

2. Даль, В. И. Пословицы русского народа / В. И. Даль [Электронный ресурс]. – Режим доступа : <http://hobbitaniya.ru/dal/dal75.php>. – Дата доступа : 03.10.18.

3. Гоголь, Н. В. Вий / Н. В. Гоголь [Электронный ресурс]. – Режим доступа : <https://ilibrary.ru/text/1002/index.html>. – Дата доступа : 02.10.18.

4. Мелетинский, Е. М. Мифологический словарь / Е. М. Мелетинский [Электронный ресурс]. – Режим доступа : <http://myths.kulichki.net/enc/item/f00/s20/a002056.shtml>. – Дата доступа : 05.10.18.

5. Достоевский, Ф. М. Хозяйка / Ф. М. Достоевский [Электронный ресурс]. – Режим доступа : <https://ilibrary.ru/text/21/p.1/index.html>. – Дата доступа : 07.10.18.

6. Энциклопедия мифологии [Электронный ресурс]. – Режим доступа : <http://godsbay.ru/orient/nuwa.html>. – Дата доступа : 07.10.18.

УДК 811.16'42:398.92

К ВОПРОСУ ОБ ИСПОЛЬЗОВАНИИ СТАТИСТИЧЕСКИХ КОРПУСНЫХ ДАННЫХ ВО ФРАЗЕОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

И. Г. Гомонова

Обосновывается необходимость использования статистических данных в ходе решения на базе корпуса текстов различных фразеологических исследовательских задач. Данное положение иллюстрируется результатами наблюдений над вхождением отдельных фразеологизмов в тексты Национального корпуса русского языка.

Ключевые слова: фразеологизм, фразеологическое исследование, корпус текстов, статистический, количественный.

The necessity of using statistical data in solving of various phraseological research problems based on corpus of texts is substantiated. This provision is illustrated by the observations on the occurrence of individual phraseological units in the texts of the Russian National Corpus.

Key words: phraseological unit, phraseological research, corpus of texts, statistical, quantitative.

При сопоставлении корпусной и традиционной лингвистики в качестве одного из отличительных признаков первой обычно называют предпочтение количественных методов. Все результаты поиска в корпусе текстов сопровождаются статистической информацией. Специализированные программы (корпусные менеджеры) позволяют получить различные статистические характеристики, связанные как с метаданными, так и с лингвистическими категориями. Статистические методы, применяемые к текстам, дают возможность количественно описать поведение различных языковых единиц, их сочетаемость с другими единицами, их распределение по текстам разных стилей, жанров, тематики, времени создания и т. д. По мнению В. П. Захарова, привлечение количественных методов к исследованию языка «позволяет существенно модифицировать представление о языковой системе и ее функционировании <...> количественная лингвистика может существенно укрепить лингвистическую теорию» [1, с. 83]. В. П. Захаров справедливо говорит о том, что «корпусно-статистический подход (разрядка наша – И. Г.) к изучению языка позволяет создавать невозможные ранее описания языка, которые можно назвать структурно-вероятностными моделями языка» [Там же].

Очевидно, что как источник достоверных статистических данных корпус текстов должен обладать достаточным объемом, сбалансированностью и богатой разметкой. Приходится с сожалением констатировать, что во многих случаях объем корпусов текстов оказывается недостаточным для получения полноценных количественных данных. Общеизвестно, что коллокации для среднечастотных и низкочастотных слов начинают выявляться только в миллиардных корпусах.

В связи с этим в последнее время распространение получила идея создания корпусов на основе ресурсов Интернета (Web as Corpus), «позволяющая сравнительно быстро и без больших трудозатрат создавать корпуса объемом несколько десятков миллиардов токенов» [1, с. 84]. Однако, решая проблему с размером корпуса, подобные источники остро ставят вопрос сбалансированности текстов. Достоверность данных страдает из-за несбалансированности таких корпусов.

Таким образом, по определению В. П. Захарова, «проблема верификации и достоверности данных, получаемых на основе статистических методов на базе корпусов – это не только проблема создания больших корпусов, но и проблема хороших и правильных корпусных данных» [Там же]. Выбирая между «большими и хорошими данными», мы делаем выбор в пользу сбалансированного и обладающего богатой экстралингвистической и лингвистической разметкой корпуса текстов, каким является Национальный корпус русского языка [2].

