

ПРОЕКТ ЧАСТОТНОГО СЛОВАРЯ РУССКИХ ИДИОМ

Баранов А. Н. (baranov_anatoly@hotmail.com),
Вознесенская М. М. (voznesh-masha@yandex.com),
Добровольский Д. О. (dm-dbrv@yandex.ru),
Киселева К. Л. (xenkis@mail.ru),
Козеренко А. Д. (akozerenko@gmail.com)

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

В русской лексикографии до сих пор не существует частотного словаря идиом. Составление такого словаря осложняется тем, что идиомы встречаются в текстах относительно редко. Предлагается проект создания такого словаря на основе словника «Словаря тезауруса современной русской идиоматики», включающего около 8000 идиом, и четырех корпусов текстов. Для получения репрезентативного корпуса текстов необходимо учесть жанровое своеобразие некоторых идиом и индивидуальные особенности языка писателя. При составлении словника предусмотрено разрешение омонимии. Словарь будет содержать информацию об относительной частотности идиом в корпусе в целом, в отдельных корпусах и в подмножествах корпусов, отражающих разные речевые жанры и типы дискурса, а также о частотности и фразеологической активности отдельных компонентов, входящих в состав идиом.

Ключевые слова: частотный словарь, идиомы, лексика, фразеология

TOWARDS A FREQUENCY DICTIONARY OF RUSSIAN IDIOMS

Baranov A. N. (baranov_anatoly@hotmail.com),
Dobrovolskii D. O. (dm-dbrv@yandex.ru),
Kiseleva K. L. (xenkis@mail.ru),
Kozerenko A. D. (akozerenko@gmail.com),
Voznesenskaia M. M. (voznesh-masha@yandex.com)

Vinogradov Russian Language Institute of the Russian Academy
of Sciences, Moscow, Russian Federation

Up to now we have not seen any frequency dictionary of Russian idioms. The reasons why this task seems so challenging are multiple. First of all, idioms appear in texts relatively rare if compared to “normal” words; next, their frequency depends heavily on the type of discourse as well as on the speaker’s (or writer’s) linguistic preferences. Our task is to compile a frequency dictionary of Russian idioms based on the list of idioms developed by our team for the Thesaurus of Modern Russian Idioms (2007). This list amounts to 8000 items and is being constantly revised and updated. We also have 4 text corpora elaborated especially for the needs of idiom analysis. They embrace the spheres of modern Russian relevant for idiomatic expressions: prose, drama, journalism, and detective fiction. We are going to take into account specifics of individual style of different authors: only those who use idioms regularly will be part of the text sample. The main type of information that the dictionary will provide is the frequency of each idiom in the corpus, taken as a whole, as well as in various genres. We will also make a list of word forms that are part of the idioms and provide statistical information on the productivity of various semantic classes and grammatical forms in Russian phraseology. This parameter, quite novel for this field, was labeled *phraseological activity*. The results of this project can link together phraseology and other aspects of language system and facilitate contrastive idiom analysis.

Key words: frequency dictionary, idioms, phraseology, corpora

Настоящая работа посвящена обсуждению проекта частотного словаря русских идиом. Поскольку словарей такого типа в лексикографии не существует, подобный проект требует серьезной теоретической и практической подготовки. Однако самые общие принципы создания фразеологического частотного словаря, вероятно, могут быть заимствованы из известных частотных словарей обычной лексики. Рассмотрим существующую практику составления частотных словарей лексики русского языка.

