

RU

Алгоритмы поиска фразеологизмов в лингвистическом корпусе с морфологической разметкой (индоевропейские языки)

Горожанов А. И.

Аннотация. Цель настоящего прикладного исследования – построить алгоритмы поиска фразеологизмов, совместимые с разработанной нами ранее моделью лингвистического корпуса с морфологической разметкой по правилам NLP-библиотеки spaCy. Научная новизна обусловлена тем, что впервые в русле корпусного подхода предлагается набор универсальных способов поиска фразеологизмов с минимальной долей ручного труда и с применением элементов сквозных цифровых технологий. В ходе исследования были описаны технические параметры подлежащих поиску фразеологизмов; изучены возможности авторского корпусного менеджера в рамках ручного и специального ручного запросов; разработаны и апробированы на материале репрезентативного корпуса текстов немецкоязычных СМИ два алгоритма для двухступенчатого поиска отдельных фразеологических единиц и их групп; приведены развернутые примеры результатов поисковых запросов. В результате доказываемся состоятельность разработанных алгоритмов, экспериментально устанавливается, что погрешность при поиске лежит в допустимом диапазоне 0-14,8%.

EN

Algorithms for searching phraseological units in a linguistic corpus with morphological markup (Indo-European languages)

Gorozhanov A. I.

Abstract. The aim of the applied research is to build algorithms for searching phraseological units that are compatible with our previously developed model of a linguistic corpus with morphological markup according to spaCy rules. The scientific novelty is due to the fact that for the first time, within the framework of the corpus approach, a set of universal ways to search for phraseological units is proposed with a minimum amount of manual labor and using elements of end-to-end digital technologies. During the study, the technical parameters of phraseological units to be searched were described; the capabilities of the author's corpus manager within the framework of manual and special manual queries were examined; two algorithms for a two-stage search for individual phraseological units and their groups were developed and tested on the basis of a representative corpus of texts from German-language media; detailed examples of search results were provided. As a result, the consistency of the developed algorithms has been proved and it has been experimentally established that the search error lies in the acceptable range of 0-14.8%.

Введение

Современное состояние корпусной лингвистики, ориентированной на активное использование сквозных технологий цифровой экономики (большие данные, искусственный интеллект), обязывает современных исследователей стремиться к поиску решений, которые бы значительно сократили долю ручного труда при отборе и разметке материала и существенно повысили бы уровень автоматизации всех связанных с обработкой естественного языка процессов.

Среди технологий искусственного интеллекта большую важность имеют не только искусственное зрение и распознавание и синтез речи, но и обработка естественного языка (*Natural Language Processing – NLP*), которая играет важную роль в реализации национальной программы «Цифровая экономика Российской Федерации» (Бирюков, Терещенко, 2023, с. 106).

К числу языковых явлений, которые представляют сложность для автоматического (программного) обнаружения, относятся фразеологизмы и шире – различные словосочетания, например, для английского языка: *X [det] Y of Z => San Gimignano is the Manhattan of the Middle Ages etc.* / Сан-Джиминьяно – это Манхэттен Средневековья и др. (здесь и далее – перевод наш. – А. Г.) (Голубкова, Бухарова, 2023, с. 13).

О высоком интересе к проблеме фразеологии свидетельствует значительное количество новейших публикаций, рассматривающих это явление с различных точек зрения: педагогики (Гусейнова, 2023; Давыдова, 2023), гендерной идентичности (Нефедова, 2023), эквивалентности (Парина, 2020; Юсипова, 2022), лингвокультуры (Никитина, 2023; Соколова, Старикова, 2023; Титова, 2023) и сравнения различных лингвокультур (Киселева, 2023). Эти, а также классические работы В. В. Виноградова (1977) и Н. М. Шанского (2012), посвященные классификации фразеологизмов, формируют теоретическую базу нашего исследования.

