

Мир науки. Социология, филология, культурология <https://sfk-mn.ru>

World of Science. Series: Sociology, Philology, Cultural Studies

2024, Том 15, № 3 / 2024, Vol. 15, Iss. 3 <https://sfk-mn.ru/issue-3-2024.html>

URL статьи: <https://sfk-mn.ru/PDF/19FLSK324.pdf>

5.9.5. Русский язык. Языки народов России (филологические науки)

Ссылка для цитирования этой статьи:

Бакшеева, М. Г. Электронный корпус текстов как основа для словаря сочетаемости глаголов с семантикой пространства / М. Г. Бакшеева, Ю. В. Исламова // Мир науки. Социология, филология, культурология. — 2024. — Т. 15. — № 3. — URL: <https://sfk-mn.ru/PDF/19FLSK324.pdf>

For citation:

Baksheeva M.G., Islamova Yu.V. Electronic corpus of texts as a basis for a dictionary of verb collocations with spatial semantics. *World of Science. Series: Sociology, Philology, Cultural Studies*. 2024;15(3): 19FLSK324. Available at: <https://sfk-mn.ru/PDF/19FLSK324.pdf>. (In Russ., abstract in Eng.)

УДК 8

Бакшеева Марина Геннадьевна

ФГБОУ ВО «Югорский государственный университет», Ханты-Мансийск, Россия

Доцент

Кандидат филологических наук, доцент

E-mail: Baksheeva.Marina@yandex.ru

ORCID: <https://orcid.org/0000-0002-3248-165X>

РИНЦ: https://elibrary.ru/author_profile.asp?id=267672

Исламова Юлия Валерьевна

ФГБОУ ВО «Югорский государственный университет», Ханты-Мансийск, Россия

Доцент

Кандидат филологических наук, доцент

E-mail: islyv@yandex.ru

ORCID: <https://orcid.org/0000-0003-3752-8418>

РИНЦ: https://elibrary.ru/author_profile.asp?id=771153

Электронный корпус текстов как основа для словаря сочетаемости глаголов с семантикой пространства

Аннотация. Корпусная лингвистика в настоящее время является одной из востребованных отраслей языкознания. В данной статье рассказывается об опыте использования электронного корпуса текстов (художественных и нехудожественных, письменных и устных) и ономастических единиц, характеризующих пространственные отношения в русском и угорских языках, в качестве основы для словаря сочетаемости глаголов с пространственной семантикой. Электронный корпус «Пространство и язык», созданный в рамках реализации научного проекта, посвященного анализу пространственных отношений в вербальных традициях народов современной России и Венгрии, доступен каждому человеку, в том числе и неспециалисту, технически слабо подготовленному пользователю, поскольку имеет очень простой интерфейс, к положительным сторонам также следует отнести доступность и легкую пополняемость новыми материалами. Подобный ресурс, выявляющий и систематизирующий языковые средства, выражающие семантику пространства, — важный источник лингвистической информации, помогающий, в частности, в решении проблем лексико-грамматической сочетаемости глаголов пространства. В процессе работы авторским коллективом была предложена схема лексикографического описания сочетаемости глаголов, в которую внесены некоторые коррективы по сравнению с традиционной формой построения дефиниции словарей подобного типа: кроме сочетаемости, в структуру словарной статьи

каждого глагола предлагается ввести элементы словника-толкователя глаголов. В статье приведен пример словарной дефиниции из материалов к словарю сочетаемости глаголов с пространственной семантикой, составленный на базе текстов, размещенных на ресурсе «Пространство и язык». Указанный корпус и опыт описания сочетаемости глагольных единиц, безусловно, будут полезны широкому кругу исследователей для решения разнообразных языковых задач.

Ключевые слова: электронный корпус; текст; глагол; семантика пространства; сочетаемость; словарь; словарная дефиниция; русский язык

Введение (актуальность)

Одним из самых перспективных направлений развития языкознания в настоящее время является корпусная лингвистика. Применение современных технических средств в лингвистических исследованиях актуально, поскольку текстовые базы данных позволяют ученым получать недоступную ранее информацию о тексте в целом и об отдельных его составляющих разных уровней языка [1–5]. Яркий пример совместной работы лингвистов и представителей информационных технологий — Национальный корпус русского языка.¹

Цель статьи — рассказать об опыте использования электронного корпуса текстов (художественных и нехудожественных, письменных и устных) и ономастических единиц, характеризующих пространственные отношения в русском и угорских языках, в качестве основы для словаря сочетаемости глаголов с пространственной семантикой.

