выручают некоторые сегменты Интернета, которые корпусами можно называть лишь метафорически. Корпуса формируются не стихийно, а создаются намеренно; их параметры задаются и контролируются, технические затруднения при поиске носят случайный характер и поисковые возможности при развитии корпуса могут меняться лишь в лучшую сторону. Не то с текстовыми массивами. Здесь мы знаем лишь самые общие характеристики содержащихся в них текстов, однако и их часто вполне достаточно. Важнейшими текстовыми массивами для работы с лексикой оказываются Библиотека Максима Мошкова (БМ) — большое собрание литературных и стилистически близких к ним текстов, и стихийно формирующаяся русскоязычная блогосфера.

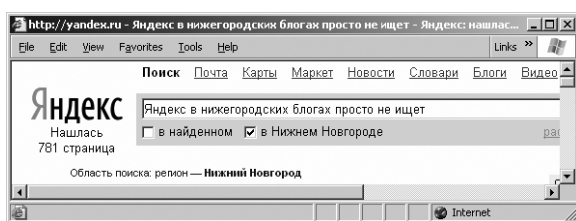
Важным достоинством БМ является ее разделение на несколько подмассивов, стилистика которых существенно различна. В первую очередь значимо противопоставление разделов «Собрание классики» (az.lib.ru), «Современная русская проза» (lib.ru/PROZA) и «Самиздат» (zhurnal.lib.ru). «Современность» в соответствующем разделе БМ трактуется достаточно широко и охватывает значительную часть советского периода. Собранием действительно современного профессионального литературного творчества является другой сегмент Интернета — «Журнальный зал» (magazines.russ.ru), где сосредоточены журнальные публикации с 1990-х гг.

«Самиздат» представляет собой очень большое по объему собрание современных самостоятельных произведений разного жанра, обычно невысоких художественных достоинств; многие авторы имеют достаточно смутные представления о литературной норме, доля разговорной и просторечной лексики в авторском тексте этого массива заметно выше, чем в только что упомянутых, поэтому в целом лексикон «Самиздата» близок к современному разговорному узусу.

Язык блогов во многом является отражением повседневного молодежного словоупотребления. Теоретически Яндекс допускает поиск в блогах с заданием отдельных параметров и их комбинаций: региона, пола и возраста блоггеров, а также с выделением конкретного фрагмента блогосферы (livejournal.com, diary.ru и т. п.). На практике же Яндексу не удаются поиски в блогах одного из крупнейших городов России.



Столь же безнадежны поиски чего бы то ни было у блоггеров Великого Новгорода, Нижнего Тагила, Нового Уренгоя и других городов, пишущихся с пробелом. Проблема теоретически вряд ли серьезна, поскольку при общем (не блоговом) поиске, где регион также можно ограничить, Яндекс без труда справляется с двусловными топонимами:



Не особенно дружественная политика Яндекса по отношению к любителям лексической статистики постоянно ужесточается. С самого начала появления поиска по блогам Яндекс вынуждал пользователей получать информацию блоками по 10 записей, уже года два, как он отказывается показывать 1001-ю и последующие найденные единицы, а с лета 2008 г. поиск с разделением по возрасту стал невозможен — со второй страницы результатов Яндекс сбивается с ограничения по возрасту и выдает лишь общий результат.

Не так уж редко кодифицированные **грамматические характеристики** отдельных единиц противоречат повседневному узусу.

Во всех толковых словарях фигурирует слово *корректив* (м. р.). Анализ текстовых массивов Интернета показывает, что если в 1930-е гг. и ранее вносился *корректив*, то к середине XX в. вносятся стали *коррективы*, в дальнейшем нейтрализация по числу способствовала смене рода на женский. К настоящему времени использование этого слова в единственном числе мужского рода стало безусловной архаикой, сейчас в словаре стоило бы писать: «**корректива**, ж. (реже **корректив**, м.), обычно мн. ...» [см. подробнее: Беликов, в печати₂].