1. Частотные словари русской лексики: принципы составления

Среди существующих частотных словарей современного русского языка следует упомянуть в первую очередь «Частотный словарь русского языка» (1977) под редакцией Л. Н. Засориной, «Статистический словарь языка русской газеты (1990-е годы)», «Частотный словарь современного русского языка (на материалах Национального корпуса русского языка)» (2009), а также «Частотный словарь современного русского языка» под ред. Л. Лённгрена (1993). Первым современным частотным словарем русского языка следует, по-видимому, считать Частотный словарь (1963) Э. Штейнфельдт, основанный на корпусе советской литературы для школьников объемом 1 млн. словоупотреблений и преследовавший в первую очередь учебные цели. Наиболее известен «Частотный словарь русского языка» (1977) под редакцией Л. Н. Засориной, который до недавнего времени был основным русским частотным словарем. Этот словарь, так же как и «Частотный словарь современного русского языка» (1993) под ред. Л. Лённгрена, опирается на относительно небольшой по современным меркам корпус, насчитывающий около 1 млн. словоупотреблений, в котором представлены тексты с начала XX века и до 1960-х годов. Словарь охватывает четыре речевых жанра: художественная литература, разговорная речь, научные и публицистические тексты, а также деловая речь; за единицу словника словаря принята лексема. Поскольку основная часть публицистических текстов и в значительной мере литературных произведений, вошедших в состав корпуса, были идеологически маркированы, словарь в целом преимущественно отражает официальный русский язык советского периода.

Специализированные статистические словари представляют отдельные жанры: язык художественной литературы (Josselson 1953), науки (Степанова 1970), публицистики (Шайкевич и др. 2008). Так, статистический словарь языка русской газеты (2008) исследует только публицистику 1990-х годов и опирается на несколько корпусов газетных текстов общей численностью более 50 млн. словоупотреблений. Отметим, что в этот словарь в качестве вокабул включены не только отдельные слова, но и фразеологические словосочетания — в основном коллокации, а также прагматемы и некоторые идиомы (ср. включение в словарь таких единиц как *круглый стол*, *по крайней мере*, *Филькина грамота*), однако подобных вокабул всего несколько десятков.

Существуют также словари, содержащие статистическую информацию по языку одного писателя — например, Грибоедова (Поляков 1999), Достоевского (Шайкевич и др. 2003), Цветаевой (Белякова и др. 1996). Статистический словарь языка Достоевского (2003) построен на корпусе текстов в 2.9 млн. словоупотреблений, разбитом на подкорпусы трех основных жанров — художественная литература, критика и публицистика, письма. Наиболее частотные лексемы также распределены по периодам творчества автора. Статистическая информация в словаре представлена в виде множества таблиц: таблица частоты употребления лексем по трем жанрам, обратный частотный словарь, распределение некоторых частых лексем художественной литературы по периодам творчества, таблица частоты некоторых бинарных сочетаний, таблицы текстуальных связей некоторых слов, таблицы лексических маркеров персонажей, таблица грамматических признаков и др. В результате объем полученных статистических данных оказался настолько велик, что авторы приняли решение выпустить словарь в гибридном виде: однотомная книга, показывающая лишь часть таблиц, и сопровождающий ее компакт-диск, содержащий полную информацию.

Последний по времени появления и самый полный на сегодняшний день Частотный словарь (2009) опирается на коллекцию текстов Национального корпуса русского языка общим объемом 92 млн. словоупотреблений, представляющую русский язык последних шестидесяти лет. В корпусе собраны тексты различных функциональных стилей: художественная литература, публицистика, учебно-научная, официально-деловая, производственно-техническая и церковно-богословская литература, электронная коммуникация, реклама, бытовая литература, а также устная непубличная речь. Структура словарной статьи словаря предусматривает разные виды статистической информации, в том числе общую частоту употребления слова (количество употреблений на миллион слов корпуса) и ранг (номер слова в общем частотном списке всех слов словника), причем систему рангов имеют также словоформы, числа, буквы и знаки препинания.

2. Частотные словари фразеологии: *terra incognita*

В современной лексикографической практике отсутствуют частотные словари фразеологии. Это объясняется в первую очередь тем, что значительная часть идиом относительно низкочастотна. Кроме того, многие фразеологизмы характерны лишь для некоторых типов дискурса и речевых жанров. Например, идиома *город на Неве* является типичным журнализмом: в корпусе публицистики¹ она встречается 167 раз, а в корпусе русской прозы — всего 6. Наконец,

¹ Здесь и далее речь идет о корпусах, созданных в отделе экспериментальной лексикографии ИРЯ РАН в середине 1990-х гг. Корпус публицистики 1990-х годов включает более 24 млн. словоупотреблений. Корпус русской прозы охватывает художественную литературу от 1960-х до 2000-х годов и включает более 33 млн. словоупотреблений.