О соотношении значения слова и его сочетаемости достаточно подробно написано в работах Д.Н. Шмелева, Ю.Д. Апресяна, В.Г. Гака, П.Н. Денисова, В.В. Морковкина, Г.А. Золотовой, Н.З. Котеловой и др. Сочетаемость определенных форм русского глагола, его функционирование в разных типах текста рассматриваются Г.А. Хайрутдиновой [6] и И.Ю. Хандархзаевой [7]. Однако в значительно меньшей степени разработана проблема взаимодействия значения слова и его сочетаемости в лексикографическом аспекте, на что указывает Архипова Н.Г. [8].

Методы

Исследование осуществлялось в рамках функционально-семантического подхода. Языковой материал анализировался от функции к средствам выражения. Описание категории пространства в функциональном плане является оригинальным и новым, поскольку ранее в лингвистике внимания к функциональной стороне языка удостоивались лишь отдельные понятийные категории и языковые сферы. Поставленные цель и материал исследования обусловили необходимость использования различных лингвистических методов. Проведенное исследование выполнено в рамках компьютерных методов изучения языка. Сбор языкового материала осуществлялся методом сплошной выборки из письменных источников (фольклорных, художественных и учебных текстов, СМИ, грамматик и словарей). Методом структурного моделирования в проекте составлен список семантических конструкций исследуемых языков, выражающих пространственные отношения. В целях исследования содержательной стороны лексики с пространственным значением привлекался метод компонентного анализа значения.

¹ Национальный корпус русского языка. URL: <https://web.ruscorpora.ru> (дата обращения: 15.02.2024).

Результаты

В 2021–2022 гг. авторским коллективом был реализован научный проект, посвященный анализу пространственных отношений в вербальных традициях народов современной России и Венгрии, выявлению и систематизации языковых средств, выражающих семантику пространства, установлению универсального и идиоэтнического в пространственных системах языков различных генетических и типологических общностей.

В числе практических результатов проектной работы — база данных (электронный корпус, включающий более 10 000 текстов разных стилей и жанров, примеров, отражающих функционирование лексики с пространственной семантикой в различных сферах общественно-речевой практики в различные периоды с привлечением фольклорных, художественных и учебных текстов, а также грамматик и словарей).²

Разработанный и наполненный текстами электронный корпус является средой для получения новых научных данных, осмысление которых представляется приоритетным для современного лингвистического описания и абсолютно необходимым в научной деятельности современного исследователя. Электронный корпус текстов — это важный источник лингвистической информации, который поможет решать разнообразные проблемы лексико-грамматической сочетаемости, а также трудности, связанные с выбором оптимальной грамматической конструкции.

Тексты разных по структуре языков, стилей и жанров доступны для дальнейшего изучения широкому кругу исследователей. Созданный ресурс доступен каждому человеку, в том числе и неспециалисту, исследования с его помощью может проводить любой, в том числе и технически слабо подготовленный пользователь. Корпус имеет очень простой интерфейс, к положительным сторонам также следует отнести доступность и легкую пополняемость новыми материалами. При этом его роль важна для лингвистики прикладной, решающей задачи автоматической обработки текстов на естественных языках. Кроме того, форма подачи текстов, включающих языковые единицы, имеющие отношение к тематической группе «Пространство», приведена к единообразию.

Новым и перспективным видится включение в корпус текстов из разноструктурных языков — русского и одного из угорских — хантыйского. Этот факт позволяет создать практически оптимальные условия для проведения исследования проблем передачи различных языковых значений в сопоставляемых языках, открывает новые возможности и, безусловно, повышает уровень объективности сопоставительных лингвистических исследований.

В числе отобранных лексических единиц с семантикой пространства — глаголы (однозначные и многозначные), всесторонний анализ которых в рамках контекстов, включенных в электронный корпус, позволил выявить способы передачи пространственных отношений в разноструктурных языках (русском и венгерском), определить их специфику, функциональную нагрузку и семантическую емкость в коммуникативной деятельности, в вербальных традициях народов современной России и Венгрии, что позволило прийти к выводу: локальность выступает как универсальный организующий центр речи, т. к. в любом высказывании фиксируется пространственная сущность мира.