Непосредственно поиску фразеологизмов в лингвистических корпусах также посвящено немало исследований. Например, Е. Б. Кротова (2022, с. 88-89) приводит примеры запросов к корпусам *Национальный корпус русского языка*, *DeReKo* и *Sketch Engine* с целью нахождения фразеологизмов-конструкций и отмечает, что хотя имеющихся в корпусных менеджерах инструментов может быть достаточно, при широком варьировании в заполнении слота приходится прибегать к дополнительному анализу полученных из корпуса данных, в том числе с помощью *Excel*-таблиц или инструментов обработки естественного языка. Этот вывод свидетельствует о том, что проблема поиска фразеологизмов является нерешенной даже в таких «эталонных» разработках, как национальные корпуса, и говорит о высокой степени актуальности темы нашей работы.

Итак, для постижения поставленной цели нам необходимо решить следующие задачи:

- 1) определить характер фразеологизмов, подлежащих поиску;
- 2) рассмотреть все возможные алгоритмы поиска обозначенных фразеологизмов средствами авторского корпусного менеджера;
- 3) внести программные модификации в корпусный менеджер, позволяющие увеличить скорость и улучшить качество поиска фразеологизмов;
- 4) провести апробацию алгоритма, полученного в ходе решения третьей задачи.

В ходе решения первой задачи применяются методы анализа и описания; вторая задача решается с помощью эксперимента, объектом которого является корпусный менеджер; при решении третьей задачи привлекается метод моделирования в части разработки формальной модели (программного кода); четвертая задача решается также с помощью эксперимента – апробации формальной модели.

Основным лингвистическим материалом исследования является сбалансированный лингвистический корпус текстов онлайн-версии немецкого журнала *Spiegel* объемом 3 034 739 токенов, собранный нами в 2023 г.

Техническим инструментом исследования послужило программное обеспечение для генерации лингвистического корпуса и соответствующий корпусный менеджер, которые были разработаны в лаборатории фундаментальных и прикладных проблем виртуального образования Московского государственного лингвистического университета (Свидетельство о государственной регистрации программы для ЭВМ № 2023683209 Российская Федерация. «Генератор сбалансированного лингвистического корпуса и корпусный менеджер»: № 2023682269: заявл. 25.10.2023: опубл. 03.11.2023 / А. И. Горожанов; заявитель – федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный лингвистический университет». EDN JHFXUV).

В практическом отношении результаты исследования могут быть полезны при интерпретации текста, при анализе идиостиля автора. Кроме того, полученные алгоритмы послужат предметом рассмотрения в рамках чтения учебных дисциплин в области корпусной лингвистики и профессионально ориентированного (лингвистического) программирования.

Обсуждение и результаты

Согласно Лингвистическому энциклопедическому словарю, «фразеологизм (фразеологическая единица) – общее название семантически связанных сочетаний слов и предложений, которые, в отличие от сходных с ними по форме синтаксических структур, не производятся в соответствии с общими закономерностями выбора и комбинации слов при организации высказывания, а воспроизводятся в речи в фиксированном соотношении семантической структуры и определённого лексико-грамматического состава» (Фразеологизм. <http://rus-yaz.niv.ru/doc/dictionary/linguistic-encyclopedia/articles/232/frazeologizm.htm>).

Существуют различные классификации фразеологизмов (см., например, классификации В. В. Виноградова и Н. М. Шанского), однако в рамках нашего исследования важна скорее «техническая» классификация, которая позволит построить оптимальные универсальные алгоритмы поиска различных фразеологических единиц в корпусах текстов на различных языках. Наиболее «простыми» в этой связи являются фразеологизмы, представляющие собой неизменяющиеся последовательности, например: русск. *на воре шапка горит*, нем. *aller Anfang ist schwer* (всякое начало трудно), англ. *as flat as a pancake* (плоский/ровный как блин), исп. *¡No pasarán!* (Они не пройдут!) или лат. *manus manum lavat* (рука руку моет). Процедура их поиска сводится к простому нахождению составляющих их символов, если только они не подвергаются намеренному искажению автором, ср.:

Ein Gespenst geht um in Europa, ein Gespenst des Populismus („Abbau von Grenzen in den Köpfen“. <https://www.ekglaatzten.de/2018/03/14/abbau-von-grenzen-in-den-koepfen>). / Призрак бродит по Европе, призрак популизма –

из исходного:

Ein Gespenst geht um in Europa, ein Gespenst des Kommunismus (Маркс К., Энгельс Ф. Манифест коммунистической партии. 1848). / Призрак бродит по Европе, призрак коммунизма.

