

Научная статья

УДК 81'33

DOI 10.25205/1818-7935-2023-21-4-18-36

Стилистические пометы «книжное» и «разговорное» во фразеологических словарях

Екатерина Михайловна Арбузова¹, Олег Владимирович Дурандин²,
Юлия Алексеевна Карабасова³, Маргарита Андреевна Климова⁴

¹ПАО «Сбербанк»,
Нижний Новгород, Россия

²Независимый исследователь,
Нижний Новгород, Россия

³ООО «ОК “Вета”»,
Нижний Новгород, Россия

⁴Национальный исследовательский университет «Высшая школа экономики»,
Нижний Новгород, Россия

¹emarbuzova@sberbank.ru, <https://orcid.org/0009-0008-7147-1893>

²oleg.durandin@gmail.com

³karabasova@veta.expert, <https://orcid.org/0009-0007-8307-9293>

⁴mfokina@hse.ru, <https://orcid.org/0000-0003-2249-6396>

Аннотация

Статья посвящена анализу распределения стилистических помет *книжное* и *разговорное* во фразеологических словарях, их взаимосвязи с другими пометами и верификации соответствующих помет русскоязычных фразеологизмов с помощью корпусно-ориентированного подхода. Анализ проведен на материале электронного датасета, включающего четыре фразеологических словаря. В результате исследования выявлены значительные различия между словарями в наборе помет, в наполненности стилистической информацией, в востребованности конкретных помет, в стилистической характеристике отдельных фразеологизмов. Для определения статистически значимой взаимосвязи помет *книжное* и *разговорное* с другими пометами был использован метод индуктивной статистики «хи-квадрат». Значимой признана взаимосвязь помет *книжное* и *высокое, устаревшее* и *книжное*, а также взаимосвязь пометы *разговорное* и эмоционально-экспрессивных помет. Основные случаи несовпадения, затрагивающие пометы *книжное* и *разговорное*, включают квалификацию фразеологизма как разговорного в одном словаре и как просторечного – в другом, а также несовпадение пометы *книжное*, с одной стороны, и *просторечное* или *разговорное*, с другой стороны. Для выявления фразеологизмов, стилистические свойства которых, а именно принадлежность к книжным или разговорным, обоснованы количественными характеристиками в соответствующих дискурсах, был применен корпусно-ориентированный подход. В результате анализа встречаемости фразеологизмов в корпусах художественных текстов и разговорной речи было обнаружено, что менее половины исследованных единиц тяготеет к преобладающему употреблению в корпусе, соответствующему присвоенной помете. Отдельно рассмотрены спорные случаи, представляющие собой фразеологизмы, которые в части словарей охарактеризованы как книжные, в то время как в других словарях имеют помету *разговорное*. Подобные фразеологизмы продемонстрировали преимущественное тяготение к употреблению в книжной речи. Опыт применения корпусно-ориентированного подхода к книжным и разговорным фразеологизмам позволил сделать вывод о том, что данный подход может быть использован для присвоения помет и их верификации при условии учета технических ограничений и гетерогенности словников фразеологических словарей.

© Арбузова Е. М., Дурандин О. В., Карабасова Ю. А., Климова М. А., 2023

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2023. Т. 21, № 4
Vestnik NSU. Series: Linguistics and Intercultural Communication, 2023, vol. 21, no. 4

Ключевые слова

фразеология, фразеологический словарь, русский язык, стилистические пометы, корпусно-ориентированный подход

Для цитирования

Арбузова Е. М., Дурандин О. В., Карabasова Ю. А., Климова М. А. Стилистические пометы «книжное» и «разговорное» во фразеологических словарях // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2023. Т. 21, № 4. С. 18–36. DOI 10.25205/1818-7935-2023-21-4-18-36

Stylistic Labels *Bookish* and *Colloquial* in Phraseological Dictionaries

Ekaterina M. Arbuzova¹, Oleg V. Durandin²,
Yulia A. Karabasova³, Margarita A. Klimova⁴

¹PJSC “Sberbank”, Nizhniy Novgorod, Russia

²independent scholar, Nizhniy Novgorod, Russia

³Veta Expert Group, Nizhniy Novgorod, Russia

⁴HSE University, Nizhniy Novgorod, Russia

¹emarbuzova@sberbank.ru, <https://orcid.org/0009-0008-7147-1893>

²oleg.durandin@gmail.com

³karabasova@veta.expert, <https://orcid.org/0009-0007-8307-9293>

⁴mfokina@hse.ru, <https://orcid.org/0000-0003-2249-6396>

Abstract

The article deals with discrepancies between stylistic labels of phraseological units in various dictionaries, which is considered to be one of the deep-rooted issues in lexicography. The authors compiled an electronic dataset of four phraseological dictionaries of the Russian language and then analysed the use of the stylistic labels *knizhnoje* (*bookish*) and *razgovornoje* (*colloquial*) in them. The results demonstrate considerable discrepancies between the dictionaries both in the set of stylistic labels and in the density of stylistic information. In addition, the relevance of stylistic labels, in particular that of *razgovornoje* (*colloquial*), varies considerably. The authors then statistically explored the connection between the stylistic labels *knizhnoje* (*bookish*) or *razgovornoje* (*colloquial*) and other stylistic characteristics using the chi-square test. Only a few pairs of stylistic labels have been found to possess a statistically significant connection, among them *knizhnoje* (*bookish*) and *vysokoje* (*elevated*), *ustarevshee* (*archaic*) and *knizhnoje* (*bookish*) as well as the combination of *razgovornoje* (*colloquial*) and emotive and expressive labels. The main differences in the stylistic labelling of one phraseological unit in different dictionaries have been found to include the discrepancy between the labels *razgovornoje* (*colloquial*) and *prostorechnoje* (*vernacular*) as well as assigning the label *knizhnoje* (*bookish*) in one dictionary and the label *razgovornoje* (*colloquial*) or *prostorechnoje* (*vernacular*) in the other. In the final part of the article, the authors apply a corpus-oriented approach to verify the stylistic labels *knizhnoje* (*bookish*) or *razgovornoje* (*colloquial*) by comparing the frequency of the corresponding phraseological units in corpora of fiction and colloquial speech respectively. According to the results, less than a half of the analysed phraseological units correspond to the stylistic label assigned to them in the dictionary. The debatable cases in terms of being *knizhnoje* (*bookish*) or *razgovornoje* (*colloquial*) were then analysed and found to predominantly occur in the corpus of fiction. On the basis of the findings it is concluded that the corpus-oriented approach can be used for assigning and verifying stylistic labels as long as the technical limitations and the heterogeneity of the units in phraseological dictionaries are taken into account.

Keywords

phraseology, phraseological dictionary, the Russian language, stylistic labeling, corpus-oriented approach

For citation

Arbuzova E. M., Durandin O. V., Karabasova Yu. A., Klimova M. A. Stylistic Labels *Bookish* and *Colloquial* in Phraseological Dictionaries. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2023, vol. 21, no. 4, pp. 18–36. DOI 10.25205/1818-7935-2023-21-4-18-36

Введение

В связи с активным взаимодействием лексикографии и компьютерной лингвистики меняется роль лексикографа: по словам Д. В. Андриановой, «он должен выступить в роли аналитика базы данных по отношению к уже существующему текстовому словарю, чтобы адаптировать этот материал под новый формат – формат электронной базы данных» [Андрианова, 2019. С. 9]. Данное исследование посвящено использованию подобной электронной базы данных, объединяющей несколько словарей фразеологизмов, для анализа стилистических помет.

Задача стилистики в области фразеологии – это описание стилистических характеристик фразеологизмов, их классификация, а также выявление правил, по которым стилистические пометы могут приписываться фразеологизмам в словаре [Баранов, Добровольский, 2014. С. 162]. Отсутствие единообразия классификаций стилистических помет и правил их присвоения отмечалось многими учеными [Ниннин, 2012; Есетова, 2017; Скляревская, 2016; Аюпова, 2008]. Признавая объективность причин непоследовательных и противоречивых словарных стилистических оценок и невозможность их полной унификации, Г. Н. Скляревская тем не менее отмечает необходимость стандартизации стилистических квалификаций и преодоления интуитивного принципа присвоения помет [Скляревская, 2016. С. 611–612]. В качестве объективного инструмента обоснования стилистических помет успешно опробован корпусно-ориентированный подход, основанный на частотности фразеологизмов в жанрово-специфических корпусах [Козеренко и др., 2013]; подобный подход применяется и в настоящем исследовании.