В процессе работы была предложена схема лексикографического описания сочетаемости глаголов, в которую внесены некоторые коррективы по сравнению с традиционной формой построения дефиниции словарей подобного типа, она может быть применена для создания других словарей сочетаемости. Кроме сочетаемости, в структуру словарной статьи каждого глагола предлагается ввести элементы словника-толкователя

² Пространство и язык. URL: <http://языкипространство.рф/> (дата обращения: 15.02.2024).

глаголов, выражающих пространственные отношения в русском и венгерском языках. Созданные дефиниции включены в «Материалы для словаря сочетаемости глаголов, выражающих пространственные отношения в русском и венгерском языках».

Пример словарной дефиниции многозначного глагола (русский язык):

Располагаться, располагаюсь, располагаешься, располагаются несовершенный вид.

1. Занимать место; размещаться. «Москва предложила компании два здания для офиса, и венесуэльская сторона уже выбрала, где будет располагаться представительство компании».³

Что (в качестве субъекта — нарицательные и собственные имена, слова и словосочетания в форме ед. и мн. числа, входящие в состав тематических групп: «Помещение», «Территория», «Место», «Организация», «Объект», «Абстрактный или конкретный): школа, ресторан, клуб, деревня, дом, квартира, комната, кабинет, казино, ферма, аптека, салон связи, «Мастерстрой», АОН, мысли, языки, видеокамеры, дороги, игры, плиты, прилавки, парки, клиники, станции метро, избирательные участки, клубы, агентства, места для рыбалки.

Инфинитив может использоваться с вспомогательными словами «будет, станет, может, должен» в составном глагольном сказуемом.

Располагаться неодушевленные предметы могут внутри какого-то объекта, на территории, в непосредственной близости от кого/чего-либо, определенным образом, до определенного времени. Локатив в предложении обычно соответствует обстоятельству места наиболее частотно: располагаться где (предлог «в» с предложным падежом) — в центре, в здании, в районе; (предлог «на» с предложным падежом) — на окраине, на территории, на полу, на берегу, на карте, на столе; (предлоги «около»/«возле»/«у» с родительным падежом) — около/возле/у магазина, около/возле/у библиотеки, около/возле/у [станции] метро, около/возле/у школы, около/возле/у здания; (предлог «под» с творительным падежом) — под Москвой, под горой; (предлог «за» с творительным падежом) — за городом, за селом, за углом, за рекой, за лесом; (предлог «напротив» с родительным падежом) — напротив дома, напротив школы, напротив театра; (с наречием) — здесь, там, близко, неподалеку. Располагаться когда, как долго — сейчас, скоро, недолго, постоянно, в будущем, до осени, до следующего года.

2. Помещаться, усаживаться, укладываться где-либо. «Посоветовавшись, командиры спецслужб отдали распоряжение сойти с дороги и располагаться близ населенного пункта, но в таком месте, где огонь противника не мог нанести урона» [9].

Кто (в качестве субъекта — нарицательные и собственные имена существительные тематической группы «Люди в широком смысле (относительно половой принадлежности, рода деятельности, социального статуса) и их объединения/группы (по интересам, профессии, роду деятельности, полу, месту проживания, по отношению к своим детям)», «Животные» в форме ед. и мн. ч., заменяющие их местоимения, словосочетания: Алеша, Олег, Скрипицын, Томанцев, Марлен Михайлович, человек, женщина, учительница, обыватель, люди, бойцы, торговцы, лейтенант, матери, соседи, военные, миротворцы, рабочие, чудовище, паразиты, кто-то, все, они, он, она, я, ты, никто из нас, любители зимней рыбалки.

Инфинитив может использоваться с вспомогательными словами «будет, станет, может, должен, предложил» в составном глагольном сказуемом.

Располагаться одушевленные предметы могут в основном: внутри какого-то объекта; на территории; в непосредственной близости от кого/чего-либо, определенным образом; в большом количестве; до определенного времени; в период времени. Локатив в предложении

³ Россия попросила PDVSA зарегистрировать юрлицо для переноса офиса в Москву. URL: <https://www.kommersa.ru/doc/4019259?ysclid=m1olry2dpo502144653> (дата обращения: 25.09.2024).

