

АКТУАЛЬНЫЕ ИССЛЕДОВАНИЯ ЯЗЫКА В СОВРЕМЕННОМ МИРЕ

Научная статья
УДК 81'33



Идентификация средств категории отрицания в сбалансированном лингвистическом корпусе

А. И. Горожанов

Московский государственный лингвистический университет, Москва, Россия
a.gorozhanov@linguanet.ru

Аннотация. Цель исследования – разработать базовую технологию идентификации средств категории отрицания в рамках работы программного комплекса «Генератор сбалансированного лингвистического корпуса и корпусный менеджер». Применяются методы анализа, тестирования, моделирования, профессионально ориентированного программирования и эксперимента. В качестве лингвистического материала выступает корпус актуальных новостных текстов CNN объемом 1 024 583 токена. В результате исследования уточнены возможности текущей версии программного комплекса, а также написано расширение, позволяющее находить лексические единицы, находящиеся в прямом подчинении от заданной леммы.

Ключевые слова: прикладная лингвистика, корпусная лингвистика, корпусный менеджер, сбалансированный лингвистический корпус, отрицание, тексты СМИ, обработка естественного языка

Для цитирования: Горожанов А. И. Идентификация средств категории отрицания в сбалансированном лингвистическом корпусе // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2025. Вып. 2 (896). С. 75–80.

Original article

Identification of Means of the Category of Negation in a Balanced Linguistic Corpus

Alexey I. Gorozhanov

Moscow State Linguistic University, Moscow, Russia
a.gorozhanov@linguanet.ru

Abstract. The aim of the research is to develop a basic technology for identifying means of the category of negation within the framework of the software package “Balanced linguistic corpus generator and corpus manager”. The methods of analysis, testing, modeling, professionally oriented programming and experiment are applied. The language material is the corpus of recent CNN news texts with a volume of 1,024,583 tokens. As a result, the capabilities of the current version of the software package were clarified, and an extension was written that allows one to find lexical units that are directly subordinate to a given lemma.

Keywords: applied linguistics, corpus linguistics, corpus manager, balanced linguistic corpus, negation, media texts, natural language processing

For citation: Gorozhanov, A. I. (2025). Identification of means of the category of negation in a balanced linguistic corpus. Vestnik of Moscow State Linguistic University. Humanities, 2(896), 75–80. (In Russ.)

ВВЕДЕНИЕ

Обработке естественного языка посвящены в последнее время не только исследования специалистов из технических областей знания (информатика, автоматика и вычислительная техника, математика и др.), но также и лингвистов. При этом наблюдается достаточно четкая дихотомия, согласно которой можно выделить: а) исследования, опирающиеся преимущественно на результаты работы готовых решений [Гращенков, Студеникина, Паско, 2024; Лагутина К., Бойчук, Лагутина Н., 2023; Солнышкина, Макнамара, Замалетдинов, 2022]; б) исследования, которые вносят некоторый вклад в разработку программного обеспечения лингвистического назначения – с использованием свободно распространяемых библиотек обработки естественного языка и экспериментальных отечественных разработок [Гагарин, 2024; Сафина, 2024; Красикова, 2024; Степанова, 2024].

В настоящей работе мы рассматриваем возможности авторского программного комплекса «Генератор сбалансированного лингвистического корпуса и корпусный менеджер» [Горожанов, 2024а] для решения проблемы учета средств категории отрицания при анализе употребления заданных лексических единиц в сбалансированном лингвистическом корпусе, собранном посредством указанного программного обеспечения.

Материалом исследования служит сбалансированный лингвистический корпус новостных текстов CNN объемом 1 024 583 токена или 38 621 предложение, собранных с декабря 2024 по февраль 2025 года.