Или:

Не так страшен черт, как его малютки (Набоков В. В. Защита Лужина. 1929) – из исходного:

Не так страшен черт, как его малюют.

Дополнительную сложность может представлять возможность изменения компонентов фразеологизма по числу, временной форме и пр. и их следования в предложении в разном порядке, например: *делать* из мухи слона, *сделать* из мухи слона, из мухи слона *сделал*; *wie ein Schwein ins Uhrwerk schauen* (смотреть как баран на новые ворота), *wie Schweine ins Uhrwerk schauen* (смотреть как бараны на новые ворота); *tête à tête* (с глазу на глаз, наедине), *têtes à têtes* (~ наедине со многими) и т. п.

Более сложными для обнаружения являются фразеологизмы, части которых, в дополнение к прочему, могут быть разделены непредсказуемым набором символов, например:

День и ночь – сутки прочь или *день и ночь – вот и сутки прочь.*

Воскрешение Лазаря или (реже) *воскрешение бедного Лазаря.*

Vom Eisentaler Gletscher am Arlberg in Österreich nahm der Physiker Martin Fierz gleich ganz Abschied (Физик Мартин Фирц попрощался с ледником Эйзенталь на Арльберге в Австрии) или *Bei einer Trauerfeier in der Kieler Petruskirche haben rund 500 geladene Gäste und Besucher Abschied von Heide Simonis genommen* (Около 500 приглашенных гостей и посетителей попрощались с Хайде Симонис на панихиде в кильской церкви Святого Петра).

„Das Beispiel Montevideo zeigt, wie die wertvolle Ressource Wasser Gefahr läuft, privatisiert zu werden, gerade in Zeiten des Klimawandels“, sagt Nicola («Пример Монтевидео показывает, что ценный водный ресурс находится под угрозой приватизации, особенно во времена изменения климата», – сообщает Никола) или *Sie laufen nun Gefahr, vom Süden und vom Norden her in die Zange genommen zu werden* (Теперь они рискуют попасть в клещи с юга и севера) (Spiegel. <https://www.spiegel.de>).

Заметим, что такие сочетания в немецком языке, как *Abschied nehmen* (прощаться), *Gefahr laufen* (подвергаться опасности, рисковать) и т. п., могут определяться также как «функциональные структуры», которые представляют собой переходное явление от свободного сочетания к фразеологизму (Богуславская, 1997, с. 4).

Обозначив явления, которые подлежат поиску, обратимся к возможностям корпусного менеджера.

В самом простом случае – при поиске по простой последовательности – эффективно использовать ручной запрос (P3) по таблице предложений, который реализуется в синтаксисе SQL, например, для библеизма *das gelobte Land* (земля обетованная):

```
SELECT * FROM sents WHERE setttext LIKE '%das gelobte Land%'
```

Результатом является одно предложение:

Grenzgebiet Eagle Pass, Texas: Drohnenaufnahmen zeigen Familien, die den Strömungen des Rio Grande trotzen, um auf die andere Uferseite zu kommen – in das gelobte Land, die USA. / Пограничный район Игл-Пасс, Техас: на кадрах, снятых с помощью дрона, видно, как семьи бросают вызов течению Рио-Гранде, чтобы добраться до другого берега реки – в землю обетованную, США.

Более точный результат можно получить с помощью запроса по регулярным выражениям, например:

```
(gelobt)(\w+)\s(Land)
```

Предполагается, что фразеологизм может употребляться в родительном или дательном падеже, с определенным, неопределенным или нулевым артиклем. В результате мы находим в дополнение к полученному предложению:

Russland galt lange als eine Art gelobtes Land für deutsche Unternehmen – doch diese Zeiten sind offensichtlich vorbei. / Россия долгое время считалась своего рода «землей обетованной» для немецких компаний – но очевидно, что эти времена прошли.

Таким образом возможен поиск фразеологизмов, компоненты которых могут изменяться по грамматическим категориям, однако их позиции в предложении относительно друг друга считаются неизменными.