Данная работа посвящена сопоставлению четырех фразеологических словарей, объединенных в датасет, с целью выявления различий в наполненности пометами и их востребованности. На следующем этапе исследования особое внимание уделяется пометам *книжное* и *разговорное* как наиболее частотным, демонстрирующим несоответствие в разных словарях и в связи с этим требующим объективных критериев присвоения. Выявляется связь данных помет с другими стилистическими характеристиками фразеологизмов, анализируется частотность фразеологизмов с пометами *книжное* и *разговорное* в корпусах художественной литературы и устной речи.

В работе рассматриваются следующие исследовательские вопросы.

1. Как отличается набор помет и степень наполненности словарей стилистической информацией?
2. Какие пометы оказываются наиболее и наименее востребованными?
3. С какими пометами значимо взаимосвязаны пометы *книжное* и *разговорное*?
4. Каковы ключевые случаи несоответствия помет у одного и того же фразеологизма в разных словарях?
5. Для каких из фразеологизмов, снабженных пометами *книжное* и *разговорное*, стилистические свойства обоснованы количественными характеристиками в соответствующих дискурсах?

1. Обзор литературы

В современном мире лексикография вступает в активное взаимодействие с компьютерной лингвистикой в аспекте обработки неоднословных единиц (англ. multi-word expressions, multi-word units). Необходимо оговориться, что данное понятие шире понятия фразеологизма: неоднословные единицы включают также составные наименования, фразовые глаголы, составные предлоги и т. д. [Masini, 2019; Villavicencio, Idiart, 2019]. Несмотря на различия в классификациях разных авторов, в состав неоднословных единиц неизменно включаются идиомы и коллокации.

Лексикография и компьютерная лингвистика успешно дополняют друг друга при решении задач идентификации и извлечения неоднословных единиц [Villavicencio, Idiart, 2019; Dobrovolsk, 2020], однако этим не ограничиваются возможности данного взаимодействия, об-

зор которых представлен в [Gantar et al., 2019]. Применение методов обработки естественного языка может также обогатить лексикографическое описание неоднословных единиц, значительную часть которых составляют фразеологизмы, в вопросах оценки целостности их значения, определения синонимических и других отношений между ними с помощью моделей дистрибуции. Для эффективного решения данных задач необходимы аннотированные корпуса или базы данных (датасеты) неоднословных единиц, активно создаваемые для разных языков [Villavicencio, Idiart, 2019; Fellbaum, 2015].

В качестве подобного ресурса для русского языка в первую очередь выступает База данных по русской идиоматике, разрабатываемая коллективом Отдела экспериментальной лексикографии Института русского языка им. В. В. Виноградова РАН под руководством А. Н. Баранова и Д. О. Добровольского и представляющая собой репрезентативный корпус примеров по русской фразеологии (экспериментальный вариант базы данных размещен на сайте <https://www.slovari.ru/>). Следует упомянуть и о Базе данных коллокаций, созданной в Санкт-Петербургском государственном университете (<https://collocations.spbu.ru/>). Разработчики ресурса подчеркивают, что база не включает идиомы и для ее формирования не использовались фразеологические словари [Хохлова, Мамаев, 2021. С. 372]. Таким образом, для сопоставления информации о фразеологизмах в различных словарях, выявления случаев, требующих разработки объективных критериев присвоения помет, необходим отдельный датасет, который и был разработан в рамках настоящего исследования.

Непоследовательность стилистической характеристики фразеологизмов в словарях, необходимость ее уточнения и связанные с этим трудности отмечались многими исследователями [Ниннин, 2012; Есетова, 2017; Складьевская, 2016; Аюпова, 2008]. По словам А. Н. Баранова и Д. О. Добровольского, «сам “пересчет” категорий теории стилистики в систему словарных стилистических помет оказывается нетривиальной задачей» [Баранов, Добровольский, 2014. С. 163]. В ряде работ рассмотрена проблема разнообразия классификаций и множественности их оснований [Ниннин, 2012; Есетова, 2017; Аюпова, 2008]. Как следствие, основополагающие пометы *разговорное* и *книжное* присутствуют во всех классификациях, но рассматриваются учеными по-разному в связи со спецификой категоризации конкретного исследования. Так, Р. А. Аюпова [Аюпова, 2008] относит пометы *разговорное* и *книжное* к тем, что указывают на функционирование фразеологизма в определенной сфере общения. В классификации Г. О. Ниннина пометы *книжное* и *разговорное* относятся к группе помет функционально-стилистического содержания, регулирующихся фактами социального характера, однако к различным подгруппам – «книжная фразеология» и «разговорная фразеология» соответственно [Ниннин, 2012]. В рамках системы стилистических шкал А. Н. Баранова и Д. О. Добровольского [Баранов, Добровольский, 2014] помета *книжное* относится к дискурсивным пометам, наряду с *журнализм*, *жаргонизм*, *народное* и т. д., в то время как *разговорное* определяет фразеологизм с точки зрения его стилистического регистра, в оппозиции с такими пометами, как *высокое*, *сниженное*, *грубое*, *неприличное*. Более того, по мнению исследователей, будучи самой частотной стилистической характеристикой идиом, разговорное не нуждается в маркировании пометой, т. е. должно помечаться через значимое отсутствие. Данная концепция воплощена в «Фразеологическом объяснительном словаре русского языка» [Баранов и др., 2009. С. 13].

Как можно заметить, расхождения в отнесении той или иной пометы к определенному классу указывают на проблемы в теории стиля, что заставляет различных исследователей создавать свои собственные подходы к описанию стилистических характеристик.

Трудности в стилистической квалификации отдельных фразеологизмов, как представляется, могут быть связаны и со сложными семантическими и прагматическими отношениями, выражающимися во взаимосвязи некоторых помет. В частности, по наблюдениям А. Н. Баранова и Д. О. Добровольского, пометы *книжное* и *высокое* оказываются взаимозависимыми, кроме того, в качестве книжных зачастую воспринимаются устаревшие фразеологизмы. [Баранов, Добровольский, 2014].

Внимание исследователей уделено принципам присвоения помет [Есетова, 2017; Лавриненко, 2017; Козеренко и др., 2013]. Рассматривая факторы, определяющие принадлежность французских фразеологизмов к разговорному стилю, В. А. Лавриненко выделяет коннотативные значения, экспрессивность и образность фразеологической единицы, антропоцентричность метафоры, слова с конкретным значением в качестве основного лексического компонента, наличие разговорной, просторечной и арготической лексики, неясную образность [Лавриненко, 2017]. Тем не менее, как отмечают А. Н. Баранов и Д. О. Добровольский, несмотря на существующие попытки сформулировать формальные критерии, реально в практических работах для присвоения помет фразеологизмам в основном используется интуиция [Баранов, Добровольский, 2008], что обуславливает актуальность разработки и внедрения операциональных критериев отнесения фразеологизма к определенному стилистическому разряду.

Данное исследование подвергает тестированию метод, впервые предложенный в работе А. Д. Козеренко, А. Н. Баранова, М. М. Вознесенской, Д. О. Добровольского, К. Л. Киселевой [Козеренко и др., 2013]. Методика позволяет предположить стилистическую характеристику фразеологизма на основе его количественных свойств, т. е. встречаемости в жанрово-специфичных корпусах текстов. По результатам исследования 100 фразеологизмов из словника Тезауруса [Баранов и др., 2007] авторами установлено, что такой метод может служить инструментом обоснования, проверки и коррекции стилистических помет.

Несмотря на ограничения данного подхода, он, несомненно, является значимым шагом на пути преодоления интуитивного принципа присвоения помет. Наша коллекция фразеологизмов позволяет протестировать этот подход на значительно большем количестве единиц, а также в ситуациях совпадения таких помет, как *книжное* и *разговорное* у одной и той же словарной статьи.

