

оригинальная статья

Возможности Национального корпуса русского языка для определения закономерностей функционирования русских причастий

Радченко Анастасия Игоревна

Кемеровский государственный университет, Кемерово, Россия
radchana@yandex.ru

Поступила в редакцию 06.03.2023. Принята после рецензирования 31.03.2023. Принята в печать 31.03.2023.

Аннотация: Интернет рассматривается как источник количественной информации для квантитативно-морфологических исследований. Цель – выявить возможности использования количественных показателей употребления форм причастий, полученных при помощи сервиса национальных корпусов, а именно Национального корпуса русского языка, для соотношения частотности и особенностей функционирования причастий с учетом поисковых опций, предоставляемых выбранным интернет-ресурсом. Материалом послужили извлеченные методом сплошной выборки из корпуса и содержащие разные типы причастий контексты из произведений русских писателей-прозаиков. Для анализа материала использованы квантитативно-морфологический анализ, сравнительно-сопоставительный метод, описательно-аналитический метод с обобщением результатов. Сделан вывод, что использование интернет-поиска как источника лингвистического материала подходит для выявления частотности определенной формы причастий и может свидетельствовать о ее устоявшемся употреблении в речи. Основная лингвистическая ценность заключается в том, что частотность употребления причастий показывает вхождения формы причастия и ее функционирование в речи. Намечены предпосылки выявления лингвоперсоналогических особенностей употребления причастий в художественных текстах на основании частотности их употребления.

Ключевые слова: квантитативная морфология, количественные методы, Национальный корпус русского языка, причастие, художественный стиль, частотность

Цитирование: Радченко А. И. Возможности Национального корпуса русского языка для определения закономерностей функционирования русских причастий. *Виртуальная коммуникация и социальные сети*. 2023. Т. 2. № 2. С. 71–75. <https://doi.org/10.21603/2782-4799-2023-2-2-71-75>

full article

Functional Patterns of Russian Participles: National Corpus of the Russian Language

Anastasia I. Radchenko

Kemerovo State University, Kemerovo, Russia
radchana@yandex.ru

Received 6 Mar 2023. Accepted after peer review 31 Mar 2023. Accepted for publication 31 Mar 2023.

Abstract: The Internet is a source of quantitative information for quantitative morphological research. This article introduces the possibilities for using the National Corpus of the Russian Language to define the quantitative indicators of Russian participles. This approach made it possible to correlate the frequency and patterns of participles based on the search options provided by the Corpus. The research featured participial contexts extracted by continuous sampling from Russian fiction. The methodology included the quantitative-morphological analysis, comparative-comparative method, and descriptive-analytical method. Internet search proved to be a reliable source of linguistic material suitable for identifying the usage frequency of a certain participial form and its speech patterns. The method may be used to profile participial patterns in fiction texts.

Keywords: quantitative morphology, quantitative methods, National Corpus of the Russian language, participle, artistic style, frequency

Citation: Radchenko A. I. Functional Patterns of Russian Participles: National Corpus of the Russian Language. *Virtual Communication and Social Networks*, 2023, 2(2): 71–75. (In Russ.) <https://doi.org/10.21603/2782-4799-2023-2-2-71-75>

Введение

Использование количественного метода лингвистического исследования позволяет освещать глубинные характеристики текста и говорить о его содержательных специфических особенностях, таких как автор и стиль. Данный тезис подтверждается рядом количественных исследований на разных уровнях языка: фонетическом [Булгакова и др. 2015], графическом [Петрова, Соловьева 2016], лексическом [Ляшевская, Шаров 2009], грамматическом [Герд и др. 1974] и др.

Сегодня механизация вычислительных процессов значительно упростила статистический расчет: на любом информационном носителе оцифрованный текст предоставляет возможность получить количественную информацию о его содержимом. Например, кроме вводных данных на веб-странице или в текстовом файле с помощью комбинации клавиш (Ctrl + F) можно осуществить поиск лексем, сочетаний букв, словосочетаний. Данный механизм считает количество запрошенных данных, а также маркирует размещение данных в массиве текста. Существует множество интернет-сервисов, ускоряющих обработку массивов текстовой информации. Изучение вопроса о применении поисковых систем Интернета в лингвистических целях было освещено в ряде публикаций Н. Д. Голева [Голев 2010; 2011; 2015].

В данном исследовании рассмотрено получение количественной информации путем: 1) запроса в поисковых системах веб-браузера; 2) в базе национальных корпусов. Эти способы получения лингвистической информации отражают реальное функционирование языковых единиц в текстах различных типов.