Отрицание «представляет собой одно из центральных явлений в рамках актуального направления обработки естественного языка (англ. *Natural Language Processing* или NLP)» [Глушак, 2023, с. 3287]. Сложность проблемы усугубляется также необходимостью применения различных программных алгоритмов для текстов на разных языках, поскольку грамматическая категория отрицания реализуется в каждом отдельном случае своим набором языковых средств, не говоря уже о различном отношении к двойному отрицанию даже в пределах (формально) одного языка. Ср. для вариантов английского языка: «I *don't* know *nothing* about it» vs. «I *don't* know *anything* about it»; или для русского и немецкого языков: «Я *никогда никому ничего* об этом не рассказываю» vs. «Ich erzähle *nie* jemandem etwas davon».

Подчеркнем, что заявленная нами тема тесно связана с проблемой автоматического определения тональности текста, которую мы рассматриваем в ряде других работ [Горожанов, 2024б; Горожанов, 2023].

Здесь под средствами категории отрицания мы будем понимать эксплицитные конститuenty, такие как: отрицательная частица *not*, наречие *never*, местоимения *nobody*, *nothing*, приставки *un-*, *in-* и суффиксы *-less*, *-free*, т. е. фактически определенные лексемы и их части, которые могут быть обнаружены в корпусе с морфологической разметкой¹. Безусловно, такой подход не принимает во внимание имплицитные средства отрицания, учет которых неизбежно бы привел к необходимости оценки тональности контекстов, коннотативных значений отдельных языковых единиц, субъективного фактора восприятия содержания текста и, возможно, к необходимости анализа средств других категорий, например, модальности (см. [Сытько, 2023]). Тем не менее мы считаем, что опора на эксплицитные конститuenty позволяет провести анализ максимально объективно и обрисовать базовую технологию работы с отрицанием в рамках оригинального корпусного подхода.

Планировалось решить следующие задачи:

- 1) проанализировать возможности текущей версии программного комплекса «Генератор сбалансированного лингвистического корпуса и корпусный менеджер» на предмет обнаружения средств категории отрицания в англоязычном тексте (методы тестирования и анализа алгоритмов);
- 2) оценить возможности программного комплекса в части добавления программных функций по поиску отдельных лемм с учетом того, что они могут употребляться с отрицанием (методы анализа возможности внесения модификаций в структуру комплекса, описания технических параметров и моделирования новых алгоритмов);
- 3) создать и провести апробацию экспериментального программного решения (методы профессионально ориентированного программирования и эксперимента).

Высокий интерес со стороны лингвистов к проблеме использования средств категории отрицания, значительное количество публикаций по теме за последние годы, а также, что особенно важно, отсутствие универсальных технических решений проблемы учета языковых средств отрицания при анализе связных текстов, подчеркивает актуальность предлагаемого исследования.

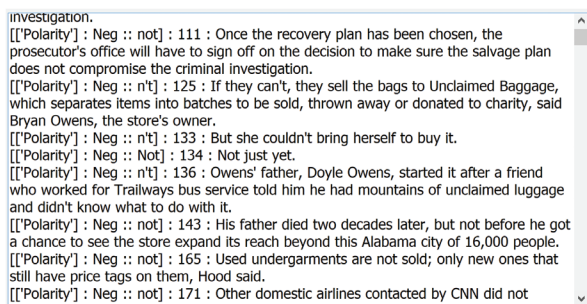
Новизна работы заключается в том, что впервые на упомянутой программной базе производится анализ заданных лемм с учетом полярности их употребления (есть отрицание / нет отрицания).

¹Наше перечисление не исчерпывает все возможные средства отрицания английского языка, даны лишь выборочные примеры.

Практическая ценность исследования состоит в создании рабочего инструмента для базовой оценки употребления средств категории отрицания в англоязычном тексте, вносится вклад в развитие функциональных возможностей программного комплекса.