2. Материал и методы

Материалом для данного исследования стала электронная коллекция, источниками для которой послужили четыре русскоязычных фразеологических словаря. В таблице 1 кратко представлена основная информация о выбранных словарях. Все перечисленные словари, кроме двуязычного словаря Д. И. Квеселевича, относятся к толковым фразеологическим словарям. Наиболее объемным и актуальным с точки зрения года выхода является словарь А. И. Федорова, включающий дополненный словник словаря А. И. Молоткова и отмеченный А. Н. Барановым и Д. О. Добровольским в числе ключевых толковых фразеологических словарей русского языка [Баранов, Добровольский, 2014. С. 152].

Сбор данных осуществлялся путем парсинга электронных словарей, представленных в формате DSL (Dictionary Specification Language). Словарные статьи собранной коллекции включают в себя варианты фразеологизма, его значения, стилистические пометы, синтаксическую роль, примеры употребления, а также наименование словаря, из которого были получены данные.

Одна словарная статья может содержать информацию о фразеологизме сразу из нескольких источников. Таким образом, итоговый объем коллекции меньше, чем совокупный объем первоначальных словарей, и составляет 17 847 статей.

Для парсинга и объединения их в электронную коллекцию был разработан скрипт на языке программирования Python с использованием ряда фреймворков.

В ходе анализа стилистических помет использовались статистические методы, а именно методы дескриптивной и индуктивной статистики и методы графического представления результатов, также осуществляемые при помощи компьютерного моделирования.

Метод анализа количественных характеристик фразеологизмов в жанрово-специфичных корпусах предполагает наличие корпусов нескольких жанров согласно исследуемым пометам.

Таблица 1

Краткая справка о словарях, вошедших в датасет

Table 1

Summary of dictionaries included in the dataset

Словарь	Автор	Год	Объем (количество единиц)
Русско-английский фразеологический словарь [17]	Д. И. Квеселевич	1998	7 567
Универсальный фразеологический словарь русского языка [18]	Т. В. Волкова	2000	4 368
Фразеологический словарь русского языка [19]	И. В. Федосов, А. Н. Лапицкий	2003	6 367
Фразеологический словарь русского литературного языка [20]	А. И. Фёдоров	2008	11 310

В данной работе внимание сфокусировано на пометах *разговорное* и *книжное*, поэтому были сформированы соответствующие корпуса: литературный и разговорной речи. Для формирования корпуса литературы были собраны произведения, высоко оцененные редакцией сайта <https://proza.ru/>. В итоговый корпус вошло 846 текстов 191 автора, опубликованные на сайте с января 2021 по декабрь 2022 г. Корпус разговорной речи собран путем извлечения и обработки текстов и диалогов российских телесериалов за период с 2007 по 2022 г. с портала <https://www.opensubtitles.org/ru>. В корпусе оказались тексты 240 сериалов, снятых на русском языке с ориентацией на русскоязычную аудиторию.

Сбор и лемматизация корпусов осуществлялись с помощью языка программирования Python. Объем каждого из них – чуть более миллиона слов. Несмотря на технические ограничения, связанные с возможностями персонального компьютера, и в корпус литературной, и в корпус разговорной речи вошли работы разнообразных тематик и жанров. На наш взгляд, данные факторы, наряду с временным разнообразием сериалов в составе корпуса разговорной речи и количеством различных авторов прозы, позволяют говорить о репрезентативности корпусов, компенсирующей небольшой объем.

3. Результаты и обсуждение

3.1. Набор помет и степень наполненности словарей стилистической информацией

Всего в нашем датасете встречаются 73 различные пометы, однако ни в одном словаре не содержатся все из них. Наибольшее их количество включает словарь Д. И. Квеселевича – 45 различных помет. Словарь Т. В. Волковой использует всего на одну меньше – 44 пометы, затем словарь А. И. Федорова с 35 пометами, и, наконец, в словаре И. В. Федосова и А. Н. Лапицкого встречается всего 16 различных помет. Интересно также, что только 9 помет используются каждым автором: *бранное*, *ироническое*, *книжное*, *неодобрительное*, *презрительное*, *пренебрежительное*, *просторечное*, *разговорное*, *шутливое*.

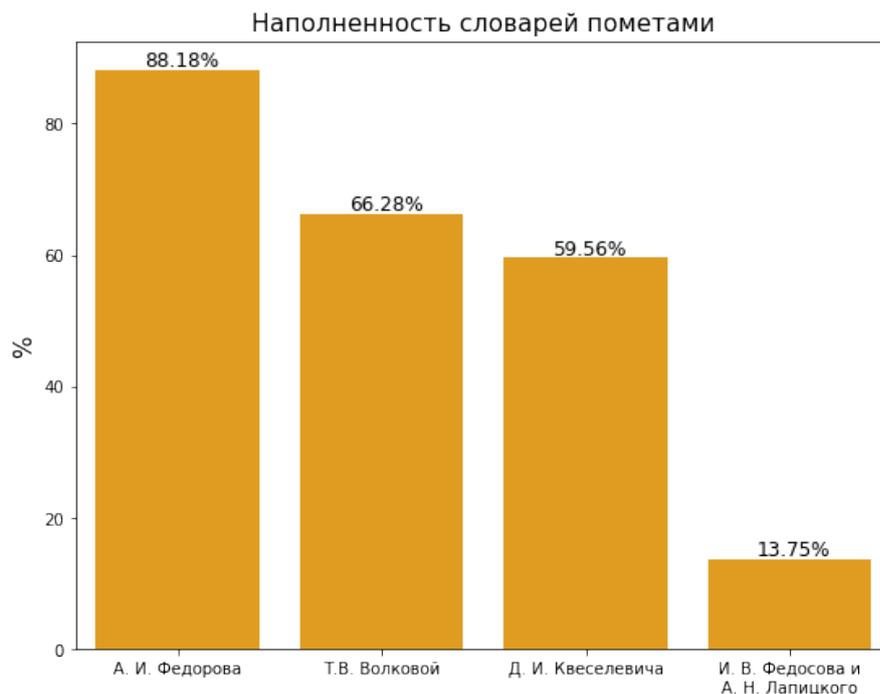


Рис. 1. Наполненность словарей пометами

Fig. 1. Fullness of dictionaries with labels

На рис. 1 представлена информация о том, у скольких процентов словарных статей есть хотя бы одна стилистическая помета. Как мы видим, словари очень сильно различаются по наполненности стилистической информацией. Так, в словаре А. И. Фёдорова для наибольшего количества фразеологизмов (88 %) присутствует стилистическая информация, в то время как в словаре И. В. Федосова и А. Н. Лапицкого такая информация есть всего чуть больше, чем у 10 % фразеологизмов.

В объединенной коллекции чаще всего фразеологизмы имеют две различные пометы (примерно 34 %), например *баба-сквернавка* (*устар.*, *бран.*). Почти так же часто встречаются фразеологизмы с одной пометой (31 %), к примеру *вороново крыло* (*экспрес.*). Около 25 % фразеологизмов в коллекции не имеют помет вовсе, например, *вести под руки* или *ставить (поднимать) на ноги*. Максимальное количество различных помет у одного фразеологизма может достигать семи единиц, таких фразеологизмов всего семь, один из них – *Вожжа (шлея) под хвост попала* (*груб.*, *неодобрит.*, *прост.*, *ирон.*, *вульг.*, *грубо-прост.*, *разг.*).

3.2. Востребованность помет

На рис. 2 приведены наиболее частотные пометы в рамках каждого из словарей.

При сравнении распределений самых частотных помет можно заметить, что они сильно различаются для представленных словарей. Так, в словарях Д. И. Квеселевича и Т. В. Волковой с заметным отрывом лидирует помета разговорное. В словаре А. И. Федорова она занимает вторую позицию, а в словаре И. В. Федосова и А. Н. Лапицкого она вообще не присутствует в десятке самых частотных.