употребление многих фразеологизмов определяется предпочтениями автора. Так, характерной особенностью идиостиля Достоевского является склонность к употреблению грамматического фразеологизма *по крайней мере* (685 вхождений по 35 художественным произведениям), при том что *по меньшей мере* у него практически не употребляется (три примера по 35 художественным произведениям).

Из сказанного следует, что для репрезентативности исходного корпуса требуется как достаточное количество текстов, так и их тематическое и жанровое разнообразие. Однако недостаточно механически увеличить объем корпуса и обеспечить большее тематическое разнообразие составляющих его текстов, что обычно рекомендуется для достижения репрезентативности в корпусных исследованиях. Существенно также учесть тот вклад, который вносит каждый конкретный тип дискурса в общий объем речевых произведений языка. Эта проблема пока не решена.

Наконец, имеются и объективные трудности формирования достаточно полного словника идиом, отражающего современный узус. Действительно, если представительный словник отдельных лексем можно получить, используя имеющиеся словари и относительно небольшие по объему выборки текстов, то идиомы употребляются довольно редко. Имеющиеся фразеологические словари, во-первых, неполны и, во-вторых, существенно отстают от современного узуса. Извлечение идиом из текстов теоретически возможно, но практически неосуществимо — имеющиеся программы автоматической обработки выявляют устойчивые сочетания слов, однако результирующий список из-за большого объема и значительного информационного шума требует существенной редактуры и постоянного общения к контексту для установления того, использовано ли словосочетание идиоматично или нет.

3. Предпосылки создания частотного словаря идиом

В настоящее время в связи с развитием методов корпусной лингвистики, а также накоплением текстового материала появилась возможность решить многие из этих проблем. Во-первых, в результате многолетних исследований сформирован словник современных идиом, который может претендовать на полноту. Имеется в виду словник Словаря-тезауруса современной русской идиоматики (2007), который постоянно уточняется и дополняется. На данном этапе он насчитывает около 8000 русских идиом.

Во-вторых, в ходе работы над этим проектом были созданы три корпуса текстов, отражающие специфику различных сфер использования русского языка. В корпусе «Русская проза» представлены разнообразные речевые жанры — нарратив, бытовые диалоги, внутренняя речь (вплоть до потока сознания) и др. В корпусе публицистики собраны газетные и журнальные тексты второй половины 1990-х годов. Устная речь представлена опосредованно, в том виде, в котором она присутствует в корпусе «Русский детектив». Для более полного отражения этого типа дискурса в настоящее время создается корпус

«Русская драматургия». Другие корпуса будут расширены за счет включения в них материала 2000-х годов.

4. Проблемы определения частотности идиом

Известно, что корпуса достаточно правдоподобно отражают частоту употребительных единиц языка, в то время как реальную частоту низкочастотных единиц отразить не удастся. Поскольку идиоматика в целом относится скорее к низкочастотной части лексикона, возникает серьезная проблема. Она усугубляется тем, что разные говорящие в разной степени склонны к идиоматизации речи. Если один говорящий скажет *он пытался меня обмануть* или *он говорил мне неправду*, другой предпочтет иную форму: *он вешал мне лапшу на уши* или *он гнал пургу/лепил горбатого* и т.п. Идиоматичность (и соответственно, фразеологичность) речи — фактор индивидуальный, что хорошо видно по имеющимся источникам. Так, в Базе данных по современной идиоматике, включающей более 50 000 контекстов употребления идиом, произведения некоторых писателей представлены большим количеством контекстов, а других — весьма незначительным. Например, из произведений В. Аксенова в базе имеются 603 контекста употребления идиом, Л. Петрушевской — тоже 603, Э. Тополя — 657, братьев Стругацких — 576, Саши Соколова — 542, Е. Гинзбург — 704, А. Солженицына — 652, В. Сорокина — 477, В. Войновича — 469, Ю. Трифонова — 471, Абрама Терца — 525, В. Приставкина — 473, В. Дудинцева — 435, С. Каледина — 343, Д. Быкова — 308, Б. Окуджавы — 310. С другой стороны, контекстов использования идиом в произведениях А. Марининой обнаружено 39, Л. Улицкой — 15, В. Шефнера — 5. Понятно, что в определенной степени это можно объяснить объемами представленных в базе произведений. Однако и при учете этого обстоятельства число употреблений идиом у «идиоматичных» и «неидиоматичных» авторов различается в десятки раз. Так, общий объем произведений В. Войновича в базе данных составляет в словоупотреблениях 428 400 токенов, а общий объем произведений А. Марининой, учтенных в базе данных, составляет 1 376 237 токенов. Тем самым, относительная частота употребления идиом в текстах В. Войновича составляет 0,001, а в текстах А. Марининой — 0,000028.