обычно соответствует обстоятельству места, наиболее частотно: располагаться где (предлог «в» с предложным падежом) — в классе, в здании, в поезде, в центре, в любом направлении пространства, в судьбе; (предлог «на» с предложным падежом) — на другом конце стола, на улице, на ельнике, на полях; (предлоги «около»/«возле»/«у» с родительным падежом): около/возле/у стола, около/возле/у магазина, около/возле/у школы, около/возле/у здания, около/возле/у населенного пункта; (предлог «под» с творительным падежом) — под святыми образами; под горой; (предлог «за» с творительным падежом) — за шкафом, за городом; за горой, за углом, за мостом, за рекой; (предлог «напротив» с родительным падежом) — напротив дома, напротив школы, напротив леса; (с наречием) — здесь, там, тут, близко, далеко. Располагаться каким образом: группами, кто как умел. Располагаться когда, как долго: постоянно, во время боя.

3. Устраиваться, обосновываться где-либо на жительство, на отдых. «— Мой Трубников, в какой спальне мне можно располагаться?» [10].

Кто (в качестве субъекта — слова тематических групп «Люди в широком смысле, их объединения/группы», «Животные»; нарицательные и собственные имена, существительные и местоимения в форме ед. и мн. ч): Чжан Бао и Чан Лин, женщина, мужичок, рабочие, дежурный, героиня, войска, охотники, партизаны, я, мы, лошади.

Одушевленные предметы могут располагаться, обосновываться в пространстве, в определенном месте, на определенной территории на ночлег, отдых, особым образом. Локатив в предложении обычно соответствует обстоятельству места, наиболее частотно: располагаться где (предлог «в» с предложным падежом) — в городе, в селе, в спальне, в кресле, в хоромах, в гостинице, в отеле, в гостинице, в горах; (предлог «на» с предложным падежом) — на окраине, на ночлег, на нарах, на отдых, на верхней полке в поезде, на полу, на берегу; (предлоги «около»/«возле»/«у» с родительным падежом) — около/возле/у озера, около/возле/у реки, около/возле/у горы, около/возле/у водопада, около/возле/у печи; (предлог «под» с творительным падежом) — под пологом, покровом ночи, под солнцем, под Питером, под горой; (предлог «за» с творительным падежом) — за лесом, за городом, за стеной; за рекой; за границей, за морем. Располагаться каким образом: свободно, поудобнее, с комфортом, похозяйски, никого не стесняя, как у себя дома. Располагаться как долго: ненадолго, на ночь, до утра, до рассвета.

Этикетное употребление в императиве: Располагайтесь поудобнее!

Нерегулярная сочетаемость:

1. С отрицательной частицей НЕ и устаревшей формой существительного СТАТЬ в значении «необходимость, надобность, нужда» — «Иноку не стать располагаться в доме столь великолепном».
2. В значении «располагать кем-то по своему усмотрению, иметь в своем распоряжении» — «распоряжаться», «делать с ним то, что считает нужным» — «Родитель своим детищем может располагаться как ему угодно».

Синонимы: размещаться, находиться, помещаться, тесниться, ютиться; частично — устраиваться.

Фразеологизмы: нет.

Обсуждение

С помощью описанного корпуса можно очень быстро найти много примеров на интересующее исследователя явление, в первую очередь — значительное количество текстов

(художественных и нехудожественных, письменных и устных) и ономастических единиц, характеризующих пространственные отношения в русском и угорских языках. Включение в корпус культурно значимых текстов на хантыйском языке представляет собою источник, обладающий собственной ценностью. Практическая значимость исследования состоит в том, что полученные результаты — тексты разных по структуре языков, стилей и жанров — доступны для дальнейшего изучения широкому кругу исследователей.

Проблема сочетаемости слов, то есть возможность использования слова в предложении или высказывании, регулируемая системой и нормой языка и выражающаяся во взаимодействии явлений разных языковых уровней, является одной из наиболее актуальных проблем современной лингвистики. Интерес к расположению предметов в вертикальной и горизонтальной плоскостях, удалённости одного предмета от другого или от говорящего (наблюдателя), статике или динамике предмета, направленности движения, валентности глаголов местонахождения с семой статичности вызван спецификой средств выражения локативности, функционирующих в разноструктурных языках — русском и венгерском.

Составленный словарь предназначен для широкого круга читателей, т. к. является первым опытом лексикографического описания сочетаемости глаголов пространства в русском и венгерском языках. Собранные и описанные материалы, безусловно, могут способствовать как дальнейшим научным изысканиям в области сочетаемости единиц языка, так и служить источником развития речи студентов в процессе изучения русского/венгерского языка как иностранного. Разработанная методика описания может использоваться в изучении категории пространства в других языках. Выявленный языковой материал может быть полезен преподавателям в качестве источника развития речи в ходе изучения русского, хантыйского и венгерского языков как иностранных.