В этой связи интересным представляется специальный ручной запрос (P3C), для которого нами был разработан особый формальный язык, являющийся своеобразным «посредником» между человеческим языком и языком SQL. При этом пользователь корпусного менеджера может обращаться к любым морфологическим атрибутам токенов, а также вводить логические операторы *AND*, *OR*, отрицание *NOT* и группирующие скобки.

Специальный ручной запрос призван находить последовательности токенов в корпусе, причем для токенов можно задавать любые параметры, которые имеются в базе данных. Например, для поиска последовательности *Hilfe leisten* (оказывать помощь), где глагол *leisten* может стоять в любой форме, валиден следующий запрос:

```
tokenpos='NOUN' AND tokenlemma='Hilfe'#1;;tokenpos='VERB' AND tokenlemma='leisten'#1
```

Приведенная запись означает, что ищется одно существительное с леммой *Hilfe*, за которым следует один глагол с леммой *leisten*.

В результате было найдено 10 предложений с этой конструкцией, например:

Wie Cherchi berichtete, war der Fahrer eines anderen Busses der Erste, der nach dem Sturz Erste Hilfe leistete und einen Feuerlöscher hinunterwarf. / Как сообщил Черчи, водитель другого автобуса первым оказал первую помощь после падения и сбросил огнетушитель.

Или

Ein Hotelangestellter habe vor dem Eintreffen des Rettungsdienstes Erste Hilfe geleistet. / Сотрудник отеля оказал первую помощь до приезда экстренных служб.

Несмотря на результативность, заметим, что специальный ручной запрос эффективнее всего применяется для обнаружения различного вида конструкций типа:

- «прилагательное + прилагательное + существительное»,
- «слово-артикл + прилагательное + существительное»,
- «слово-артикл + любое слово + существительное»,
- «существительное + существительное в род. падеже» и мн. др., –

т. е. в случаях, где известна позиция токенов относительно друг друга, но могут варьироваться их словоформы.

Перейдем к решению третьей задачи нашего исследования. Составим требования к программному модулю для поиска фразеологизмов в корпусе, который бы учел факт нахождения компонентов фразеологизма в различных частях предложения. Ожидается, что пользователь будет вводить леммы, из которых состоит фразеологизм, а программа будет проверять наличие всех этих лемм в каждом предложении корпуса. Такой метод может иметь погрешность, однако представляется максимально универсальным. В частности, он может быть применен на материале различных языков.

После внесения программных модификаций в код корпусного менеджера была проведена серия экспериментов, которые показали результативность алгоритма под рабочим названием «Фразеология», однако и выявили его недостаток – низкую скорость работы.

В первой версии программа производила поиск всех заданных компонентов во всех предложениях корпуса. Но поскольку вероятность нахождения фразеологизма появляется только в тех предложениях, в которых все компоненты встречаются одновременно, достаточно определить наличие самого редко встречающегося компонента, а затем искать остальные в оставшемся подкорпусе. Самый редкий компонент, который определяется пользователем самостоятельно, следует в поисковом запросе первым. Например, в немецком выражении *eine Lanze für j-n brechen* (вступить за кого-л.) наиболее редким будет существительное *Lanze* (копье).

В соответствии с этими замечаниями была изготовлена вторая версия алгоритма, которая показала значительную экономию времени при выполнении тех же запросов. Например, при поиске уже упоминавшегося сочетания *Abschied nehmen* (прощаться) для лингвиста, знающего немецкий язык, очевидно, что первый компонент является встречающимся реже, чем второй. В нашем корпусе фиксируется наличие всего лишь 10 лемм *Abschied*. Таким образом, глагол *nehmen* ищется на второй ступени реализации алгоритма не в 175 266 предложениях, а только в 10.

Тестирование второй версии алгоритма «Фразеология» показало высокую скорость, но также позволило сформулировать предложения по повышению его эффективности в части количества подлежащих поиску фразеологизмов. Нами было выдвинуто требование о возможности нахождения путем старта одной поисковой процедуры N фразеологизмов, где N ограничено только лишь возможностями исследователя (мощностью программно-аппаратного обеспечения и объемом подготовленного списка фразеологизмов).