Таким образом, степень идиоматичности речи, проявляющаяся в частоте использования фразеологии, является индивидуальной чертой идиолекта каждого говорящего и пишущего. Данный факт необходимо учесть при создании корпуса для частотного фразеологического словаря. Действительно, если в нем будут широко представлены авторы, практически не употребляющие идиомы, то картина будет искажена, поскольку некорректно оценивать частоту какого-то явления у тех носителей языка, для речи которых это явление нетипично в принципе. Поэтому при подготовке корпусов художественной прозы, драматургии и детективной литературы следует исключить из них тексты тех авторов, которые не склонны к использованию фразеологии (для корпуса публицистики такой отбор по очевидным причинам невозможен). Порог частоты использования идиом для включения того или иного автора в корпус будет определен экспериментально.

5. Формирование словника

Основу словника словаря составят идиомы, зафиксированные в «Словаре-тезаурусе», за исключением идиом с обценными компонентами. Это объясняется тем, что данные о частоте использования таких идиом были бы заведомо искажены: обценная идиоматика является феноменом устной речи, ее отражение в литературных произведениях, даже включающих значительные фрагменты бытовых диалогов, крайне нерегулярно из-за табуированности этой группы слов русского языка.

Входом словарной статьи будет служить идиома в начальной форме. Начальная форма будет зависеть от типа идиомы — она может быть именной, адъективной, адвербиальной, глагольной и т.п. Будет использоваться та форма, которая представлена в «Словаре-тезаурусе...», поскольку для словника этого словаря уже была проделана соответствующая работа. Вопрос о том, будут ли варианты одной идиомы представлены в одной словарной статье или в разных, пока не решен. В том случае, если они будут сведены в одну словарную статью, предстоит решить вопрос о критериях выбора заглавного варианта: это может быть наиболее частотный вариант, морфологически более простой вариант, наиболее полный по компонентному составу вариант и др.

Следует отметить, что, как и при создании частотных словарей лексики, этот этап работы включает разрешение омонимии. Ср. две омонимичные идиомы *на учёте*:

НА УЧЁТЕ I. (кто-л.) Человек, мыслящий, думающий и т.п. ненормально или испытывающий затруднения в мыслительной деятельности и, как следствие, ведущий себя неадекватно из-за психического заболевания, которое документально фиксируется в соответствующем медицинском учреждении или из-за своей очевидности так осмысляется.

Мужчина, явно на учёте, схватив ружье и, по предположениям, коробку дроби, закрылся в одной из комнат квартиры и категорически отказывался выходить. [Корпус Публ.]

НА УЧЁТЕ II. (что-л. у кого-л.) *нейтр.* Чего-л. так мало, что это требует постоянного контроля его наличия — в том числе письменной фиксации или так осмыляется.

Жанна — простая чертежница, у нее каждый рубль на учёте. [Корпус Публ.]

В тексте словаря полное толкование не предусмотрено, однако будет представлен краткий комментарий к значению, который позволит пользователю отождествить соответствующие употребления.