ЛИТЕРАТУРА

1. Клышинский Э.С., Кочеткова Н.А., Литвинов М.И., Максимов В.Ю. Автоматическое формирование базы сочетаемости слов на основе очень большого корпуса текстов // Компьютерная лингвистика и интеллектуальные технологии. М.: Изд-во РГГУ, 2010. С. 181–186.
2. Осипов Г.А. О перспективах корпусных исследований и практике их применения в лингвистике // Вестник Адыгейского государственного университета. Сер.: Филология и искусствоведение. Майкоп, 2023. Вып. 1. С. 118–124.
3. Птущенко Е.Б. Перспективные информационные технологии как инструмент познания в лингвистике // Вестник Адыгейского государственного университета. Сер.: Филология и искусствоведение. Майкоп, 2017. Вып. 2. С. 102–108.
4. Рыков В.В. Корпус текстов как реализация объектно-ориентированной парадигмы // Труды международного семинара «Диалог-2002». URL: <https://www.dialog-21.ru/digest/2002/articles/rykov/> (дата обращения: 15.02.2024).
5. Хурум Р.Ю. Использование информационно-коммуникационных технологий в лингвистике как фактор формирования информационных компетенции // Вестник Адыгейского государственного университета. Сер.: Филология и искусствоведение. Майкоп, 2017. Вып. 4. С. 133–138.
6. Хайрутдинова Г.А. Функционирование видо-временных форм глаголов в повествовательных и описательных текстах // Русский глагол: Лексико-грамматические исследования. — Казань: Изд-во КГПИ, 1986. С. 91–99.

7. Хандархаева И.Ю. О сочетаемости глаголов совершенного вида с перфектным и аористическим значением с другими видовременными глагольными формами // Вестник бурятского государственного университета. Улан-Уде, 2014. Вып. 10(4). С. 69–74.
8. Архипова Н.Г. Проблемы русской нормативной лексикографии. Благовещенск: Амурский государственный университет, 2002. 70 с.
9. Кожевников В. М. Щит и меч: роман: в 2 кн. / Вадим Кожевников. — Москва: ЭКСМО-Пресс, 1998. Кн. 1. — 1998. — 590 с.
10. Прашкевич Г., Богдан А. Человек «Ч». М.: Вагриус, 2001. 301 с.

Baksheeva Marina Gennad'evna

Yugra State University, Khanty-Mansiysk, Russia

E-mail: Baksheeva.Marina@yandex.ru

ORCID: <https://orcid.org/0000-0002-3248-165X>

RSCI: https://elibrary.ru/author_profile.asp?id=267672

Islamova Yulia Valer'evna

Yugra State University, Khanty-Mansiysk, Russia

E-mail: islyv@yandex.ru

ORCID: <https://orcid.org/0000-0003-3752-8418>

RSCI: https://elibrary.ru/author_profile.asp?id=771153

Electronic corpus of texts as a basis for a dictionary of verb collocations with spatial semantics

Abstract. Corpus linguistics is currently one of the most popular branches of linguistics. This article describes the experience of using an electronic corpus of texts (fiction and non-fiction, written and oral) and onomastic units characterizing spatial relations in Russian and Ugric languages as a basis for a dictionary of verb compatibility with spatial semantics. The electronic corpus «Space and Language», created within the framework of the implementation of a scientific project dedicated to the analysis of spatial relations in the verbal traditions of the peoples of modern Russia and Hungary, is accessible to everyone, including a non-specialist, a technically poorly trained user, since it has a very simple interface; the positive aspects also include accessibility and easy replenishment with new materials. Such a resource, which identifies and systematizes linguistic means expressing the semantics of space, is an important source of linguistic information, helping, in particular, in solving problems of lexical and grammatical compatibility of verbs of space. In the course of the work, the authors proposed a scheme for the lexicographic description of verb compatibility, which included some adjustments compared to the traditional form of constructing definitions for dictionaries of this type: in addition to compatibility, it is proposed to introduce elements of a dictionary-interpreter of verbs into the structure of the dictionary entry for each verb. The article provides an example of a dictionary definition from materials for the dictionary of compatibility of verbs with spatial semantics, compiled on the basis of texts posted on the resource «Space and Language». The specified corpus and experience of describing the compatibility of verbal units will certainly be useful to a wide range of researchers for solving various language problems.

Keywords: electronic corpus; text; verb; semantics of space; combinability; dictionary; dictionary definition; Russian language