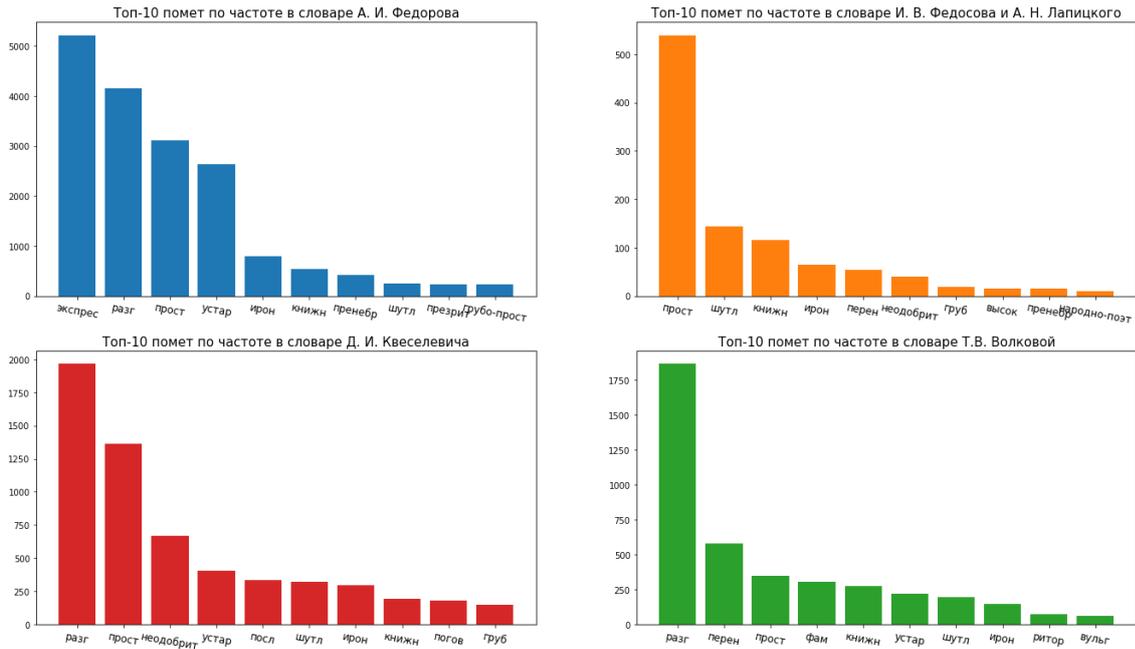


Рис. 2. Рейтинги 10 самых популярных помет в каждом словаре
 Fig. 2. The top ten popular labels in each dictionary

Только 4 пометы (*просторечное, ироническое, книжное, шутливое*) присутствуют в топах по частотности во всех словарях, при этом они занимают различные позиции.

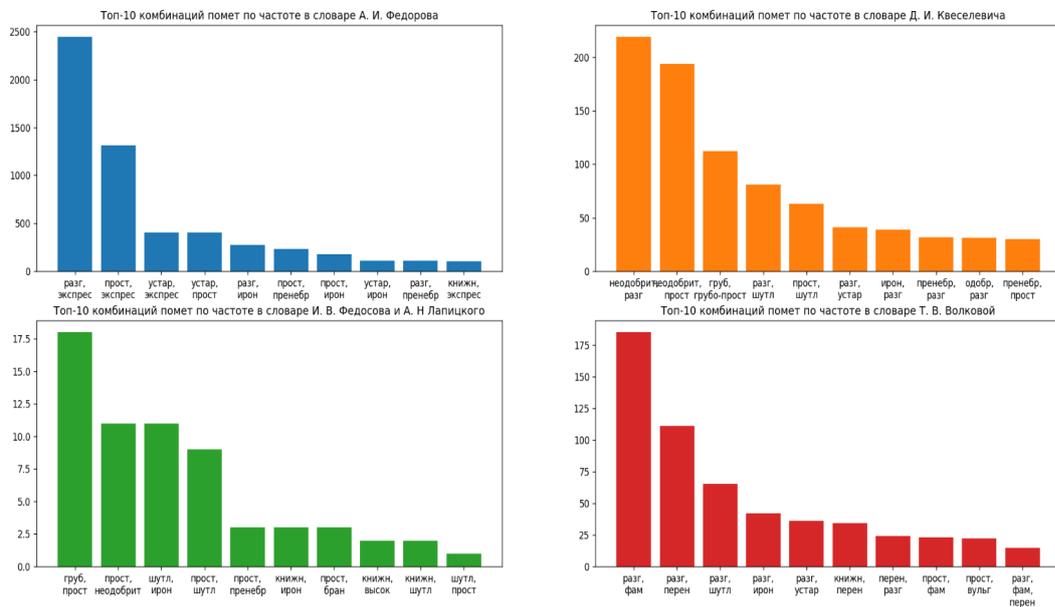


Рис. 3. Рейтинги 10 популярных комбинаций помет в каждом словаре
 Fig. 3. The top ten popular label combinations in each dictionary

На рис. 3 приведено распределение самых частотных комбинаций помет. Поскольку во всех словарях встречается в среднем не более двух помет на один фразеологизм, то самые

частотные комбинации имеют также объем не более двух помет, за единственным исключением в словаре Т. В. Волковой.

Как правило, самые частотные пометы образуют самые частотные комбинации. Так, самая популярная помета в словаре А. И. Федорова (*экспрессивное*) образует комбинации с тремя другими наиболее популярными пометами этого словаря (*разговорное*, *просторечное*, *устаревшее*). Аналогично в словаре Д. И. Квеселевича первые три пометы по популярности образуют входящие в топ комбинации.

Помета *разговорное* частотно встречается в трех из четырех рассматриваемых словарей и так же часто встречается в сочетании с большим разнообразием других помет, а именно с пометами *экспрессивное*, *ироническое*, *неодобрительное*, *шутливое*, *пренебрежительное*, *фамильярное*, *переносное*. Помета *книжное* встречается реже и с более ограниченным кругом помет: *экспрессивное*, *ироническое*, *шутливое* и *переносное*. Можно заметить, что большая часть компонентов частотных комбинаций для помет *разговорное* и *книжное* пересекаются. Так, обе пометы можно часто встретить в сочетании с *экспрессивное*, *ироническое*, *шутливое*, *переносное*. *Книжное*, однако же, не отличается частотной встречаемостью с пометами *неодобрительное*, *пренебрежительное*, *фамильярное*.

3.3. Взаимосвязь помет «книжное» и «разговорное» с другими пометами

Создание электронного датасета позволило не только проследить совместную встречаемость помет через частотные комбинации, но и статистически оценить ее значимость. Рассмотрим сочетания интересующих нас помет *книжное* и *разговорное* с другими пометами и воспользуемся тестом хи-квадрат [Corder, Foreman, 2014. С. 172–186], чтобы определить статистическую значимость. В качестве нулевой гипотезы будет использоваться допущение о независимости пары стилистических помет внутри одного словаря.

В рамках вычислительного эксперимента для всех фразеологизмов, содержащихся в каждом словаре, была рассчитана совместная встречаемость помет (в форме contingency table), после чего был применен тест хи-квадрат.

Как уже отмечалось, словари в рамках датасета различаются по набору помет, используемых авторами, поэтому не все интересующие нас пары помет были найдены в каждом рассмотренном словаре. Более того, тест имело смысл проводить лишь в тех случаях, когда была выявлена одновременная встречаемость обеих помет в рамках одного словаря, поскольку в случае, если пометы не встречались одновременно ни в одном фразеологизме в рамках словаря, мы исходили из их независимости друг от друга.

В качестве уровня значимости для проверки статистической гипотезы был принят 0,05. Таким образом, если полученное в результате применения критерия *p-value* было менее принятого уровня значимости, гипотеза о независимости стилистических помет отклонялась.