6. Типы информации в частотном словаре идиом

6.1. Частота по корпусам

В настоящее время рано говорить обо всех компонентах словаря и о деталях структуры словарной статьи. Укажем лишь те типы информации, которые заведомо будут отражены в словаре. Прежде всего, это относительная частота употребления идиомы во всех четырех корпусах, вместе взятых. Будет указана также относительная частота употребления в каждом из корпусов: Русская проза; Русская драматургия; Русская публицистика; Русский детектив.

Поскольку корпус «Русский детектив» и корпус «Русская драматургия» в значительной степени отражают особенности функционирования разговорной речи, то будет отдельно фиксироваться и относительная частота употреблений в этих двух корпусах. Кроме того, целесообразно дать читателю относительную частоту употребления во всех корпусах, подвергшихся нормированию по степени «фразеологичности» (Русская проза, Русская драматургия и Русский детектив).

Как было уже отмечено выше, многие идиомы привязаны к конкретному речевому жанру. В то же время есть и идиомы, которые не ограничены жанровыми характеристиками. Так, идиомы *рано или поздно*, *ровным счетом*, *все до одного*, *без устали*, *между делом* нейтральны, то есть велика вероятность того, что они в равной мере употребительны в разных типах дискурса. Тем самым, встает задача упорядочить идиомы не только по их относительной частоте, но и по их распространенности в разных корпусах. Для этого в существующих частотных словарях используются разные статистические инструменты: в Чешском национальном корпусе это показатель Average Reduced Frequency (ARF) [Čermak, Křen 2005], в других частотных словарях — коэффициент А. Жуйана [Gries 2008]. Один из этих инструментов будет использован и в планируемом словаре.

6.2. Частотность компонентов идиом

Предполагается включить в словарь информацию о количественных характеристиках компонентов идиом, т. е. о том, сколько раз то или иное слово употребляется в составе разных идиом русского языка. Подобные данные имеют ценность как для исследования фразеологии, давая объективную информацию о лексическом составе идиом, так и для установления связей с лексикой, а также с морфологией и словообразованием.

Технология создания частотного словника компонентов идиом предусматривает следующие этапы. Все идиомы разбиваются на составляющие их компоненты, полученные словоформы сводятся к исходным формам лексем (леммам), которые снабжаются числовым индексом, показывающим частоту

их употребления. В большинстве случаев правила сведения словоформ к исходной форме те же, что приняты в современных словарях (см., например, Частотный словарь русского языка (1977): 11–19). Так, для существительного это им. падеж ед. ч., для прилагательного — им. падеж ед. ч. мужского рода, для глагола — форма инфинитива. Но есть и некоторые отличия, связанные со спецификой фразеологического материала. Степени сравнения прилагательных и наречий, видовые пары глаголов, причастия и деепричастия, супплетивные формы (*человек/люди, ребенок/дети*) считаются разными леммами. Уникальные компоненты представлены в словнике в том виде, в каком они употреблены в составе идиом: *зги (не видно ни зги), скажу (на всем скажу)*. Предусматривается разделение омонимов, например, *да* как союз и как частица будут иметь разные входы и, соответственно, разную частотность. Полученный таким образом словник компонентов будет включать все слова, входящие в состав идиом — знаменательные слова (в том числе имена собственные), уникальные компоненты и заимствованные слова, а также служебные слова — предлоги, союзы, частицы и др.

6.3. Фразеологическая активность

Словник компонентов идиом будет представлен в двух вариантах — алфавитно-частотном и собственно частотном (по убыванию частотности). С помощью Указателя, содержащего отсылки от любого компонента к идиоме, пользователь сможет самостоятельно извлечь информацию об идиомах, в состав которых входит тот или иной компонент, а также о том, в каких словоформах он реализуется и насколько часто встречаются эти словоформы. Само присутствие в словнике того или иного компонента является признаком его «фразеологической активности», а частотный показатель позволяет судить о степени этой активности. Понятие «фразеологической активности» может рассматриваться как новый параметр исследования фразеологии.