Созданный в соответствии с этими требованиями алгоритм получил рабочее название «Фразеология 2». Оба новых типа поиска активируются в одноименных подменю основного меню «Инструменты» графического интерфейса пользователя корпусного менеджера (см. Рисунок 1).

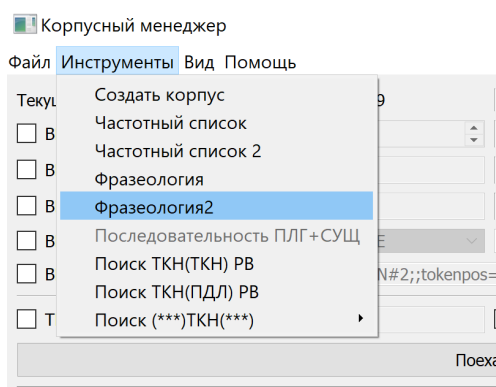


Рисунок 1. Графический интерфейс пользователя. Меню «Инструменты»

Четвертая задача работы заключалась в апробации созданных алгоритмов. Для более обстоятельной проверки «Фразеологии» был реализован ряд запросов, из которых мы приведем примеры для сочетаний *Arbeit leisten* (исполнять работу), *Aufsehen erregen* (вызвать переполох), *aus der Reihe tanzen* (выбиваться из общей картины) и *Spuren hinterlassen* (оставлять следы).

По запросу *leisten, Arbeit* на первой ступени было отобрано 273 предложения, содержащие компонент *leisten*. В этом подкорпусе поиск компонента *Arbeit* выдал 27 соответствий, из которых 23 предложения содержали заданное сочетание, например:

Das kalifornische Gesetz verbietet es Arbeitgebern, Arbeitnehmern aufgrund ihres Geschlechts, ihrer Rasse oder ihrer ethnischen Zugehörigkeit weniger zu zahlen als Kollegen, die „im Wesentlichen ähnliche Arbeit“ leisten. / Закон Калифорнии запрещает работодателям платить работникам меньше, чем их коллегам, выполняющим «по существу аналогичную работу», из-за их пола, расы или этнической принадлежности.

В четырех случаях предложения содержали одновременно леммы *leisten* и *Arbeit*, однако не в составе фразеологической единицы, например:

Helen Ackah, Penderlin „Ich kann mir das wirklich nicht leisten, weil ich fünf Tage die Woche zur Arbeit fahre“. / Хелен Ака, живет далеко от работы: «Я действительно не могу себе этого позволить, потому что езжу на работу пять дней в неделю».

Таким образом, поиск сочетания *Arbeit leisten* выявил погрешность на уровне 14,8%, что рассматривается нами как допустимый показатель.

По запросу *Aufsehen, erregen* вначале были отобраны 87 предложений, количество которых далее было сведено до 14. Из них все 14 содержали заданный фразеологизм, т. е. погрешность отсутствовала, например:

Dabei erregte die Aktion zum Abschluss des Turniers in Australien und Neuseeland international Aufsehen und sorgte selbst in spanischen Regierungskreisen für Empörung. / Акция по окончании турнира в Австралии и Новой Зеландии вызвала международный резонанс и даже вызвала возмущение в правительственных кругах Испании.

По запросу *tanzen, Reihe, aus* было найдено одно совпадение:

Das Lager des republikanischen Vorsitzenden des Repräsentantenhauses, Kevin McCarthy, warf den Rechtsaußen der Partei vor, aus der Reihe zu tanzen. / Лагерь республиканского лидера палаты представителей Кевина Маккарти обвинил правых экстремистов в том, что они не вписываются в общую картину.

Здесь небольшой количественный результат не позволяет сформировать какое-либо представление о величине погрешности.

Наконец, запрос *Spur, hinterlassen* вывел данные о 156 леммах *Spur*. На второй ступени поиска количество предложений сократилось до 17, причем погрешность в этом случае также отсутствовала. Все предложения содержали заданную фразеологическую единицу, например:

Der Beginn der Industrialisierung hingegen, auch ein denkbarer Start des Anthropozäns, hat in Gesteinsschichten in vielen Regionen der Welt keine Spuren hinterlassen, teilte die AWG mit. / «Однако начало индустриализации, а также возможное начало антропоцена не оставило никаких следов в слоях горных пород во многих регионах мира», – сообщает AWG.