Глаголы *лазить* (*лажу*, *лазишь*,...) и *лазать* (*лазаю*, *лазаешь*,...) всеми словарями признаются синонимичными и описываются обычно в одной статье; в московском словаре Шведовой [ТСРЯ] второй снабжается пометой *разг.*, что имеет естественное объяснение: «на слух» *лазать* в Москве говорят заметно реже, чем в Петербурге. Но в действительности положение с отдельными словоформами этих глаголов различно. Судя по блогам (2007–2008), в петербургском узусе преобладают личные формы «от *лазать*», соотношение: *лазаешь/лазишь* — 52/30, *лазает/лазит* — 204/112; но с заметно более частотными инфинитивами положение обратное: 797/1120. В московских блогах преобладание «строго нормативного» инфинитива выражено очень явно: за IV квартал 2008 соотношение: *лазить/лазать* составило 696/167, но с личными формами происходят странные вещи, за 2007–2008 гг. *лазаешь/лазишь* — 115/140, а *лазает/лазит* — 647/590. «Правильная форма» *лажу* быстро устаревает и не всегда используется даже в старшем поколении; показательна реакция одного известного ученого (не русиста), чл.-корр. РАН: «Говорю *лазию*, пишу *лазаю*». — «А *лажу*?» — «Ну, это просторечие, *Из кичмана не вылажу* какое-то».

У глагола *определиться* в ОШ было значение «определить своё местонахождение, положение» с пометой *спец.* и примером *Лётчик определился с помощью приборов*; в обновленном издании [ТСРЯ] это значение дополнено: «...; вообще установить, решить что-н. для себя» и снабжено речением *О. в своих планах, целях, намерениях*. Это значение ввел в общерусский обиход М. С. Горбачев, у которого стандартным управлением было *определиться по чему-л.*, однако в общем употреблении сразу стали конкурировать разнообразные модели (тип управления иногда связан с зависимым существительным, но во многих случаях взаимозаменяем [Беликов 2004]). В бумажных СМИ в самом начале 1990-х основной стала модель *определиться в чём-л.*, но уже к середине 1990-х она стала вытесняться моделью *определиться с чем-л.*, которая к настоящему времени оказалась вне конкуренции. Это легко выявляется в блогосфере, поиск на *определиться* /+4 (*планы* | *цели* | *намерения*) по 2008 г. выявил 157 случаев *определиться с* и 25 *определиться в*. «Исконное» *определиться по* с этими словами в блогах не встретилось, хотя в СМИ оно еще попадает: *Госу-*

дарству российскому необходимо четко определиться по своим внешнеполитическим целям и задачам («Время новостей»; 4.07.2006); Г. Бержуашвили считает необходимым для обоих государств [Грузии и Украины] определиться по своим интеграционным планам («Интерфакс», 2.04.2007). С другими управляемыми словами соотношение оказывается иным, но при возможности конкуренции «с-управление» заведомо преобладает, соотношение с- и в-управления при аналогичном поиске с набором понятия | термины | определения составляет 56/23, с теми же словами в ед. ч. — 22/2¹. Указание на тип управления — необходимая часть словарной статьи глагола, но при этом должны быть проиллюстрированы все возможности, при ограничениях на объем единственно упоминаемым может быть только наиболее универсальное² и частотное. Резон для соединения «старого» и «нового» словоупотреблений в одно значение неясен; причиной могли, конечно, быть интроспективно полученные конструкты типа *Лётчик определился в своих планах с помощью приборов или *Лётчик определился со своими целями по приборам, но среди многих сотен просмотренных примеров сходные структуры не встретились.

Обращение к интернет-массивам позволяет довольно точно определить время и темпы конкретных словарных изменений. Изменения эти могут иметь разный характер: лексическая единица может «просто» устареть и выйти из употребления, может, наоборот, проявить территориальную или социальную экспансию, а может замениться другой, внешне сходной. Словарями эти процессы фиксируются далеко не всегда адекватно.