ХОД ИССЛЕДОВАНИЯ И ЕГО РЕЗУЛЬТАТЫ

Рассмотрим *возможности текущей версии программного комплекса* «Генератор сбалансированного лингвистического корпуса и корпусный менеджер». При разметке генератором токены снабжаются характеристикой *polarity* («полярность»), которая может принимать значения «pos» («положительная») и «neg» («отрицательная»). При этом в качестве отрицательных для английского языка¹ маркируются исключительно частицы *not*. Поиск по параметру «Совокупный учет»: «Атрибуты = polarity» + «Значения = neg» в корпусе показал наличие 5 524 токенов (0,54 % от общего количества) в 5 102 предложениях (13,21 % от общего количества) (см. рис. 1):

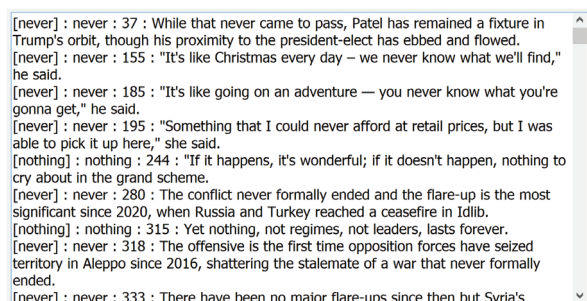


```
Investigation.
[["Polarity"] : Neg :: not] : 111 : Once the recovery plan has been chosen, the
prosecutor's office will have to sign off on the decision to make sure the salvage plan
does not compromise the criminal investigation.
[["Polarity"] : Neg :: n't] : 125 : If they can't, they sell the bags to Unclaimed Baggage,
which separates items into batches to be sold, thrown away or donated to charity, said
Bryan Owens, the store's owner.
[["Polarity"] : Neg :: n't] : 133 : But she couldn't bring herself to buy it.
[["Polarity"] : Neg :: Not] : 134 : Not just yet.
[["Polarity"] : Neg :: n't] : 136 : Owens' father, Doyle Owens, started it after a friend
who worked for Trailways bus service told him he had mountains of unclaimed luggage
and didn't know what to do with it.
[["Polarity"] : Neg :: not] : 143 : His father died two decades later, but not before he got
a chance to see the store expand its reach beyond this Alabama city of 16,000 people.
[["Polarity"] : Neg :: not] : 165 : Used undergarments are not sold; only new ones that
still have price tags on them, Hood said.
[["Polarity"] : Neg :: not] : 171 : Other domestic airlines contacted by CNN did not
```

Рис. 1. Фрагмент вывода запроса по совокупному учету «Отрицательная полярность»

Следует обратить внимание на то, что программа безошибочно идентифицировала *not* даже в сокращенных вариантах (*couldn't, hasn't* и пр.).

Прочие конститuenty, выраженные отдельным словом (*never, nobody, nothing* и др.), можно найти, используя поиск по набору лемм (см. рис. 2):



```
[never] : never : 37 : While that never came to pass, Patel has remained a fixture in
Trump's orbit, though his proximity to the president-elect has ebbed and flowed.
[never] : never : 155 : "It's like Christmas every day — we never know what we'll find,"
he said.
[never] : never : 185 : "It's like going on an adventure — you never know what you're
gonna get," he said.
[never] : never : 195 : "Something that I could never afford at retail prices, but I was
able to pick it up here," she said.
[nothing] : nothing : 244 : "If it happens, it's wonderful; if it doesn't happen, nothing to
cry about in the grand scheme.
[never] : never : 280 : The conflict never formally ended and the flare-up is the most
significant since 2020, when Russia and Turkey reached a ceasefire in Idlib.
[nothing] : nothing : 315 : Yet nothing, not regimes, not leaders, lasts forever.
[never] : never : 318 : The offensive is the first time opposition forces have seized
territory in Aleppo since 2016, shattering the stalemate of a war that never formally
ended.
[never] : never : 333 : There have been no major flare-ups since then but Syria's
```

Рис. 2. Фрагмент вывода запроса по набору лемм

¹Уточнение сделано в связи с тем, что текущая версия программного комплекса работает с русским, английским, немецким, французским, испанским и китайским языками.