Цель – выявить возможности использования количественных показателей употребления форм причастий, полученных при помощи интернет-ресурсов, в частности Национального корпуса русского языка¹ (НКРЯ), для соотношения частотности и особенности функционирования причастий. Демонстрируется возможность использования интернет-ресурсов и предоставляемых ими количественных показателей употребления причастий в произведениях русской художественной литературы для решений лингвистических задач.

Методы и материалы

Для сбора материала применен метод сплошной выборки (отбор причастий всех типов для определения частотности их употребления в текстах). Методы работы с материалом: количественно-морфологический анализ, сравнительно-сопоставительный метод (использован для выявления частоты показателей использования причастий в произведениях русских прозаиков), описательно-аналитический метод с обобщением результатов.

Материалом стали извлеченные из НКРЯ и содержащие разные типы причастий контексты из произведений русских писателей-прозаиков: Чудесное око, Продавец воздуха, Человек-амфибия, Остров погибших кораблей, Голова профессора Доуэля (А. Р. Беляев); Мертвые души (Н. В. Гоголь); Обрыв, Обломов, Обыкновенная история (И. А. Гончаров); Братья Карамазовы, Подросток, Бесы, Идиот, Игрок, Преступление и наказание, Нечочка Незванова, Бедные люди (Ф. М. Достоевский); Доктор Живаго (Б. Л. Пастернак); Двенадцать стульев, Золотой теленок (Е. Петров и И. Ильф); Пошехонская старина. Житие Никанора Затрапезного, пошехонского дворянина, Современная идиллия, Господа Головлевы, История одного города, Губернские очерки (М. Е. Салтыков-Щедрин); Петр Первый, Хождение по мукам, Аэлита (Закат Марса), Гиперболоид инженера Гарина, Похождения Невзорова, или Ибикус (А. Н. Толстой); Воскресение, Анна Каренина, Война и мир (Л. Н. Толстой); Новь, Дым, Отцы и дети, Дворянское гнездо, Накануне, Рудин (И. С. Тургенев); Пролог, Что делать? (Н. Г. Чернышевский).

Результаты

Данная работа является решением сопутствующей задачи к исследованию причастий как функционально-речевых единиц [Радченко 2022]. Процедура поиска, осуществляемая через интернет-браузеры, имеет значение при изучении употребления форм причастий в разнообразных текстах, представленных в Сети. Количество, предоставляемое поисковой системой, может говорить (1) об освоенности в речи данной формы причастия, (2) о функциональных рамках его употребления. Такие гипотезы

¹ Национальный корпус русского языка. URL: <https://ruscorpora.ru> (дата обращения: 10.12.2022).

выдвинуты на основании контрастных количественных результатов употребления причастий, полученных в ходе запросов в поисковой системе Google, например на следующие формы:

читавший – 151000,
читающий – 1400000,
читанный – 50800,
читаемый – 890000.

Однако материал, полученный в результате интернет-поиска, не всегда предоставляет сведения по окказиональным причастиям. По запросу *недосозданы* предоставляется результат по прилагательному *недосозданный* – 1030 употреблений. Но данный пример не исключает результатов частотности употребления окказиональных причастий, например, *вымечтанный*, чей результат 3930 употреблений. Данные показывают, что такие причастия не зафиксированы словарями русского языка, но активно употребляются в речи.

Окказиональные причастия часто остаются не отмеченными и в НКРЯ. Однако на современном этапе количественная морфология активно изучается на базе языковых корпусов. Корпусы, являясь продуктом деятельности лингвистов, представляют собой статистические данные о языке факте. Корпусная лингвистика позволяет рассчитать грамматическую форму причастия в текстах нужного стиля, нужного автора и т.д. Также основанием для выбора НКРЯ как источника информации служит наличие опции снятия грамматической омонимии в тексте, что позволяет при возможных неточностях получить максимально объективные результаты.

В своем исследовании мы опираемся на данные именно НКРЯ (табл.). Для вычисления коэффициента частотности общее количество причастий во всех произведениях писателя (колонка 4 таблицы) делилось на количество слов во всех проанализированных произведениях автора (колонка 3 таблицы).

Разные количественные данные коэффициентов употребления причастий позволяют выдвинуть гипотезу о возможности идентификации идиостиля автора. Больше всего причастий использует Б. Л. Пастернак, меньше всего – И. А. Гончаров. Лидирующую позицию Б. Л. Пастернака в употреблении причастий можно объяснить двумя факторами:

- 1) высокая интертекстуальность произведения (много отсылок к Евангелию, содержащему большое количество старославянизмов – генетических «предков» причастия);
- 2) Б. Л. Пастернак – не только прозаик, но и поэт, а поэтической речи в большей степени свойственно наличие в ней причастий, часто выполняющих функцию определения.