В словаре под редакцией Д. Н. Ушакова имелись слова *бестоварье* 'недостаток, отсутствие товаров' с пометой *газет.*, *ведьмак* 'знахарь, колдун, оборотень' с пометой *обл.* и *решешник* 'учебное пособие, содержащее подробные решения задач, помещенных в каком-н. задачнике; ключ к задачнику' с пометой *школьн. арг.* Последнее слово позднее в толковые словари не включалось, то ли как устаревшее, то ли как имеющее слишком узкогрупповое распространение. *Ведьмак* в достаточно объемном (130 тыс. слов) БТС отсутствует, но в новом БАСе представлен, по-прежнему с пометой *обл.* *А бестоварье* в двух названных академических словарях получило помету *разг.* Между тем *бестоварье* в БМ встречается лишь в 8 текстах, датированных 1917–1928 годом (три художественных: «Разгром» Фадеева, «Третья столица» Пильняка, «Возвращение Мюнхгаузена» С. Кржижановского), а также в два современных, но в обоих случаях цитируются документы той же эпохи. В блогосфере это «разговорное слово» появилось лишь четырежды, причем толь-

ко в цитатах 1920-х гг. Что касается *ведьмака*, то обнаружившаяся в последние десятилетия всенародная любовь ко всему мистическому и потустороннему, сделала это слово чрезвычайно частотным: если в разделе «Классика» БМ оно встретилось лишь в 4 произведениях, то в современной литературной периодике («Журнальный зал» по 2008 г. вкл.) — в 19 текстах; в повседневном узусе *ведьмаки* упоминаются чаще многих общеизвестных политиков: в блогах за первую неделю декабря 2008 г. *ведьмаки* встретились 203 раза. Разумеется, никакого «областного» налета у этого слова уже нет. *Решешник* же достаточно давно стал упоминаться вне связи со школой, ср. у Аркадия Штейнберга о рыбаке: *Он перелистывает, как решешник, / Волну, волну... Ответа нет как нет* («Взморье», 1932); позднее это слово метафорически использовалось в стихах Борисом Слуцким и Сашей Соколовым. Что касается основного значения, то оно ежедневно по нескольку раз фигурирует в блогосфере, причем и в связи с вузовским обучением.

В словарях до сих пор отсутствует наречие *пешкодралом* и до самого последнего времени не было фразеологизма (*не*) *брать в голову*,³ устаревшее *пешедралом* и практически вышедшее из употребления (*не*) *забирать в голову* не имеют в словарях ограничительных помет. *Пешкодралом* впервые появляется в конце XIX в., в современной профессиональной прозе *пешкодралом* и *пешедралом* равночастотны, в «Самиздате» преобладание нового варианта уже вполне выражено (55/35), а в блогосфере можно говорить о его победе (по 2008 г. вкл. соотношение составляет 2114/337). Экспансия нового варианта фразеологизма была гораздо более быстрой: он появился и победил в течение второй половины XX в.: в БМ-классике соотношение (*не*) *забирать* и (*не*) *брать в голову*, составляет 22/1 (единственный пример с *не брать в голову* находится в относительно случайной для этого раздела литературоведческой повести о Лескове Л. А. Аннинского «Несломленный»), а в «Журнальном зале — 3/150.

Детальный анализ блогов позволяет увидеть разницу в скорости лексических изменений в различных территориальных и социальных группах. Как известно, в настоящее время происходит переход от выражений типа как *не_фига* *делать, мне это по_фигу* к как *не_фиг* *делать, мне на это по_фиг*⁴; знак «_» символизирует здесь нерелевантность пробела в орфографической записи⁵. Подтверждается не только вполне естественное предположение, что инновация

¹ Примеры типа *определяться с семантикой основных понятий, определиться в значении базовых понятий*, естественно, в результат не включены.

² *Определиваться* в несовместимо, например, с временными отрезками, ср. *определиваться с отпуском* (*...в отпуске).

³ Впервые фиксируется в Словаре-тезаурусе... [2007].

⁴ По материалам СМИ этот процесс описан в [Беликов 2008].

⁵ В действительности носители новой модели существенно сильнее ощущают наречный характер выражения, что отражается в орфографии, ср. статистику по московским блогам 2004 г.: *по фигу*: 426, *пофигу*: 923; *по фиг*: 153, *пофиг*: 1667.

в большей степени характерна для младших возрастов, но также выявляется такой нетривиальный факт: женщины используют новую модель чаще мужчин. Проиллюстрировать это придется лишь «архивными» поисками июня—июля 2008 г., (как говорилось, выявление возрастной статистики по блогам с середины 2008 стало практически невозможным). Вот данные по Петербургу, где смена *по_фигу* на *по_фиг* идет несколько быстрее, чем в Москве.