Как видно из табл. 2, для ряда случаев выявлена статистически значимая зависимость, хотя анализируемые пометы встретились вместе в конкретном словаре лишь один раз. Из-за единичной совместной встречаемости эти результаты нельзя признать полностью достоверными. Поэтому будем рассматривать наблюдения, у стилистических помет которых относительная совместная встречаемость превышает 0,1 %. Такие случаи выделены в таблице полужирным. В частности, значимой признана взаимосвязь помет *книжное* и *высокое*, *устаревшее* и *книжное*, что подтверждает наблюдения А. Н. Баранова и Д. О. Добровольского [Баранов, Добровольский, 2014. С. 175–178]. Тем не менее для каждой пары статистически значимая совместная встречаемость обнаружена только в одном словаре, что говорит о разной силе взаимосвязи данных помет в представлении разных лексикографов. Значимая взаимосвязь выявлена между пометой *разговорное* и эмоционально-экспрессивными пометами *фамильярное*, *шутливое*,

Таблица 2

Статистическая значимость совместной встречаемости помет

Table 2

Statistical significance of co-occurrence of labels

Пара стилистических помет	Автор(ы) словаря	p-value, Chi-squared test	Результат	Совместная встречаемость, абс. (%)
Книжн.-высок.	А. И. Федоров	< 0,0001	Зависимость	45 (0,397)
<i>Книжн.-высок.</i>	И. В. Федосов, А. Н. Лапицкий	< 0,0001	Зависимость	3 (0,045)
<i>Книжн.-высок.</i>	Д. И. Квеселевич	0,486	Независимость	2 (0,045)
Устар.-книжн.	А. И. Федоров	< 0,0001	Зависимость	60 (0,52)
<i>Устар.-книжн.</i>	И. В. Федосов, А. Н. Лапицкий	< 0,0001	Зависимость	1 (0,033)
<i>Устар.-книжн.</i>	Д. И. Квеселевич	0,26	Независимость	3 (0,68)
<i>Устар.-книжн.</i>	Т. В. Волкова	0,58	Независимость	3 (0,147)
<i>Разг.-фам.</i>	А. И. Федоров	0,786	Независимость	3 (0,026)
<i>Разг.-фам.</i>	И. В. Федосов, А. Н. Лапицкий	< 0,001	Зависимость	1 (0,033)
Разг.-фам.	Т. В. Волкова	< 0,001	Зависимость	110 (5,418)
Разг.-шутл.	А. И. Федоров	0,048	Зависимость	52 (0,458)
<i>Разг.-шутл.</i>	И. В. Федосов, А. Н. Лапицкий	0,027	Зависимость	1 (0,033)
Разг.-шутл.	Д. И. Квеселевич	0,003	Зависимость	49 (1,112)
Разг.-шутл.	Т. В. Волкова	0,001	Зависимость	47 (2,315)
<i>Разг.-шутл.-фами- льярн.</i>	И. В. Федосов, А. Н. Лапицкий	< 0,001	Зависимость	1 (0,033)
<i>Разг.-шутл.-ирон.</i>	И. В. Федосов, А. Н. Лапицкий	< 0,001	Зависимость	1 (0,033)
<i>Разг.-шутл.-ирон.</i>	Д. И. Квеселевич	0,734	Независимость	1 (0,033)
<i>Разг.-ирон.</i>	А. И. Федоров	0,645	Независимость	209 (1,844)
<i>Разг.-ирон.</i>	И. В. Федосов, А. Н. Лапицкий	< 0,001	Зависимость	1 (0,033)
<i>Разг.-ирон.</i>	Д. И. Квеселевич	0,454	Независимость	35 (0,794)
Разг.-ирон.	Т. В. Волкова	0,002	Зависимость	34 (1,675)
Разг.-презрит.	А. И. Федоров	0,002	Зависимость	38 (0,335)
<i>Разг.-презрит.</i>	И. В. Федосов, А. Н. Лапицкий	< 0,001	Зависимость	1 (0,033)
<i>Разг.-презрит.</i>	Д. И. Квеселевич	0,675	Независимость	1 (0,022)
<i>Разг.-презрит.</i>	Т. В. Волкова	0,692	Независимость	1 (0,049)
Разг.-пренебр.	А. И. Федоров	0,03	Зависимость	90 (0,794)
<i>Разг.-пренебр.</i>	И. В. Федосов, А. Н. Лапицкий	< 0,001	Зависимость	1 (0,033)
<i>Разг.-пренебр.</i>	Д. И. Квеселевич	0,173	Независимость	27 (0,612)
<i>Разг.-пренебр.</i>	Т. В. Волкова	0,07	Независимость	6 (0,296)
<i>Книжн.-торж.</i>	И. В. Федосов, А. Н. Лапицкий	< 0,001	Зависимость	1 (0,033)

ироничное, презрительное, пренебрежительное, что соответствует выводам предыдущих исследований [Ниннин, 2012]. При этом наиболее последовательно зависимость прослеживается для пары *разговорное* и *шутливое* (найдена в трех из четырех словарей).

3.4. Несовпадение помет

Для выявления степени различия в практике присвоения помет составителями различных словарей мы сравнили количество совпадений и несовпадений помет у фразеологизмов, снабженных ими как минимум в двух словарях, таких оказалось 5661. Для каждого фразеологизма мы посчитали количество общих помет во всех словарях, в которых он представлен, поделив их на общее число помет у данной единицы, затем сгруппировали (просуммировали) фразеологизмы по каждому результату. Как видно из рис. 4, в подавляющем большинстве случаев (в более чем 75 %) пометы для одного и того же фразеологизма полностью не совпадают в разных словарях. Лишь у приблизительно 5 % фразеологизмов пометы совпадают полностью во всех словарях, где этот фразеологизм присутствует.



Рис. 4. Доля совпадающих стилистических помет у фразеологизмов
Fig. 4. Proportion of overlapping stylistic labels in idioms

Таким образом, мнения лексикографов о стилистической характеристике фразеологизмов значительно различаются. Дальнейший сопоставительный анализ позволил выявить основные случаи несовпадения, затрагивающие пометы *книжное* и *разговорное*. Они могут быть сгруппированы следующим образом.

1. Квалификация фразеологизма как разговорного в одном словаре и как просторечного – в другом. Будучи распространенными типами помет, *разговорное* и *просторечное* часто не совпадают. Например, *бедовая голова*, *буйная голова*, *виданное ли дело*, *голова трежит*, *дать по носу*, *как заведенный*, *на карачках*, *умываться кровью* и др. снабжены пометой *разговорное* в словаре А. И. Федорова и пометой *просторечное* в словаре Д. И. Квеселевича. Интересно отметить, что имеет место и обратная ситуация: фразеологизмам *войти в раж*, *выбиваться из круга*, *дело не клеится*, *отмочить штуку* и др. присвоена помета *разговорное* в словаре Д. И. Квеселевича и *просторечное* в словаре А. И. Федорова. Данные наблюдения свидетельствуют не только о возможности разного понимания признаков, стоящих за указанными пометами, но и о тонкой грани между этими стилистическими характеристиками.

2. Несовпадение пометы *книжное*, с одной стороны, и *просторечное* или *разговорное*, с другой стороны. Подобное резкое расхождение в регистрающей характеристике фразеологизмов наблюдается в следующих примерах: *в разрезе*, *кипеть ключом* (разг. в словаре А. И. Федорова, *книжн.* в словаре Т. В. Волковой), *вращать глазами* (прост. в словаре А. И. Федорова, *книжн.* в словаре Т. В. Волковой), *за семью печатями*, *по сей день* (разг. в словаре А. И. Федорова, *книжн.* в словаре Д. И. Квеселевича), *по сей день* (*книжн.* в словаре А. И. Федорова, *прост.* в словаре И. В. Федосова и А. Н. Лапицкого).

3.5. Квантитативные характеристики фразеологизмов с пометами «книжное» и «разговорное» в корпусе

Для исследования принадлежности фразеологизмов к книжным или разговорным исходя из практики их использования в соответствующем дискурсе был использован метод подсчета вхождений фразеологизмов в разговорный и художественный корпус.

В табл. 3 представлено статистическое обобщение результатов корпусного исследования фразеологизмов, отмеченных пометой *разговорное* и пометой *книжное*.

Таблица 3

Статистические показатели корпусных данных

Table 3

Statistical summary of the corpus data

Метрика	Помета	
	Разговорное	Книжное
Точность	37,59 %	47,69 %
Среднее	Разговорный: 24,65 Художественный: 17,25	Разговорный: 3,35 Художественный: 8,2
Медиана	Разговорный: 3 Художественный: 3	Разговорный: 2 Художественный: 2
Мода	Разговорный: 1 Художественный: 1	Разговорный: 1 Художественный: 1
Итого	1793	195

Здесь *Разговорный* означает корпус разговорной речи, а *Художественный* – корпус художественной литературы. Среди 3118 фразеологизмов, охарактеризованных как разговорные, в корпусах встречаются лишь 57,5 %, что может отчасти быть объяснено редкостью использования фразеологизмов в речи и относительно небольшими объемами корпусов. Данные также говорят о том, что 37,59 % этих фразеологизмов действительно встречаются чаще в разговорном корпусе. Из 470 книжных фразеологизмов коллекции хотя бы одно вхождение в корпус было найдено для 195 единиц, при этом в литературных текстах, т. е. в художественном корпусе, чаще, чем в разговорных, можно встретить 47,69 % этих фразеологизмов.