Для технической апробации алгоритма «Фразеология 2» нами был составлен комплексный поисковый запрос, содержащий небольшой произвольный набор фразеологизмов с различной частотой употребления в современном немецком языке:

- *Lanze, brechen, für* для *eine Lanze für j-n brechen* (вступить за кого-л.);
- *Canossa, nach* для *nach Canossa gehen* (идти в Каноссу – инцидент покаяния императора Генриха IV перед папой Григорием VII);
- *Scherflein, beitragen* для *sein Scherflein beitragen* (внести свою лепту);
- *Samariter, barmherzig* для *barmherziger Samariter* (добрый самаритянин);
- *Stich, lassen* для *im Stich lassen* (бросить на произвол судьбы);
- *Geige, erst, spielen* для *erste Geige spielen* (играть первую скрипку);
- *Beschluss, fassen* для *Beschluss fassen* (принимать решение);
- *Entscheidung, treffen* для *Entscheidung treffen* (принимать решение [при выборе из вариантов]);
- *pleite, gehen / pleite, sein* для *pleite gehen / pleite sein* (обанкротиться).

Результаты выполнения запроса представим в табличной форме (см. Таблицу 1).

Таблица 1. Результаты поиска по алгоритму «Фразеология 2»

Запрос	Найдено (ед.)	Погрешность (%)
Lanze, brechen, für	0	–
Canossa, nach	0	–
Scherflein, beitragen	0	–
Samariter, barmherzig	0	–
Stich, lassen	20	0
Geige, erst, spielen	0	–
Beschluss, fassen	0	–
Entscheidung, treffen	58	1,7
pleite, gehen / pleite, sein	6	0

Примечание: поиск по *pleite, gehen / pleite, sein* проводился как два отдельных поиска, которые выдали 2 предложения с *pleite gehen* и 4 предложения с *pleite sein*. В Таблице 1 показан сводный результат.

Заметим также, что поиск заданных фразеологизмов показывает очень низкую погрешность, величина которой, впрочем, может меняться в зависимости от частотности употребления компонентов фразеологизма в языке в целом (ср. 14,8% для *Arbeit leisten*). По нашему мнению, при работе с NLP-библиотеками допустима погрешность в пределах 20%. Тем не менее одной из важнейших задач в этом направлении является стремление к уменьшению значения этого параметра.

Заключение

Поставленная в ходе исследования цель была достигнута – нами были созданы два алгоритма поиска фразеологических единиц, обозначенные как «Фразеология» и «Фразеология 2», которые встраиваются в модель корпусного менеджера с морфологической разметкой по правилам *sraSu*. При этом, во-первых, была приведена краткая «техническая» классификация фразеологизмов, которая учитывает такие критерии, как порядок следования компонентов относительно друг друга и возможность их изменения по грамматическим категориям. Во-вторых, были обстоятельно изучены возможности ручного и специального ручного запросов к таблице предложений. Первый использует синтаксис *SQL*, для второго применяется специальный «промежуточный» язык. Оба типа запроса эффективны для нахождения фразеологизмов с закрепленным порядком следования компонентов. В-третьих, в исходный код корпусного менеджера были добавлены модули «Фразеология» и «Фразеология 2» для поиска фразеологических единиц с «плавающей» позицией компонентов внутри предложения. При этом для повышения скорости работы программы была разработана модель двухступенчатого поиска – сначала по первому заданному компоненту, затем по остальным, в уже полученном после прохождения первой ступени подкорпусе. Наконец, созданные алгоритмы прошли апробацию, по результатам которой мы можем заключить, что они не показывают технических сбоев и достаточно верно выявляют заданные фразеологизмы при условии грамотного составления поискового запроса оператором. Максимальная погрешность составила 14,8%, что может быть объяснено частотностью употребления компонентов в современном немецком языке.

В качестве перспектив исследования обозначим составление банков данных фразеологизмов для алгоритма «Фразеология 2», а также апробацию разработанного программного обеспечения на материале других языков.