Произведения И. А. Гончарова, напротив, отличаются реалистической направленностью. Несмотря на то что И. А. Гончаров был менее религиозен, чем, например, Ф. М. Достоевский или Л. Н. Толстой, в его произведениях присутствуют христианские мотивы. Однако в отличие от упомянутых писателей И. А. Гончаров не насыщал романы дидактизмом, поскольку не ставил перед собой задачу наставления читателей. Мы разделяем позицию В. И. Мельник: «Его [Гончарова] тип писательского утверждения христианства был иным, чем у Гоголя»

Табл. Частотность употребления причастий русскими писателями-прозаиками
Tab. Participles in Russian fiction: usage frequency

| Место в рейтинге | Автор | Общее количество слов в произведениях | Количество причастий | Коэффициент частотности |
|------------------|-----------------------|---------------------------------------|----------------------|-------------------------|
| 1 | Пастернак Б. Л. | 149162 | 4466 | 0,0299 |
| 2 | Толстой А. Н. | 640386 | 15334 | 0,0239 |
| 3 | Толстой Л. Н. | 858176 | 18958 | 0,0220 |
| 4 | Петров Е., Ильф И. | 185996 | 4059 | 0,0218 |
| 5 | Беляев А. Р. | 191111 | 3904 | 0,0204 |
| 6 | Гоголь Н. В. | 115143 | 2207 | 0,0191 |
| 7 | Салтыков-Щедрин М. Е. | 538651 | 7971 | 0,0147 |
| 8 | Достоевский Ф. М. | 1212176 | 15538 | 0,0128 |
| 9 | Тургенев И. С. | 306326 | 3715 | 0,0121 |
| 10 | Чернышевский Н. Г. | 279262 | 3060 | 0,0109 |
| 11 | Гончаров И. А. | 467174 | 4951 | 0,0105 |

или Достоевского. У Гончарова не было религиозного пафоса Достоевского и Гоголя, но глубочайшая евангельская вера – несомненно, была. <...> Евангельские установки у Гончарова... залегают в смысловых пластах его произведений» [Мельник 2008: 245].

Намечены предпосылки выявления лингвоперсоналогических особенностей употребления причастий в художественных текстах на основании частотности употребления причастий. В перспективе исследования предполагается разработка типологии языковых личностей по частотности использования видов причастий и специфике их функционирования.

Заключение

Использование интернет-поиска как источника лингвистического материала подходит для выявления частотности определенной формы причастий и может свидетельствовать о ее устоявшемся употреблении в речи. Однако простой поиск по отдельным формам причастий не дает абсолютно точных количественных данных – результаты этих данных выше, чем реальное число употреблений причастий, т.к. веб-поиск не может снять проблему полисемии и омонимии, проблему многократных повторений, цитирований, ссылок и словарей, размещенных на сайтах. Несмотря на данный фактор, сама

возможность поиска в интернет-браузерах форм причастий представляется возможной и информативной, т.к. между формами причастий обнаруживаются разные количественные данные употреблений и наблюдается закономерность этих употреблений. Соответственно, результаты сбора языкового материала при помощи поисковой системы Google, предположительно, и других браузеров тоже имеют лингвистическую ценность. И основная ценность состоит в том, что частотность употребления причастий показывает вхождения формы причастия и ее функционирование в речи.

Однако для данного исследования использование НКРЯ является лучшим источником лингвистической информации из-за наличия ряда технических факторов обработки информации, а также опциональных возможностей поисковой системы, что увеличивает научную достоверность результатов и делает возможным задачу использования интернет-ресурса в лингвистических целях.

Конфликт интересов: Автор заявил об отсутствии потенциальных конфликтов интересов в отношении исследования, авторства и / или публикации данной статьи.

Conflict of interests: The author declared no potential conflicts of interests regarding the research, authorship, and / or publication of this article.