Значение моды в обоих случаях принимает единицу, а значит фразеологизмы чаще всего встречаются лишь раз в корпусе, а среднее число употреблений разговорных фразеологизмов заметно выше, чем у книжных. Это может быть обусловлено тем, как в литературных произведениях авторы наделяют своих персонажей живой речью с использованием разговорных фразеологизмов, что дает высокие результаты как в разговорном корпусе, так и в литературном.

Однако среди тех единиц, которые имеют помету *книжное*, но чаще встретились в разговорном корпусе, действительно много фразеологизмов: фразеологических сращений (*в поте лица*),

фразеологических единств (*точка зрения, поле зрения, потерпеть фиаско, идти навстречу*), фразеологических выражений (*вопрос жизни и смерти, пути господни неисповедимы*). В то же время единицы, оснащенные пометой *разговорное*, но встретившиеся чаще в литературном корпусе, в большем количестве случаев являются, к примеру, служебными словами и частями речи: союзные образования (*а то, и то, и есть*), частицы (*и так, всё же*), наречные образования (*до конца, на потом, на ходу*), вводные выражения (*скорее всего, к счастью, с радостью*). С одной стороны, такие единицы не являются фразеологизмами в общепринятом понимании, поскольку они не соответствуют ряду признаков, например, не все входящие в них единицы являются лексемами [Baldwin and Kim, 2010]. С другой стороны, некоторые исследователи относят подобные случаи к грамматическим фразеологизмам [Баранов, Добровольский, 2014. С. 84], а многие авторы включают их в состав фразеологических словарей. В данном исследовании мы придерживаемся широкого понимания фразеологизма, поэтому не стали исключать такого рода единицы из рассмотрения, однако признаем, что вопрос о границах фразеологизма является дискуссионным. Приведенные примеры подчеркивают проблему присвоения авторами рассматриваемых помет, а также затрагивают вопрос о принципах составления словарей.

Еще одна проблема, поднятая в данной статье, – это одновременное присваивание фразеологизмам помет *разговорное* и *книжное* разными авторами. В нашей коллекции встречается 41 такой случай, причем для 11 из них не было найдено ни одного употребления в корпусах (*без руля и без ветрил, буря в стакане воды, во сто крат, кровь от крови, лить слёзы, накалить атмосферу, ни на йоту, под пятой, превратиться в слух, сидеть на пище святого Антония, сойти в могилу*). При подсчете вхождений остальных фразеологизмов в имеющиеся корпуса получилась следующая статистика:

- с преимуществом в разговорном корпусе – 5 единиц (12,2 %): *на почве, нет на свете, омыwać руки, падать духом, по пунктам*;
- с преимуществом в литературном корпусе – 23 единицы (56,1 %): например, *бить в набат, волею судьбы, воспарить духом, кипеть ключом, плыть по течению*;
- с одинаковыми результатами для обоих корпусов: 2 единицы (4,88 %): *за семью печатями, сиамские близнецы*.

Таким образом, при спорных ситуациях данные корпуса говорят скорее о книжном характере этих единиц, нежели о разговорном. На гистограмме ниже (рис. 5) отражена сумма использованных авторами помет в спорных случаях.

Заметное преобладание в случае с пометой *разговорное* имеется в словаре А. И. Федорова, что означает, что этот автор чаще других характеризует фразеологизм как разговорный, в то время как другие характеризуют его как книжный. С другой стороны, помета *книжное* чаще используется Т. В. Волковой в тех случаях, где другим или другими авторами используется помета *разговорное*.

На рис. 6 отражено отношение случаев, когда данные корпуса согласовывались с пометой автора, к общему количеству помет, присвоенных автором в спорных ситуациях.

Как можно заметить, чаще в исследуемых случаях данные согласуются со словарем И. В. Федосова и А. Н. Лапицкого (в 4 из 9 случаев). Немного меньше доля совпадений у словаря Т. В. Волковой – 42,86 % согласования, но в абсолютных числах это 12 случаев из 28. На последнем месте здесь оказался словарь А. И. Федорова, однако случаев согласования у него больше, чем у всех остальных авторов – 13.

В четырех случаях А. И. Федоровым один и тот же фразеологизм был отмечен разговорным и книжным в разных значениях: *во сто крат, класть жизнь, на страже, накалить атмосферу*, что подчеркивает необходимость унификации и формализации принципов присвоения фразеологизмам стилистических помет.

У корпусного метода, и использования лемматизированного корпуса в частности, есть свои ограничения. В первую очередь это совпадение в лемматизированном виде фразеологизмов и простых словосочетаний:

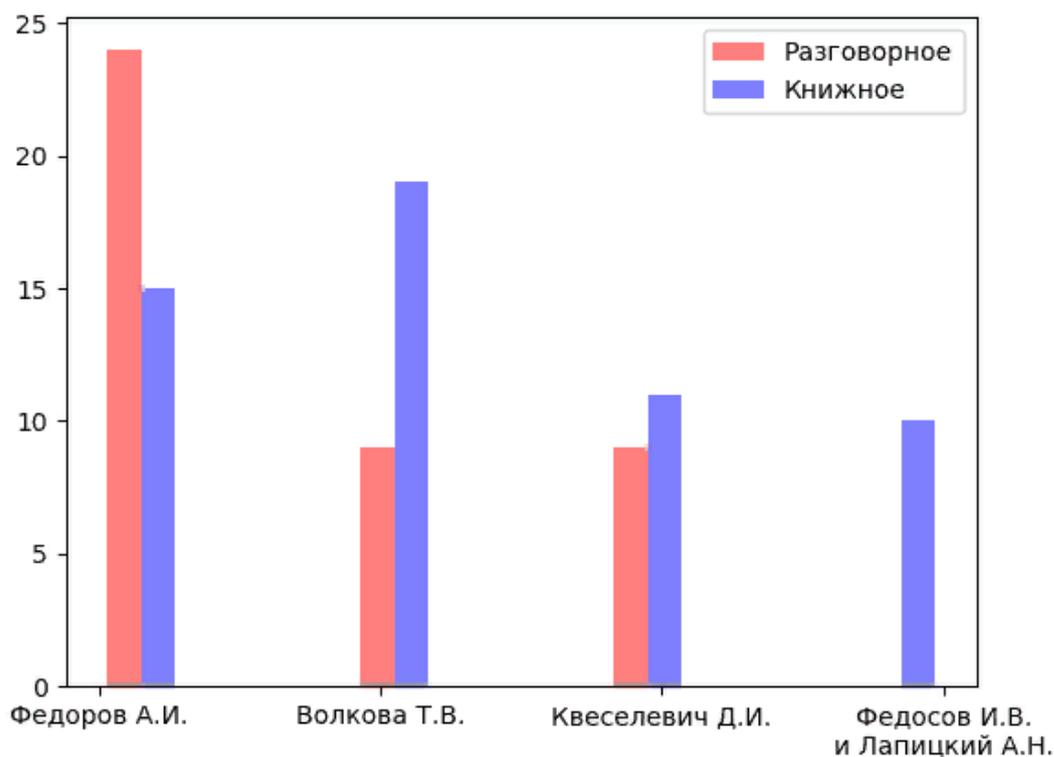


Рис. 5. Использование авторами словарей помет *книжное* и *разговорное* в спорных случаях
Fig. 5. Dictionary authors' use of the labels *bookish* and *colloquial* in controversial cases