Тем самым становится возможным выявление центра и периферии рассматриваемого множества единиц. Классификация слов, входящий в состав идиом, по различным основаниям может дать интересные результаты, характеризующие фразеологию русского языка как систему. Словник компонентов и их количественная характеристика могут служить, например, основой выделения различных семантических классов слов, входящих в состав идиом. Можно будет также исследовать «фразеологическую активность» отдельных грамматических и морфологических категорий (например, суффиксов и префиксов). Таким образом, частотный словник компонентов идиом может служить «мостиком», соединяющим фразеологию с другими аспектами языковой системы.

Полученные данные окажутся чрезвычайно полезными и при сопоставительном изучении разных языков. Так, сравнение словников и частотных

характеристик может выявить как черты, сближающие фразеологические системы двух языков, так и важные содержательные различия².

7. Заключение

Создание частотного словаря идиом — новая теоретическая и прикладная задача. Ее решение позволит:

- определить относительную частоту употребления подавляющего большинства русских идиом в целом по корпусу и в разных типах текстов;
- выявить группы общеупотребительных, редких, стилистически связанных фразеологизмов, а также «авторских фреквенталий»;
- оценить распространенность в идиоматике омонимии;
- получить количественное представление о грамматических типах идиом;
- наглядно представить лексический компонентный состав русских идиом, а также степень фразеологической активности отдельных компонентов идиом и типов компонентов.

Литература

1. Баранов А. Н., Вознесенская М. М., Добровольский Д. О., Киселева К. Л., Козеренко А. Д. Фразеологический объяснительный словарь русского языка. — Москва, Эксмо, 2009.
2. Баранов А. Н., Добровольский Д. О. Аспекты теории фразеологии. — Москва, Знак, 2008.
3. Баранов А. Н., Добровольский Д. О., Киселева К. Л., Козеренко А. Д. при участии М. М. Вознесенской и М. М. Коробовой, под редакцией А. Н. Баранова и Д. О. Добровольского. Словарь-тезаурус современной русской идиоматики. — Москва, Мир энциклопедий Аванта+, 2007.
4. Белякова И. Ю., Оловянникова И. П., Ревзина О. Г. (сост.). Словарь поэтического языка Марины Цветаевой. — Москва, Дом-музей Марины Цветаевой, 1996.
5. Засорина Л. Н. (ред.). Частотный словарь русского языка. — Москва, Русский язык, 1977.
6. Лённгрен Л. (ред.). Частотный словарь современного русского языка. — Uppsala : Acta Univ. Ups., Studia Slavica Upsaliensia 32, 1993.
7. Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка) — Москва, Азбуковник, 2009.
8. Поляков А. Е. Электронный словарь языка писателя (на примере языка А. С. Грибоедова) // Труды Международного семинара

² См., например, реализацию подобного подхода при сравнительно-сопоставительном изучении фразеологии русского и немецкого языков в [Райхштейн 1980: 87–99].

- Диалог-99 по компьютерной лингвистике и ее приложениям. — Таруса, 1999.
9. Райхштейн А. Д. Сопоставительный анализ немецкой и русской фразеологии. — Москва, Высшая школа, 1980.
 10. Степанова Е. М. Частотный словарь общенаучной лексики. — Москва, Изд-во МГУ, 1970.
 11. Шайкевич А. Я., Андриющенко В. М., Ребецкая Н. А. Статистический словарь языка Достоевского. — Москва, Языки славянских культур, 2003.
 12. Шайкевич А. Я., Андриющенко В. М., Ребецкая Н. А. Статистический словарь языка русской газеты (1990-е годы)/ Рос. акад. наук. Ин-т русского языка им. В. В. Виноградова. Т. 1. — Москва, Языки славянских культур, 2008.
 13. Штейнфельдт Э. Частотный словарь современного русского литературного языка. — Таллин, 1963.
 14. Čermak F., Křen M. New generation corpus-based frequency dictionaries: the case of Czech // International journal of corpus linguistics, 10, 2005.
 15. Gries S. Th. Dispersions and adjusted frequencies in corpora // International journal of corpus linguistics, 13, 4, 2008.
 16. Josselson H. The Russian word count and frequency analysis of grammatical categories of Standard Literary Russian. — Detroit: Wayne University Press, 1953.