Литературы / References

- Булгакова Е. В., Шолохов А. В., Томашенко Н. А. Метод идентификации дикторов на основе сравнения статистик длительностей фонем. *Научно-технический вестник информационных технологий, механики и оптики*. 2015. Т. 15. № 1. С. 70–77. [Bulgakova E. V., Sholokhov A. V., Tomashenko N. A. Speakers' identification method based on comparison of phoneme lengths statistics. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2015, 15(1): 70–77. (In Russ.)] <https://elibrary.ru/tjyjsx>
- Герд А. С., Капорулина Л. В., Колесов В. В., Рускова М. П., Черепанова О. А. Именное склонение в славянских языках XI–XIV вв. Л.: ЛГУ, 1974. 236 с. [Gerd A. S., Kaporulina L. V., Kolesov V. V., Ruskova M. P., Cherepanova O. A. *Nominal declension in Slavic languages of the XI–XIV centuries*. Leningrad: LSU, 1974, 236. (In Russ.)]
- Голев Н. Д. Лексическая реализация как функциональная характеристика словообразовательной системы русского языка и количественные параметры ее описания. *Осьм' десятъ: сборник научных статей к 80-летию И. С. Улуханова*, отв. ред. М. А. Малыгина. М.: Азбуковник, 2015. С. 183–193. [Golev N. D. Lexical realization as a functional characteristic of the word-formation system of the Russian language and quantitative parameters of its description. *Osm' desyat': collection of scientific articles for the 80th anniversary of I. S. Ulukhanov*, ed. Malygina M. A. Moscow: Azbukovnik, 2015, 183–193. (In Russ.)] <https://elibrary.ru/vnnufr>
- Голев Н. Д. Лексическое функционирование словообразования и лексико-словообразовательная системность русского языка в свете статистики Интернета. *Вестник Томского государственного университета. Филология*. 2011. № 1. С. 22–31. [Golev N. D. Lexical functioning of word-building and the lexical-word-building system of the Russian language in the light of Internet statistics. *Vestnik Tomskogo Gosudarstvennogo Universiteta. Filologiya*, 2011, (1): 22–31. (In Russ.)] <https://elibrary.ru/negsif>

- Голев Н. Д. Поисквые системы Интернета как лингвистический источник (на примере решения некоторых теоретических и прикладных вопросов русского словообразования). *Новые явления в славянском словообразовании: система и функционирование*: XI Междунар. науч. конф. (Москва, 24–26 марта 2009 г.) М.: МАКС Пресс, 2010. С. 197–212. [Golev N. D. Internet search engines as a linguistic source in solving some theoretical and applied issues of Russian word formation. *New phenomena in Slavic word formation: system and functioning*: XI Intern. Sci. Conf., Moscow, 24–26 Mar 2009. Moscow: MAKS Press, 2010, 197–212. (In Russ.)] <https://elibrary.ru/seaaql>
- Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка на материалах Национального корпуса русского языка. М.: Азбуковник, 2009. 1087 с. [Lyashevskaya O. N., Sharov S. A. *Frequency dictionary of the modern Russian language based on the materials of the National Corpus of the Russian language*. Moscow: Azbukovnik, 2009, 1087. (In Russ.)] <https://elibrary.ru/rtaoxb>
- Мельник В. И. О последних новеллах И. А. Гончарова (проблема религиозного смысла). *И. А. Гончаров*: IV Междунар. науч. конф., посв. 195-летию со дня рождения И. А. Гончарова. (Ульяновск, 13–17 июня 2007 г.) Ульяновск: Ника-дизайн, 2008. С. 236–246. [Melnik V. I. The last novels by I. A. Goncharov: religious meaning. *I. A. Goncharov*: Proc. IV Intern. Sci. Conf. dedicated to the 195th anniversary of I. A. Goncharov, Ulyanovsk, 13–17 Jun 2007. Ulyanovsk: Nika-desain, 2008, 236–246. (In Russ.)]
- Петрова А. Л., Соловьева А. А. Исследование распределений букв, передающих гласные, в итальянских, русских, французских и немецких песенных текстах методами математической статистики. *Математика и междисциплинарные исследования – 2016*: Всерос. науч.-практ. конф. молодых ученых с Междунар. уч. (Пермь, 16–19 мая 2016 г.) Пермь: ПГНИУ, 2016. С. 261–265. [Petrova A. L., Solovieva A. A. Analysis of vowel letter distribution in musical lyrics via use of mathematical statistics methods (a case study of Italian, Russian, French and German songs). *Mathematics and Interdisciplinary Research – 2016*: Proc. All-Russian Sci.-Prac. Conf. of young scientists with Intern. participation, Perm, 16–19 May 2016. Perm: PSU, 2016, 261–265. (In Russ.)] <https://elibrary.ru/wxuodr>
- Радченко А. И. Причастия как функционально-речевые единицы (на материале произведений русской прозы). *Культура и текст*. 2022. № 3. С. 167–177. [Radchenko A. I. Participles as functional speech units (based on the material of works of Russian prose). *Kultura i tekst*, 2022, (3): 167–177. (In Russ.)] <https://doi.org/10.37386/2305-4077-2022-3-167-177>