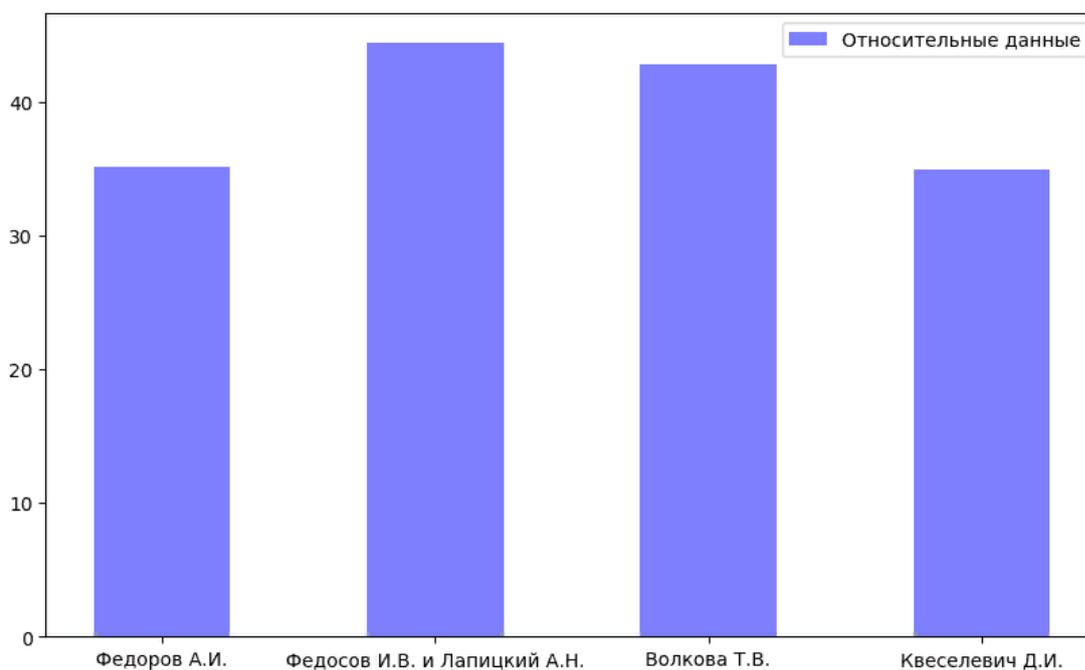


Рис. 6. Доля случаев согласования корпусных данных для каждого автора
Fig. 6. The proportion of corpus data agreement for each author

1. а) ...Как жила старая, я даже и не представляю. Хоронили её **всем миром**. Я тоже участие принял... [лем. **весь мир**]

б) ...Он жаждет исполнить свою мечту – заставить **весь мир** склонить головы...

2. а) ...она сама умудрялась при этом быть почти всегда **на нуле** и куда утекало все... [лем. **на ноль**]

б) ...его слух уловил сказанную учительницей фразу: «И запомните раз и навсегда: **на ноль** делить нельзя!»...

Сложность также вызывает учет дистантного расположения элементов фразеологизма:

1. а) ...Да ну тебя, **язык без костей**...

б) ...**Язык** у неё был, как говорится, **без костей**...

2. а) ...Мы несколько часов вместе с **Наумом ломали голову** над твоей психограммой...

б) ...**ломала** она себе **голову** и ничего не могла придумать...

Еще одна особенность, не позволяющая учесть все использования фразеологизма в корпусе, это вариативность порядка его слов:

а) ...**Хорошо язык подвешен**, любит друг мой поболтать...

б) ...**Весь такой обходительный. И язык хорошо подвешен**. Он нас быстро очаровал...

в) ...и даёт послушать людей с **хорошо подвешенными языками**, которые...

Более того, в примере в) присутствует также грамматическая вариативность, а именно использование полной формы причастия вместо краткой, и, как следствие, другой синтаксический статус фразеологизма.

Отметим также случаи совпадения фразеологизированных выражений с предложно-падежной формой слова. Например, в нашей коллекции встречаются такие фразеологизмы, как *в руки, на руках, в головах, на голову*, которые в корпусах встречаются как в своем фразеологическом значении (например, ...*ускользнувшая из профессионального сообщества в руки* [в значении 'в собственность'] *научно-популярных журналистов...*, ...*не пускали наблюдателей, так как у них на руках* [в значении 'в наличии'] *не было мандатов...*), так и в качестве омонимичной предложно-падежной формы (...*шаман взял в руки бубен...*, ...*Илья отнес на руках Евдокию в комнату...*).

Тем не менее подобные примеры не оказали существенного влияния на итоговый результат значительно, так как количество единиц, затронутых указанными ограничениями, в общем объеме электронной коллекции мало.

Таким образом, результаты применения корпусно-ориентированного подхода показали разную эффективность для различных групп фразеологизмов, словники которых во фразеологических словарях необходимо признать гетерогенными с точки зрения типов содержащихся в них неоднословных единиц – не только фразеологизмов в строгом понимании, но и коллокаций. Работа с некоторыми единицами затрудняется из-за технических ограничений, с другими – ввиду низкой встречаемости фразеологизмов в целом. Увеличение объема корпусов и фокус на отдельных типах фразеологизмов в будущем могут повысить эффективность подхода.

Заключение

Данное исследование было посвящено проблеме необходимости унификации стилистических помет *книжное* и *разговорное* во фразеологических словарях и направлено на анализ распределения помет и их взаимосвязи на материале нескольких фразеологических словарей, объединенных в электронный датасет. Датасет объединил четыре словаря, что позволило использовать количественные и статистические методы на большом объеме материала.

Результаты продемонстрировали значительную разницу между словарями как в наборе помет, так и в наполненности стилистической информацией. Помета *разговорное* является самой частотной в двух из четырех словарей, но не входит в десятку самых популярных

в одном из словарей, что является лишь одной иллюстрацией различий в востребованности конкретных помет. Были выявлены и значительные несоответствия в характеристике одного фразеологизма в разных словарях. Ключевые несовпадения, затрагивающие пометы *книжное* и *разговорное*, включают квалификацию фразеологизма как разговорного в одном словаре и как просторечного – в другом, а также присвоение фразеологизму пометы *книжное* в одном словаре и *просторечное* или *разговорное* – в другом.

Сложные семантические и парадигматические отношения между отдельными пометами отражают объективную взаимосвязь между стилистической, дискурсивной, эмоционально-экспрессивной и др. характеристиками фразеологизмов, отмеченную предыдущими исследователями: в частности, статистически подтверждена связь помет *книжное* и *высокое*, *устаревшее* и *книжное*. Примечательна взаимосвязь, выявленная для пометы *разговорное* и эмоционально-экспрессивных помет, однако ни в одном случае совместная встречаемость не признана значимой во всех проанализированных словарях.

Корпусно-ориентированный подход, направленный на верификацию помет *книжное* и *разговорное* с помощью данных частотности фразеологизмов в корпусах художественной литературы и разговорной речи соответственно, продемонстрировал, что менее чем в половине случаев фразеологизм тяготеет к преобладающему употреблению в корпусе, соответствующему присвоенной помете. Спорные случаи, получившие неоднозначную квалификацию в словарях, продемонстрировали преимущественное тяготение к характеристике *книжное*. Технические ограничения и проблема гетерогенности анализируемых словников могут быть в перспективе решены за счет увеличения объема корпусов и локализации исследования, в частности его сосредоточения на фразеологизмах в отличие от других типов коллокаций.

Полученные результаты свидетельствуют о значительных несоответствиях в стилистической характеристике фразеологизмов во фразеологических словарях и могут быть использованы для унификации системы помет и принципов их присвоения.

Список литературы

- Андрианова Д. В.** Изменение лексикографических принципов представления фразеологизмов и устойчивых сочетаний слов в «Словаре современного русского литературного языка» и в «Большом академическом словаре русского языка» // Вопросы лексикографии. 2019. № 15. С. 71–82.
- Аюпова Р. А.** Лексикографическое описание функционально-стилистического аспекта фразеологической коннотации (на материале английского, русского и татарского языков) // Вестник университета Российской академии образования. 2008. № 5. С. 78–82.
- Баранов А. Н., Вознесенская М. М., Добровольский Д. О., Киселева К. Л., Козеренко А. Д.** Фразеологический объяснительный словарь русского языка. М., 2009. 704 с.
- Баранов А. Н., Добровольский Д. О.** Аспекты теории фразеологии. М.: Знак, 2008. 656 с.
- Баранов А. Н., Добровольский Д. О., Киселева К. Л., Козеренко А. Д.** Словарь-тезаурус современной русской идиоматики. М., 2007.
- Баранов А. Н., Добровольский Д. О.** Основы фразеологии (краткий курс). М.: ФЛИНТА, 2014. 312 с.
- Волкова Т. В.** Универсальный фразеологический словарь русского языка. М.: Вече, 2000. 463 с.
- Есетова Н. С.** Проблема стилистической характеристики фразеологизмов // Язык и культура: Сборник материалов XXVII Международной научно-практической конференции. Новосибирск, 2017. С. 187–194.
- Квеселевич Д. И.** Русско-английский фразеологический словарь. М.: Рус. яз., 1998. 704 с.