В процентном выражении присутствие *never, nobody* и *nothing* составило 0,05 % для токенов и 1,37 % для предложений.

Поиск по приставкам и суффиксам производился с помощью функции определения символов в начале и конце токена. Здесь следует принимать во внимание погрешность, поскольку ищется не морфема, а простая последовательность знаков, например, для *un-* будет найдено не только *unorthodox, unclear, unjustly*, но и *under, unique* и т. п. Однако поиск по суффиксам дает практически нулевую погрешность: только *unless* при поиске по *-less* (поиск по *-free* выдал всего лишь один токен *carefree*).

Таким образом, мы установили, что текущая версия программного комплекса вполне удовлетворительно справляется с задачей поиска эксплицитных отрицаний в англоязычном тексте.

Далее была проведена оценка возможности программного комплекса в части добавления программных функций по поиску отдельных лемм с учетом того, что они могут употребляться с отрицанием.

Важным отличием программы-генератора является опция разметки так называемой зависимости, «т. е. числового указания на номер токена, который является главным для текущего токена» [Горожанов, 2024а, с. 27]. В нашем случае это позволяет, например, определить к какому слову относятся отрицания *no* или *not*, при этом главный и подчиненный токены могут находиться на расстоянии друг от друга (см. рис. 3):

4671	161	they	PRON	they	Case=...	NULL	4672	/
4672	161	have	VERB	have	Mood=...	NULL	4668	/
4673	161	no	DET	no		NULL	4675	/
4674	161	identifying	VERB	identify	Aspec=...	NULL	4675	/
4675	161	information	NOUN	infor...	Numb=...	NULL	4672	/
4676	161	.	PUNCT	.	Punct=...	NULL	4668	/

Рис. 3. Фрагмент таблицы токенов базы данных корпуса с указанием главного слова

Мы видим, что *no* удалено от главного слова *information* на одну позицию, но всё же числовой маркер верно указывает на их взаимосвязь. Заметим, что установление зависимостей происходит автоматически при разметке корпуса генератором на основе правил spaCy.

Поскольку такие данные заложены в разметке лингвистического корпуса, их *возможно* использовать, внося некоторые дополнения в программный код корпусного менеджера.

Итак, на следующем этапе исследования в корпусный менеджер была введена новая функция поиска зависимых слов по заданным параметрам. При этом был задействован принцип

масштабирования, или универсальности, программного решения, чтобы в будущем можно было применять этот инструмент для иных целей, выходящих за рамки анализа средств категории отрицания.

Функция использует стандартный интерфейс с добавлением только одного подпункта в меню «Коннотации». Для старта алгоритма оператору необходимо заполнить строки лемм и токенов исходной формой главного слова и искомыми лексическими единицами в исходной форме, которые должны находиться в прямой зависимости от него. В рамках поиска средств отрицания для английского языка запрос может иметь следующий вид: главное слово *agree*, зависимое слово *not*. При реализации такого запроса был получен представленный ниже результат (см. рис. 4):

```
agree : 3723 : Whomever would commit such a crime against a minor should be locked away, would you not agree?"
agree : 4805 : And leaders can't agree on how to tame this monster we have created.
agree : 6087 : "It's just that perfect thing when you're on set and a director will give you a note that maybe you don't agree with, or I know that she doesn't particularly like, and it's just that like familiar glance at each other of like, 'can't wait to talk about that later,'" the British star said.
agree : 6266 : What Jeffries had found in Johnson was a partner who he did not always agree with, but came to trust.
agree : 9883 : It's easy to cancel some Americans when you don't agree with them, but Septima Clark, one of the characters in McMahon's book, illustrates the folly of such an approach.
agree : 17432 : "We can't agree to a situation where the released prisoners will murder our people in the future."
Токенов: 6
Предложений: 6
Предложений всего: 38621
```

Рис. 4. Результат поискового запроса по лемме «agree» и зависимому слову «not»

Демонстрируемый фрагмент показывает, что программа не только верно распознала сокращенные формы, правильно отобразила кавычки (что может быть проблемой для некоторых программных решений), но и безошибочно связала глагол *agree* с относящимся к нему отрицанием *not* вне зависимости от расстояния между первым и вторым.