блоггеры Петербурга	2005, весь год		2008, январь–май	
	<i>по_фиг(y)</i>	в т. ч. <i>по_фиг</i>	<i>по_фиг(y)</i>	в т. ч. <i>по_фиг</i>
все блоггеры	1388	935 (67,4 %)	2529	1847 (73,0 %)
указавшие пол:				
женщины	655	454 (69,3 %)	1602	1189 (74,2 %)
мужчины	607	397 (65,4 %)	792	546 (68,9 %)
указавшие возраст:				
11–19 лет	34	23 (67,6 %)	640	484 (75,6 %)
20–30 лет	614	406 (66,1 %)	549	381 (69,4 %)
старше 30 лет	95	57 (60,0 %)	117	70 (59,8 %)

Среди прочего эта таблица иллюстрирует неизвестные факты о быстрых переменах в половозрастной структуре блоггерского сообщества: за неполных три года доля младшей возрастной группы выросла с 5% до половины, а женщины, лишь незначительно превосходявшие мужчин, стали составлять 2/3 (при анализе стилистически различной лексики структура блоггеров может меняться существенно, но эти данные близки к средним показателям для единиц повседневного узуса). И не следует забывать, что возраст всегда указывает лишь около половины блоггеров, причем среди неуказавших доля женщин выше, чем в блогосфере в целом. Так — в Петербурге в конкретные годы; в других местах и в другое время состав блоггеров будет иным.

Кроме собственно лексических перемен, Интернет помогает отследить **изменения в сочетаемости**. Любители футбола заметили, что с недавних пор судьи все чаще не *назначают* пенальти, а *ставят*. В блогах обсчитывались вхождения глаголов *ставить/поставить* и *назначать/назначить* на расстоянии «/3» от *пенальти*. В 2006 г. из общего числа в 1106 таких контекстов с глаголом (*по)ставить* было 55 (5%), в 2007 — 7% от 1371, в 2008 — 15% от 2079. В течение 2008 г. поквартальный рост в процентах составлял 11–15–16–18. В некоторых контекстах постановка пенальти уже преобладает, так, соположение *поставить* vs. *назначить* и «ле-

вый пенальти» менялось следующим образом: 2006 — 3:13, 2007 — 3:4, 2008 — 8:3.

Похожими темпами идет экспансия глагола (*по)ставить* в спортивной прессе. В двух наиболее массовых изданиях, «Советский Спорт» и «Спорт-Экспресс», рост соотношения глаголов (*по)ставить* и *назначить* (*назначать*) в соответствующих контекстах был следующим: 1999–2000: 4/1091; 2001–2002: 13/1005; 2003–2004: 26/963; 2005–2006: 40/895; 2007–2008: 67/720. Поначалу (*по)ставить* обычно заключалось в кавычки (ср. Гусев не «поставил» пенальти ни в наши, ни в их ворота, распределив все поровну — Советский Спорт; 31.08.1999), но в последние годы кавычки редки, новая конструкция стала использоваться в заголовках: *На Украине за «локомотивский» фол против Кержакова пенальти ставят без раздумий* («На страже Родины», СПб; 14.05.2005), *В Манчестере такие пенальти не ставили* (Советский спорт; 15.05.2008).

Выражение *левый пенальти* появляется в СМИ в конце 1990-х (самый ранний пример в базе «Интергум»: *В последнем матче, который якутяне выиграли у «Атома» из Железногорска, судьи дважды высосали левые пенальти, ни за что ни про что удалили нашего игрока с поля, но все равно «Автомобилист» сумел отстоять свое хрупкое преимущество и победил 3:2 — «Якутия», 29.11.1997), но употребляется очень редко. Статистика *левых пенальти* мизерна, но отмахнуться от нее нельзя: соотношение *поставленных* и *назначенных* по всем газетам таково: 1999–2004: 1/14, 2005–06: 5/10, 2007–08: 4/1.*

Примеры на (*по)ставить штрафной* в прессе до сих пор единичны⁶, но блоггеры так пишут все чаще: за 2005–2007 на (*поставить* | *ставить*) /1 *штрафной* находится, соответственно, 1, 2 и 3 релевантных примера, а в 2008 — уже 15 (при 113 на *назначить* /1 *штрафной*).