Проиллюстрируем далее результаты наблюдений над вхождением в тексты НКРЯ фразеологизма *медвежий угол / уголок / край* (разг.) 'захолустное, отдалённое, глухое место' [З, с. 679], полученные с использованием статистики корпуса. При этом отметим, что данная информация имеет схематичный характер и подлежит дальнейшей интерпретации при системном описании функционирования фразеологизма в текстах корпуса.

Медвежий угол – 174 документа, 218 вхождений (запрос *медвежий* на расстоянии от 1 до 3 от *угол*); из них

– 82 вхождения в форме множественного числа (запрос *медвежий* на расстоянии от 1 до 3 от *угол pl*): *Мы изложили свои взгляды на педагогическую технику, которую уже нам удалось испробовать в одном из медвежьих углов, далеко от больших педагогических дорог* [А. С. Макаренко. Педагоги пожимают плечами (1927–1935)]; *Куда, по мнению бывших коллег, должны были направлять свои рукописи десятки начинающих авторов из российских медвежьих углов и тьмутараканей?* [Ю. В. Трифонов. Записки соседа (1972) // «Дружба народов», 1989];

– 34 вхождения в форме предложного падежа с предлогом *в* (запрос *в* на расстоянии 1 от *медвежий* на расстоянии от 1 до 3 от *угол* (loc | loc2)); такие наблюдения можно сделать и по другим падежным формам: *Но оказалось, что Серебряков – мыльный пузырь, не по праву занимающий высокий пост, а живые талантливые люди, дядя Ваня и Астров, в это время гноят свою жизнь в медвежьих углах обширной неустроенной России* [К. С. Станиславский. Моя жизнь в искусстве (1925–1928)]; *В мирные времена в медвежьем углу, черт знает в каких тартарах, куда месяц скачи – не доскачешь, в маленьком армейском гарнизоне провел бы свою простую армейскую жизнь неведомый капитан Иванов, так же как и тысячи его однофамильцев с ударением на «о»* [А. В. Туркул, И. С. Лукаш. Дроздовцы в огне (1937–1948)];

– 7 вхождений с компонентом *медвежий* в форме превосходной степени, являющейся результатом окачества адъектива (запрос *самый* на расстоянии 1 от *медвежий* на расстоянии от 1 до 3 от *угол*): *В самых медвежьих углах Союза растут культурные центры* [А. Е. Ферсман. Воспоминания о камне (1940)]; *Уже через несколько дней я получал читательские поздравления с этим стихотворением из самых-самых «медвежьих углов», с кораблей, находящихся в далеком плавании* [Е. Евтушенко. Плач по цензуре // «Огонек». № 5, 1991];

– 28 вхождений данного фразеологизма, распространенного прилагательным, (1) качественным, усиливающим значение и стилисти-

ческую окраску ФЕ, или (2) относительным, выражающим признак по месту (запрос А на расстоянии 1 от *медвежий* на расстоянии от 1 до 3 от *угол*): (1) *Глухие медвежьи углы* рисовались в самом поэтическом свете, как своего рода обетованная земля [Д. Н. Мамин-Сибиряк. Отрезанный ломоть (1899)]; (2) *Вспоминал Валерий Александрович, глядя на сына, как когда-то он, молоденький шахтер, сам ухаживал за Галей, как, поженившись, прожили они душа в душу тридцать пять лет, хлебнув и квартирной неустроенности, и суровой убогости сибирских «медвежьих углов», но при этом вырастив шестерых родных детей и одного, взятого из детдома* [Наезд // «Сельская новь», 2003.09.16];