Далее мы проверили алгоритм на связке *chance* и *no* (см. рис. 5):

```
chance : 4034 : For its part, the Biden administration is taking no chances over whether Jolani has the capability to manage the threat from ISIS.
chance : 15412 : Applying online takes around 20 minutes and you must complete the process in one go — there is no chance to save as you go.
chance : 28303 : But with just a few days before a chance at a historic three-peat, the Kansas City quarterback was taking no chances with his nutrition.
chance : 35441 : From across the boroughs and otherwise warring ideological factions, there is widespread agreement that the mayor has little to no chance of winning the Democratic primary in June — and that he shouldn't even try.
Токенов: 4
Предложений: 4
Предложений всего: 38621
```

Рис. 5. Результат поискового запроса по лемме *chance* и зависимому слову *not*

Результат подтвердил правильность выбранного программного решения. Формы леммы *chance* были идентифицированы верно. При наличии двойного употребления лексемы *chance* была маркирована именно форма с отрицанием (образец № 28 303).

Далее программа была доработана в части организации вывода: к лемме в квадратных скобках было добавлено зависимое слово, которое находится в найденном контексте. Кроме того, при подсчете единиц указывается количество токенов, имеющих заданные зависимые слова и общее количество этих токенов (см. рис. 6):

```
money [any] : 7081 : But, the bot added, "I don't handle any money myself."
money [any] : 11736 : That strong stock performance comes as AI growth is expected to drive up demand for nuclear power, although Oklo isn't yet making any money.
money [any] : 13537 : "It was a hair-raising nightmare, you hold your breath and hope that the engineer called it a mud flow instead of a landslide, because if it's considered a landslide, then you don't get any money," she said.
money [no] : 22154 : Trump repeated his false claim that during his first term, Iran "had no money for Hezbollah, they had no money for Hamas."
money [no] : 22154 : Trump repeated his false claim that during his first term, Iran "had no money for Hezbollah, they had no money for Hamas."
money [no] : 24267 : In three previous years under Biden, USAID spent no money on condoms in the entire Middle East: A detailed federal report published last year said USAID did not provide or fund any condoms in the Middle East in the 2021, 2022 and 2023 fiscal years.
money [any] : 25846 : "But I don't want to take any money."
Токенов: 7 / 270
Предложений: 6
Предложений всего: 38621
```

Рис. 6. Результат поискового запроса по лемме *money* и зависимым словам *no*, *any* с более информативным выводом

В приведенном примере показано, что в корпусе имеется 270 упоминаний *money*, но только семь из них имеют в подчинении *no* или *any*.

Для подтверждения универсальности созданной функции мы провели дополнительное экспериментальное тестирование на материале сбалансированного лингвистического корпуса текста крупнейшей региональной газеты ФРГ «Westdeutsche allgemeine Zeitung»¹. В качестве главного слова был обозначен модальный глагол *können*, зависимой единицей выступило отрицание *nicht*. Был получен следующий результат (см. рис. 7):

```
können [nicht] : 144 : „Deshalb kann ich es auch nicht verstehen, dass so viele dieser Meinung sind.“
können [nicht] : 150 : Über den Klinikalltag sagt er: „Ohne Ausländer könnte das nicht funktionieren.“
können [nicht] : 160 : Selbst mitstimmen am 23. Februar könne er nicht, verrät Causevic kurz vor dem Abschied noch.
können [nicht] : 538 : Man kann den Betroffenen nicht alle Sorgen abnehmen.
können [nicht] : 620 : Zwar könne das Frühschwimmen im Südpool nicht stattfinden, jedoch könnten die Gäste in dieser Zeit das Sport- und Erlebnisbad Wananas nutzen.
können [nicht] : 845 : Die endgültige Fertigstellung hänge von verschiedenen Faktoren ab und könne derzeit noch nicht konkret abgeschätzt werden, heißt es in einem Bericht der Deutschen Verkehrszeitung.
können [nicht] : 904 : Luft konnten sich die Essener im Kampf um den Klassenerhalt trotz der imposanten Serie nicht verschaffen.
können [nicht] : 939 : Wie es mit ihm am Wochenende aussieht, kann ich noch nicht sagen.“
können [nicht] : 1017 : Der offensive Mittelfeldspieler kann aufgrund von Rückenbeschwerden in
```