Выше речь шла о лексике, равномерно распространенной по русскоязычному пространству. Но существуют тысячи регионально маркированных лексических и фразеологических единиц различного стилистического статуса (нормативные, разговорные, жаргонные). Уточнения ареалов их распространения без обращения к различным сегментам Интернета невозможно.

Принято считать, что в паре синонимов *студень* / *холодец* (в знач. 'холодное кушанье из сгустившегося желеобразного мясного отвара с мелкими кусочками мяса') для Петербурга более характерно первое; это подтверждается и толковой лексикографией северной столицы: при немаркированном *студне*, *холодец* в МАСе имел помету *разг.*, а в БТС был «понижен» до *нар.-разг.* Между тем Интернет довольно отчетли-

⁶ Один из наиболее ранних: *Все-то у него идет под классические 0:0, как вдруг на 95-й ставят ему штрафной, не смертельный, но сильно неприятный — эх, не всегда ладится законтачить с последним судейским поколением!* (Футбол-review; 16.06.1999).

во указывает на явную экспансию холодца в Петербурге; в младших возрастах в этом значении холодец уже более употребим. В блогах по окт. 2008 вкл. соотношение слов *студень* и *холодец* в Москве составляло 1772:4634, в Петербурге — 1330:1026. Но при этом семантический шум для слова *холодец* в обоих городах равномерен и невелик, а за *студнем* в молодежном узусе скрываются также значения ‘студент’ и ‘студенческий билет’, при этом в Москве доля этих студенческо-жаргонных единиц сравнительно невелика, а в Петербурге составляет около половины всех студней. Результаты выборочного анализа словоупотреблений за 2005–2008 гг. в двух городах таковы:

	«студень», всего	студень ‘студент’	студень ‘сту- денческий билет’
Блоги Санкт-Петербурга:			
октябрь 2005	18	1	8
октябрь 2006	27	3	18
октябрь 2006	33	4	14
октябрь 2008	40	5	16
«всего»	118	13	56
Блоги Москвы (те же 4 месяца):			
«всего»	165	14	16

Имея в виду различия в объемах блогосферы двух городов, вполне очевидно, что центром распространения нового жаргонизма, особенно в значении ‘студенческий билет’, является Петербург.

Интернет-блоги во многих случаях являются наиболее эффективным инструментом выявления ареалов распространения регионально маркированных единиц: *чоїс* ‘любая лапша быстрого приготовления’ и *оптaрь* ‘оптовый рынок’ находятся только в омских блогах, *садоогород* ‘садово-огородное товарищество или участок в нем; используется и в официальных контекстах — почти исключительно в Удмуртии, *ссобойка* (также *собойка*) ‘набор продуктов на работу, в дорогу; школьный завтрак, взятый из дома’ — практически только в Белоруссии.

Однако в ряде случаев семантические и стилистические особенности лексической единицы таковы, что статистический анализ блогов не дает нужных результатов.

Недавний фразеологизм «с канцелярским оттенком» *мокрая печать* (поставленная непосредственно на документ, не ксерокопированная), на первый взгляд, распространен достаточно широко. Но анализ его появления в текстах, размещенных на разных доменах первого уровня, однозначно указывает на то, что фразеологизм привязан в первую очередь к Украине; вот статистика сайтов с релевантными документами⁷ за отдельные годы:

<i>Мокрая печать</i>	...–2004	2005	2006	2007	2008	всего
Всего	44	36	68	164	450	762
В домене .ru	16	12	23	53	112	216
в том числе про Украину домене .ru	10	3	14	19	41	87
В домене .ua	7	13	22	62	187	291
В домене .md						11

Как видим, фразеологизм чаще появляется в русскоязычных текстах домена .ua, а в домене .ru значительная часть релевантных текстов касается Украины.