– ~ 65 вхождений данного фразеологизма, распространенного местоимением – притяжательным, указательным, неопределенным, определительным (запрос (SPRO | APRO на расстоянии 1 от *медвежий* на расстоянии от 1 до 3 от *угол*): *Возле залива застроить не могут, а что говорить про наш медвежий угол* [Д. Каралис. Мы строим дом (1987–2001)]; *Война обошла этот медвежий угол, хотя подкатывалась уже достаточно близко* [А. Битов. Азарт, или Неизбежность ненаписанного (1997–1998)]; *Многое было неумело, иногда грубовато, иногда провинциально-эстетично (последнее извинялось тем, что большинство стихов было подписано каким-то медвежьим углом Воронежской губернии), много было просто зелено – но все-таки книжка обращала на себя внимание, и в «Русской мысли» и «Аполлоне» Брюсов и Гумилев очень сочувственно о ней отозвались* [Г. В. Иванов. Петербургские зимы (1928)]; *Ехать в провинцию «на хлеба и в безопасность» было одно время сущим поветрием, и сколько глупого народа так сослепу-то очертя голову бросилось по всяким медвежьим углам, чтобы рано или поздно заметить о Петербурге, как о рае волшебном...* [А. М. Ремизов. Взвихренная Русь (1917–1924)].

Вариант *медвежий уголок* – 8 документов, 8 вхождений (запрос *медвежий* на расстоянии от 1 до 3 от *уголок*); из них

– 7 (!) вхождений в форме множественного числа: *Сюда, в подземелье, ежесекундно стекается подробная информация о том, что происходит в небе России: от самых удаленных «медвежьих уголков» до ее сердца – Москвы* [Олег Литвинов. За семью дверями. Репортаж с Центрального командного пункта ВВС // «Воздушно-космическая оборона», 2003.06.15];

– 3 вхождения в составе конструкции «*один из + Р. п.*»: *Наш Темрюк представляет в области радио один из медвежьих уголков Союза* [коллективный. По СССР // «Радио Всем», 1928];

– 1 вхождение с компонентом *медвежий* в форме превосходной степени: *Наш город представляет один из самых медвежьих уголков, в котором полновластно царит скука* [Внутренние известия. (От наших корреспондентов) // «Русское слово», 1895];

– 3 вхождения данного фразеологизма, распространенного прилагательным, (1) качественным, усиливающим значение и стилистическую окраску ФЕ, или (2) относительным, выражающим признак по месту: (1) *Жители пригородных данковских слобод – Сторожевой и Крюковки – в культурном отношении оказались не выше своих собратьев из глухих медвежьих уголков* [неизвестный. Холера. По телеграфу от наших корреспондентов // «Русское слово», 1910]; (2) *Но вот попадает А. А. Лозинский в один из российских медвежьих уголков, в г. Уральск, и в местном медицинском обществе делает «предварительное сообщение» об открытых им «новостях» медицины в статье «К вопросу о скрытых формах болотной лихорадки»* [Н. Ф. Федоровский. Гомеопатия и государство // Доклад общему собранию членов С.-Петербургского Общества самопомощи в болезнях 13 мая 1901 г. Вестник гомеопатической медицины].

Вариант *медвежий край* – 2 документа, 2 вхождения (запрос *медвежий* на расстоянии от 1 до 3 от *край*): *Тетки и отец считают себя древнего рода, самого коренного в этом лесном медвежьем крае* [П. Д. Боборыкин. Василий Теркин (1892)]; *Пришли мы сюда с бабой да с дитем. Огляделись. Медвежий край* [Б. Л. Горбатов. Большая вода (1939)].

Констатируем, что, несмотря на вариативность данного фразеологизма, в текстах НКРЯ преобладает его основной словарный вариант *медвежий угол* в форме единственного числа.

Одним из компонентов статистической характеристики текстов корпуса и входящих в них языковых единиц являются хронологические данные. Национальный корпус русского языка содержит информацию о времени создания представленных в нем текстов, о годах рождения авторов текстов и позволяет систематизировать эти данные при исследовании тех или иных языковых единиц (напр., определить, у какого автора и в каком произведении, по данным корпуса, анализируемая единица появилась впервые). Кроме того, НКРЯ содержит графики распределения языковых единиц по годам, содержащие информацию об их относительной частоте употребления за определенный год (показатель ipm ; частота ipm определяется как количество употреблений анализируемой единицы за год, разделенное на объем корпуса за этот год и умноженное на 1 миллион). Графики сопровож-

даются таблицами с абсолютными частотами употреблений за каждый год и общим числом слов в текстах за данный год. Гиперссылки в таблицах позволяют перейти к просмотру примеров из корпуса.