Источники | References

1. Бирюков А. А., Терещенко И. А. Перспективы применения технологии обработки естественного языка в рамках цифровой трансформации экономики РФ и построения цифрового общества // Информационное общество. 2023. № 2.
2. Богуславская И. В. Предикатные конструкции с дистантным расположением компонентов: структура и семантика глагольно-именных предикативных сочетаний в немецком языке: автореф. дисс. ... к. филол. н. СПб., 1997.
3. Виноградов В. В. Об основных типах фразеологических единиц в русском языке // Виноградов В. В. Избранные труды. Лексикология и лексикография. М.: Наука, 1977.
4. Голубкова Е. Е., Бухарова А. А. Загадки конструкции XYZ(W): корпусный подход // От слова к дискурсу: взаимодействие форм и (не)предсказуемость смыслов: тез. междунар. науч. конференции (г. Минск, 12-13 мая 2023 г.) / Минский государственный лингвистический университет. Мн.: МГЛУ, 2023.
5. Гусейнова И. А. Актуальные вопросы современной немецкой лексикологии и фразеологии: учеб. пособие для студентов бакалавриата, обучающихся по направлению подготовки 45.03.02 Лингвистика. Казань: Бук, 2023.
6. Давыдова М. А. Фразеологические единицы в теории и практике обучения РКИ на продвинутом этапе: узус и окказиональное употребление // Вестник Московского университета. Серия 19: Лингвистика и межкультурная коммуникация. 2023. № 4.
7. Киселева Л. А. Натурморфные образы в славянской и тюркской эмотивной фразеологии: параллели и контрасты // Вестник Волгоградского государственного университета. Серия 2: Языкознание. 2023. Т. 22. № 1.
8. Кротова Е. Б. Фразеологизмы-конструкции в корпусе: особенности поиска и анализа выдачи (на материале немецкого языка) // Лингвистика и методика преподавания иностранных языков. 2022. № 2 (17). <https://doi.org/10.37892/2218-1393-2022-17-2-76-90>
9. Нефедова Л. А. Гендерно инклюзивная фразеология современного немецкого языка: симметрия фразеологизмов в свете гендера. М.: МПГУ, 2023.
10. Никитина Т. Г. Библиейские образы в устойчивых сравнениях славян: типы межъязыковых соответствий и их словарная репрезентация // Русин. 2023. № 71.
11. Парина И. С. Применение корпусов параллельных текстов для исследования немецкой фразеологии в сопоставительном аспекте // Русская германистика: ежегодник Российского союза германистов (г. Коломна, 28-30 ноября 2019 г.) / Нижегородский государственный лингвистический университет им. Н. А. Добролюбова, Российский союз германистов. Коломна: ФЛИНТА, 2020. Т. XVII.
12. Соколова Т. С., Старикова Г. Н. Фразеологические единицы с названиями времен суток и их производными как реализация номинативно-деривационного потенциала исходных слов // Вестник Томского государственного университета. Филология. 2023. № 82. <https://doi.org/10.17223/19986645/82/8>
13. Титова О. В. Лексика, связанная с жилищем, во фразеологизмах и паремиях удмуртского языка // Вестник угроведения. 2023. Т. 13. № 2. <https://doi.org/10.30624/2220-4156-2023-13-2-293-301>
14. Шанский Н. М. Фразеология современного русского языка. М.: ЛИБРОКОМ, 2012.
15. Юсипова Д. Б. Корпусный подход к исследованию фразеологии: анализ фразеологизмов семантического поля Пространство // Вестник Московского государственного областного университета. Серия: Лингвистика. 2022. № 3-1. <https://doi.org/10.18384/2310-712X-2022-3-38-45>

Информация об авторах | Author information**Горожанов Алексей Иванович**¹, д. филол. н., доц.¹ Московский государственный лингвистический университет**Gorozhanov Alexey Ivanovich**¹, Dr¹ Moscow State Linguistic University¹ a_gorozhanov@mail.ru**Информация о статье | About this article**

Дата поступления рукописи (received): 09.12.2023; опубликовано online (published online): 22.01.2024.

Ключевые слова (keywords): корпусная лингвистика; корпусный менеджер; обработка естественного языка; фразеология; немецкий язык; corpus linguistics; corpus manager; natural language processing; phraseology; German language.