References

1. Baranov A. N., Voznesenskaia M. M., Dobrovol'skii D.O., Kiseleva K. L., Kozerenko A. D. (2009) Frazeologicheskii ob"iasnitel'nyi slovar' russkogo iazyka [Explanative phraseological dictionary of Russian language]. Moscow, Eksmo.
2. Baranov A. N., Dobrovol'skii D. O. (2008) Aspekty teorii frazeologii [Aspects of the Theory of Phraseology]. Moscow, Znak.
3. Baranov A. N., Voznesenskaia M. M., Dobrovol'skii D.O., Kiseleva K. L., Korobova M. M., Kozerenko A. D. (2007), edited by Baranov A. N., Dobrovol'skii D.O. Slovar'-tezaurus sovremennoi russkoi idiomatiki [Thesaurus of modern Russian idioms]. Moscow, Mir entsiklopedii Avanta+.
4. Beliakova I. Iu., Oloviannikova I. P., Revzina O. G. (eds.) (1996) Slovar' poeticheskogo iazyka Mariny Tsvetaevoi. [The dictionary of Marina Tsvetaeva's poetic language]. Moscow, Dom-musei Mariny Tsvetaevoi.
5. Čermak F., Křen M. (2005) New generation corpus-based frequency dictionaries: the case of Czech, International journal of corpus linguistics, 10.
6. Gries S. Th. (2008) Dispersions and adjusted frequencies in corpora, International journal of corpus linguistics, 13, 4.
7. Josselson H. (1953) The Russian word count and frequency analysis of grammatical categories of Standard Literary Russian. — Detroit: Wayne University Press.
8. Liashevskaja O. N., Sharov S. A. (2009) Chastotnyi slovar' sovremennogo russkogo jazyka (na materialax atSIONal'nogo korpusa russkogo iazyka) [A Frequency

- Dictionary of Modern Russian (based on the data of The National Russian Corpus]. Moscow, Azbukovnik.
9. *Lönngren L.* (ed.) (1993) *Chastotnyi slovar' sovremennogo russkogo iazyka* [A Frequency Dictionary of Modern Russian]. Uppsala, Acta Univ. Ups., Studia Slavica Upsaliensia 32.
 10. *Poliakov A. E.* (1999) *Elektronnyj slovar' jazyka pisatelja (na primere iazyka A. S. Griboedova)* [The electronic dictionary of the writer's language (the case of A. S. Griboedov)]. *Trudy Mejdunarodnogo seminara Dialog-99 po computer-noi lingvistike i ee prilozheniiam*. Tarusa.
 11. *Raikhshtein A. D.* (1980) *Sopostavitel'nyj analiz nemetskoj i russkoj frazeologii* [Comparative analysis of the German and Russian phraseology]. Moscow, Vysshaia shkola.
 12. *Shaikevich A. Ia., Andriushchenko V. M., Rebetskaia N. A.* (2003) *Statisticheskii slovar' iazyka Dostoevskogo* [The statistical analysis of the Dostoevsky's language]. Moscow, *Iazyki slavianskikh kultur*.
 13. *Shaikevich A. Ia., Andriushchenko V. M., Rebetskaia N. A.* (2008) *Statisticheskii slovar' iazyka russkoj gazety (1990-e gody)* [The statistical analysis of the language of Russian newspapers in the 1990s] V. 1. Moscow, *Iazyki slavianskikh kultur*.
 14. *Shteinfel'dt E.* (1963) *Chastotnyi slovar' sovremennogo russkogo iazyka* [A Frequency Dictionary of Modern Russian]. Tallinn.
 15. *Stepanova E. M.* (1970) *Chastotnyi slovar' obshchenauchnoi leksiki* [A Frequency Dictionary of General Scientific Notions]. Moscow, MGU.
 16. *Zasorina L. N.* (ed.) (1977) *Chastotnyi slovar' russkogo iazyka* [Frequency Dictionary of Russian Language]. Moscow, *Russkii Iazyk*.