Хронологическая обработка составляет один из этапов в процессе установления основных закономерностей репрезентации фразеологизмов в корпусах текстов. Проиллюстрируем хронологическую информацию, представленную в НКРЯ, на примере фразеологизма *без задних ног* ‘сильно усталый’ [4, с. 48]; (разг. шутл.) ‘крепко, беспробудно (спать)’ [3, с. 437].

Данный фразеологизм представлен в Основном корпусе НКРЯ ~ 100 раз. Первое вхождение в тексты корпуса датируется 1859 годом – в повести Ф. М. Достоевского «Село Степанчиково и его обитатели»: *Я бы вас и сегодня с собою пригласил, да вот как-то весь упал, раскис, совсем без задних ног сделался*. Полная хронология вхождений ФЕ в тексты НКРЯ: 1859–2015 г.г. Отметим, что в текстах XIX в. (наиболее выразительно) и первой половины XX в. фразеологизм *без задних ног* регулярно используется для характеристики разных состояний человека: опьянения, болезни, сна, усталости: – *Ты, видно, вчера пьян был... без задних ног, раккалия!*.. [Д. Н. Мамин-Сибиряк. Приваловские миллионы (1883)]; *Чахла-чахла, извелась вся, заболела и теперь без задних ног* [А. П. Чехов. В ссылке (1892)]; – *Я думал, они встретят меня, а они спят без задних ног* [С. Т. Семенов. Односельцы (1917)]; *Ты у рабочего спроси, как это его бодрит, когда он в выходной день наработается дотемна, домой возвращается без задних ног, а утром ему к станку становиться...* [В. Ф. Панова. Кружилиха. Роман (1947)]. В текстах второй половины XX в. и XXI в. данный фразеологизм употребляется в основном для характеристики сна (другие состояния характеризуются менее регулярно): *Осмотрела наши палатки, мимоходом выведала, как зовут меня и мою жену, затем сообщила, что под утро она дико замерзла и что Вадька и Барс все еще спят без задних ног* [В. Белов. Чок-получок (1980)]; – *Я, понимаешь, открываю шаланду, а она лежит себе на пакете и дрыхнет без задних ног!* [В. Кунин. Кыся (1998–2000)].

График распределения данного фразеологизма по годам демонстрирует самую высокую относительную частоту его употребления в 1958 г. (ipm 1.02412). Это два вхождения – (1) в роман Ф. Абрамова «Братья и сестры» и (2) в детскую повесть А. Мусатова «Клава Назарова»: (1) *Лежит на сеннике без задних ног... Весь дом винищем провонял*; (2) – *Клава небось дрыхнет без задних ног, – заметил Колька*.