- Козеренко А. Д. и др.** Квантитативные характеристики идиомы как показатель ее стилистических свойств // Труды международной конференции «Корпусная лингвистика – 2013». СПб.: С.-Петербургский гос. университет, 2013. С. 309–317.
- Лавриненко В. А.** Факторы, определяющие принадлежность фразеологических единиц к разговорному стилю (на материале французского языка) // Вестник Таганрогского института им. А. П. Чехова. 2017. № 2. С. 18–26.
- Ниннин Г. О.** Стилистическая характеристика фразеологизмов в русских фразеологических словарях // Вестник Тамбовского ун-та. Серия: Гуманитарные науки. 2012. № 9. С. 280–283.
- Скляревская Г. Н.** Современная русская лексикография: достижения и лакуны // Славянская лексикография. М.: Азбуковник, 2013. С. 579–615.
- Федоров А. И.** Фразеологический словарь русского литературного языка. М.: Астрель: АСТ, 2008. 878 с.
- Федосов И. В., Лапицкий А. Н.** Фразеологический словарь русского языка. М.: ЮНБЕС, 2003. 608 с.
- Хохлова М. В., Мамаев И. Д.** Разработка базы данных коллокаций: обзор золотого стандарта на примере атрибутивных словосочетаний // Труды международной конференции «Корпусная лингвистика – 2021». СПб.: Скифия-принт, 2021. С. 370–379.
- Corder G. W., Foreman D. I.** Nonparametric Statistics: A Step-by-Step Approach. John Wiley & Sons, 2014. 273 p.
- Dobrovoljc K.** Identifying dictionary-relevant formulaic sequences in written and spoken corpora // International Journal of Lexicography. 2020. Vol. 33, № 4. P. 417–442.
- Fellbaum C.** The treatment of multi-word units in lexicography // The Oxford handbook of lexicography. Oxford, 2015.
- Gantar P. et al.** Multiword expressions: between lexicography and NLP // International Journal of Lexicography. 2019. Vol. 32, № 2. P. 138–162.
- Masini F.** Multi-Word Expressions and Morphology // Oxford Research Encyclopedia of Linguistics. Oxford, 2019.
- Villavicencio A., Idiart M.** Discovering multiword expressions // Natural Language Engineering. 2019. Vol. 25, № 6. P. 715–733.

References

- Andrianova, D. V.** The changes in the lexicographic principles of representation of phraseological units and stable word combinations in the Dictionary of the Modern Russian Literary Language and in the Large Academic Dictionary of the Russian Language. *Voprosy leksikografii*, 2019, no. 15, p. 71–82. (in Russ.)
- Аурпова, Р. А.** Lexicographic description of the functional-stylistic aspect of the phraseological connotation (in English, Russian and Tatarian). *Herald of the University of the Russian Academy of Education*, 2008, no. 5, p. 78–82. (in Russ.)
- Baranov, A. N., Dobrovol'skij, D. O.** Aspects of the phraseology theory. Moscow, Znak, 2008. (in Russ.)
- Baranov, A. N., Dobrovol'skij, D. O.** Basics of phraseology (a short course). Moscow, FLINTA, 2014. (in Russ.)
- Baranov, A. N., Dobrovol'skij, D. O., Kiselyova, K. L., Kozerenko, A. D.** An idiomatic thesaurus of Modern Russian. Moscow, 2007. (in Russ.)
- Baranov, A. N., Voznesenskaya, M. M., Dobrovol'skij, D. O., Kiselyova, K. L., Kozerenko, A. D.** Phraseological Explanatory Dictionary of the Russian Language. Moscow, 2009. (in Russ.)

- Corder, G. W., Foreman, D. I.** Nonparametric Statistics: A Step-by-Step Approach. John Wiley & Sons, 2014.
- Dobrovoljc, K.** Identifying dictionary-relevant formulaic sequences in written and spoken corpora. *International Journal of Lexicography*, 2020, vol. 33(4), p. 417–442.
- Esetova, N. S.** The issue of stylistic characteristics of phraseological units. *Language and Culture: Proceedings of the XXVII International academic and practical conference*. Novosibirsk, 2017, p. 187–194. (in Russ.)
- Fedorov, A. I.** Phraseological Dictionary of the Standard Russian Language. Moscow, Astrel, AST, 2008. (in Russ.)
- Fedosov, I. V., Lapitskij, A. N.** Phraseological Dictionary of the Russian Language. Moscow, IUNVES, 2003. (in Russ.)
- Fellbaum, C.** The treatment of multi-word units in lexicography. *The Oxford Handbook of Lexicography*. Oxford, 2015.
- Gantar, P.** et al. Multiword expressions: between lexicography and NLP. *International Journal of Lexicography*, 2019, vol. 32(2), p. 138–162.
- Khokhlova, M. V., Mamaev, I. D.** Building Russian collocations database: a case study of gold standard for ADJ-NOUN collocations. *Proceedings of the international conference “Corpus linguistics – 2021”*. St. Petersburg, Skifiya-print, 2021, p. 370–379. (in Russ.)
- Kozerenko, A. D. et al.** Correlation between idiom frequency and its stylistics properties. *Proceedings of the international conference “Corpus linguistics – 2013”*. St. Petersburg, St. Petersburg University, 2013, p. 309–317. (in Russ.)
- Kveselevich, D. I.** Russian-English Phraseological Dictionary. Moscow, Rus. yaz., 1998. (in Russ.)
- Lavrinenko, V. A.** Factors defining phraseological units as part of the conversational style. *Vestnik Taganrozkogo instituta imeni A. P. Chekhova*, 2017, no. 2, p. 18–26. (in Russ.)
- Masini, F.** Multi-Word Expressions and Morphology. *Oxford Research Encyclopedia of Linguistics*. Oxford, 2019.
- Ninnin, G. O.** Stylistic characteristic of phraseology units in Russian phrasebook. *Tambov University Review. Series: Humanities*, 2012, no. 9, p. 280–283. (in Russ.)
- Sklyarevskaya, G. N.** Modern Russian lexicography: achievements and lacunas. *Slavic lexicography*. Moscow, Izdatel'skij tsentr «Azbukovnik», 2013, p. 579–615. (in Russ.)
- Villavicencio, A., Idiart, M.** Discovering multiword expressions. *Natural Language Engineering*, 2019, vol. 25(6), p. 715–733.
- Volkova, T. V.** Universal Phraseological Dictionary of the Russian Language. Moscow, Veche, 2000. (in Russ.)

Информация об авторах

Арбузова Екатерина Михайловна, data-инженер департамента данных и общих сервисов ИТ В2С ПАО «Сбербанк»

Дурандин Олег Владимирович, независимый исследователь

Карабасова Юлия Алексеевна, помощник специалиста отдела оптимизации бизнес-процессов ООО «ОК “Вета”»

Климова Маргарита Андреевна, кандидат филологических наук, доцент департамента фундаментальной и прикладной лингвистики Национального исследовательского университета «Высшая школа экономики»

Information about the Authors

Ekaterina M. Arbuzova, data-engineer, Department of Data and General IT Services B2C PJSC “Sberbank”, Nizhny Novgorod, Russia

Oleg V. Durandin, independent scholar, Nizhniy Novgorod, Russia

Yulia A. Karabasova, assistant specialist of the department of business process optimization, Veta Expert Group, Nizhniy Novgorod, Russia

Margarita A. Klimova, Candidate of Philology, Associate Professor, Department of Fundamental and Applied Linguistics, HSE University, Nizhniy Novgorod, Russia

*Статья поступила в редакцию 20.05.2023;
одобрена после рецензирования 03.09.2023; принята к публикации 04.09.2023
The article was submitted 20.05.2023; approved after reviewing 03.09.2023;
accepted for publication 04.09.2023*