На другом стилистическом полюсе находится детская лексика: тут анализ только блогов дает очень приблизительный результат в силу достаточно редкого спонтанного появления соответствующих единиц в блогах и неравномерности распространения блогосферы по русскоязычной территории. Но совмещение анализа блогов с непосредственным опросом в Интернете позволяет делать достаточно четкие выводы. Ср., например, распределение синонимичных именованной игры, самым распространенным названием которой является *вышибалы*: широко распространенное в недавнем прошлом название *круговая лапта* сохранилось только в Азербайджане, в обширном регионе от Красноярска до Хабаровска преобладает именование *выжгалы*, в Смоленской обл. — *высекалы* и т. п.⁸

В некоторых случаях региональное развитие блогосферы пока слишком незначительно, чтобы в Интернете можно было получить релевантную информацию. Так, слово *чeбэшка* с орфографическими вариантами ‘дом, не полностью обеспеченный коммунальными удобствами; квартира в таком доме’ (от сокр. ч/б = частично благоустроенный) встретилось в пяти газетах (30 текстов), причем только из Якутии [Ахметова 2008], но анализ блогов практически ничего не дает: за 2007–2008 гг. слово *чeбэшка* (*чэбэшка*, *чeбeшка* и т. п.) встретилось у 17 блоггеров в значении ‘черно-белая фотография, черно-белая фотопленка’ и лишь однажды в значении жилища — про якутский поселок Багатай, но у московского блоггера.

Приведенные выше примеры можно легко умножить. Интернет позволяет подтверждать или опровергать имеющиеся в толковых словарях сведения о функциональных свойствах лексики, выявлять новые, трудноуловимые традиционными методами особенности употребления слов и фразеологизмов.

Орехи словарей прошлого вполне объяснимы, но в XXI веке лексикограф не имеет права работать по старинке и манкировать легкодоступными текстовыми массивами.

⁷ Нерелевантными считались тексты с упоминанием *мокрой печати* как традиционного (не компьютерного) способа фотопечати.

⁸ См. подробнее на форуме «Городские диалекты» — <http://forum.lingvo.ru/actualthread.aspx?tid=110505>.

Литература

1. *Ахметова М. В.* Региональная вариативность терминов, связанных с городской недвижимостью (по материалам электронной базы периодики «Интегрум») // Компьютерная лингвистика и интеллектуальные технологии. Вып. 7 (14). По материалам Международной конференции Диалог 2008. М.: ИПИ РАН, 2008.
2. *БАС: Большой академический словарь русского языка* / Гл. ред. К. С. Горбачевич. — М.—СПб: Наука. Тт. 1–9. А–Медь. 2004–2007.
3. *Беликов В. И.* «... Справочники устаревают и требуют корректив» // Материалы XIV Филологических чтений имени проф. Р. Т. Гриб. Лесосибирск. (В печати₂).
4. *Беликов В. И.* Yandex как лексикографический инструмент // «Компьютерная лингвистика и интеллектуальные технологии». Труды Международной конференции Диалог'2004. М., «Наука», 2004.
5. *Беликов В. И.* Динамика утраты флексии в ползнаменательных наречных выражениях // Язык современного города. Тезисы докладов международной конференции Восьмые Шмелевские чтения. — М.: ИРЯ РАН, 2008.
6. *Беликов В. И.* Стереотипы в понимании литературной нормы // Стереотипы в языке, коммуникации и культуре. М.: РГГУ. (В печати₁).
7. *БТС: Большой толковый словарь русского языка.* / Под ред. С. А. Кузнецова. — СПб., 1998.
8. *МАС: Словарь русского языка: В 4 т.* / Под ред. А. П. Евгеньевой. — 2-е изд., испр. и доп. — М.: Рус. яз., 1981–1984.
9. *ОШ: Ожегов С. И., Шведова Н. Ю.* Толковый словарь русского языка. М., 1992; изд. 2-е, М., 1994; изд. 4-е, М., 1997.
10. *Словарь-тезаурус современной русской идиоматики: около 8 000 идиом современного русского языка* / Под ред. А. Н. Баранова, Д. О. Добровольского — М.: Мир энциклопедий Аванта+, 2007.
11. *ТСРЯ: Толковый словарь русского языка с включением сведений о происхождении слов* / Ин-т рус. яз. РАН. Отв. ред. Н. Ю. Шведова. — М.: Азбуковник, 2007.
12. *Ушаков: Толковый словарь русского языка.* Т. 1–4 / Под ред. Д. Н. Ушакова. М., 1935–1940.