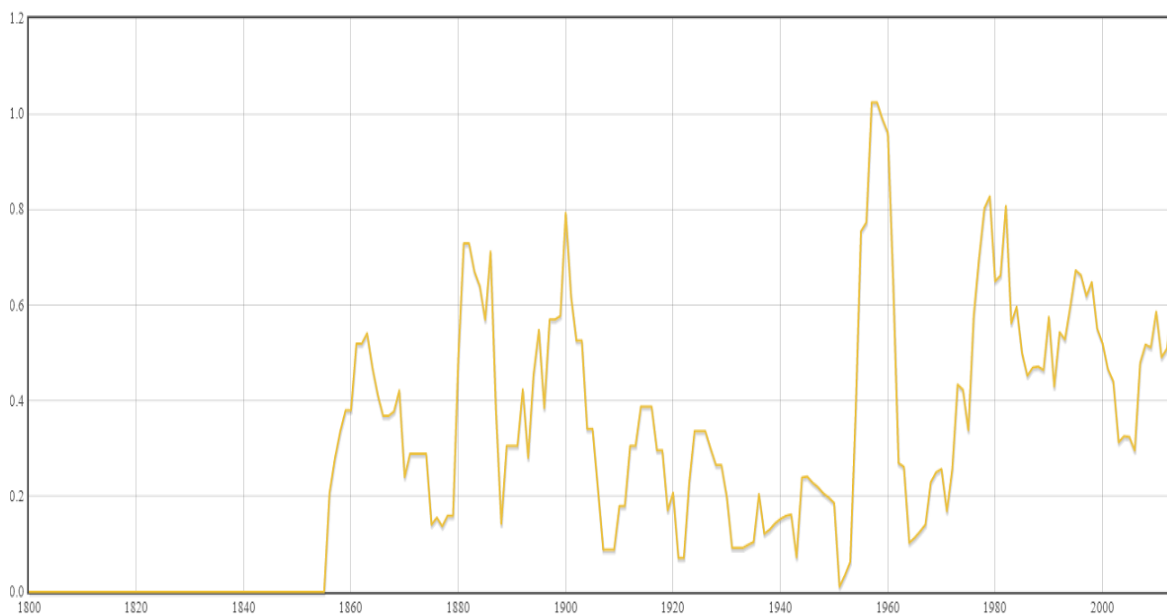


Рисунок 1

Самая высокая абсолютная частота употреблений фразеологизма *без задних ног* в текстах НКРЯ наблюдается в 1999 г. – 4 вхождения в произведения разных авторов.

Отметим, что хронологическая характеристика фразеологизмов, бытующих в корпусах текстов, представляет интерес и сама по себе, и как предпосылка для исследования особенностей репрезентации фразем в данных электронных ресурсах.

Таким образом, обращение к корпусам текстов открывает для фразеологических исследований целый ряд возможностей. Работа с корпусом позволяет определить частотность исследуемых фразеологических единиц; уточнить стандартную с точки зрения современного узуса начальную форму фразеологизма; определить набор наиболее значимых вариантов каждого фразеологизма; выявить многозначность фразеологических единиц и уточнить их конкретные значения; определить типичные для каждого фразеологизма трансформации; охарактеризовать типичное для исследуемых фразеологизмов окружение и виды контекстов, в которых их использование воспринимается как наиболее естественное [см. 5]. Более того, перечисленными возможностями не исчерпывается специфика применения корпусных методов во фразеологических исследованиях. При этом не вызывает сомнений необходимость привлечения количественных данных в ходе решения на базе корпуса текстов любых фразеологических исследовательских задач.

Список использованных источников

1. Захаров, В. П. Лингвистика больших корпусов / В. П. Захаров // Компьютерная лингвистика и вычислительные онтологии : сборник научных статей. Труды XVIII объединенной конференции «Интернет и современное общество» (IMS-2015), Санкт-Петербург, 23–25 июня 2015 г. – СПб. : Университет ИТМО, 2015. – С. 82–93.
2. Национальный корпус русского языка [Электронный ресурс]. – Режим доступа : <http://www.ruscorpora.ru/> – Дата доступа : 03.09.2018.
3. Мокиенко, В. М. Большой словарь русских поговорок / В. М. Мокиенко, Т. Г. Никитина; под общ ред. В. М. Мокиенко. – М. : ЗАО «ОЛМА Медиа Групп», 2007. – 784 с.
4. Михельсон, М. И. Русская мысль и речь. Свое и чужое. Опыт русской фразеологии : сборник образных слов и иносказаний. В 2-х т. / М. И. Михельсон. – Т. 1. – М. : Терра, 1994. – 792 с.
5. Добровольский, Д. О. Корпусы текстов и двуязычная фразеография / Д. О. Добровольский // Вестник Новосибирского гос. пед. университета . – 5 (27). – Новосибирск, 2015. – С. 23– 37.

УДК 811.161.3'35

ВЫТОКІ І ЎЖЫВАННЕ ФРАЗЕАЛАГІЗМА БОГ ЖЫВЕ (паводле матэрыялаў друку і інтэрнэта)

М. А. Даніловіч

Разглядаецца этымалогія фразеалагізма “Бог жыве”, функцыянаванне яго ў беларускай дыялектнай і літаратурнай мове. Аналізуецца нацыянальна-культурная канатацыя створанага Уладзімірам Караткевічам на базе гэтага фразеалагізма крылатага выразу “На Беларусі Бог жыве”. Па матэрыялах друку і інтэрнэта прасочваюцца найбольш тыповыя сферы яго выкарыстання.

Ключавыя словы: *фразеалагізм, крылаты выраз, этымалогія, дыялектная мова, літаратурная мова, пазамоўная сфера.*

The article deals with the etymology of the phraseological unit “Бог жыве” (“God lives”) and its functioning in dialect and standard varieties of the Belarusian language. The ethnic cultural connotation of the popular expression “На Беларусі Бог жыве” (“God lives in Belarus”) created by