Лемма können + подчиненные [nicht]

Рис. 7. Результат поискового запроса по лемме *können* и зависимому слову *nicht*

Алгоритм верно распознал отрицание при различных словоформах глагола (*könnte*, *könne*, *kann*, *konnten* и пр.) и на различном расстоянии от них, что было бы практически невозможно при «механическом» анализе, т. е. при поиске по шаблону «словоформа AND (+ n слов OR – n слов)».

¹Объем тестового микро-корпуса составил 48 540 токенов, что достаточно для первичного тестирования нашего алгоритма, но, безусловно, крайне мало для проведения современных серьезных корпусных исследований языковых явлений.

ЗАКЛЮЧЕНИЕ

Итак, мы разработали и апробировали базовую технологию идентификации средств категории отрицания в рамках оригинального корпусного подхода. При этом были полностью решены поставленные задачи. Были проанализированы возможности текущей версии программного комплекса «Генератор сбалансированного лингвистического корпуса и корпусный менеджер» на предмет обнаружения средств категории отрицания в англоязычном тексте. Были оценены возможности программного комплекса в части добавления программных функций по поиску отдельных лемм с учетом того, что они могут употребляться с отрицанием. В ходе исследования было принято решение использовать разметку зависимостей токенов. Наконец, была написана и апробирована новая функция корпусного менеджера,

которая безошибочно определила наличие у заданных лемм средств категории отрицания.

Поскольку при проектировании функции был применен принцип масштабирования, полученные результаты вышли за рамки первоначально поставленных задач. Программа способна работать не только с отрицаниями, но практически с любыми языковыми единицами. Например, возможен поиск по запросам «главное существительное – зависимое прилагательное», «главный глагол – зависимое существительное» и др.

В перспективе необходимо провести тестирование на материале других языков, уточнив особенности стандартной разметки полярности, а также соотнести полученный инструмент с исследованиями, посвященными определению параметров идиостиля, интерпретации художественного текста и оценке качества перевода.

СПИСОК ИСТОЧНИКОВ

1. Гращенков П. В., Студеникина К. А., Паско Л. И. Ограничение сочинительного острова в лингвистической компетенции больших языковых моделей // Вестник Санкт-Петербургского университета. Язык и литература. 2024. Т. 21. № 3. С. 668–688. DOI 10.21638/spbu09.2024.309. EDN HCAZUH.
2. Лагутина К. В., Бойчук Е. И., Лагутина Н. С. Автоматическая классификация русскоязычных интернет-текстов по жанрам // Искусственный интеллект и принятие решений. 2023. № 4. С. 103–114. DOI 10.14357/20718594230410. EDN BAKBAF.
3. Солнышкина М. И., Макнамара Д. С., Замалетдинов Р. Р. Обработка естественного языка и изучение сложности дискурса // Russian Journal of Linguistics. 2022. Т. 26. № 2. С. 317–341. DOI 10.22363/2687-0088-30171. EDN EQQVTI.
4. Гагарин С. Н. Базовые методики анализа языковых картин политики с помощью языка программирования Python и библиотеки NLTK (на материалах корпусов британского парламентского дискурса) // Филологические науки в МГИМО. 2024. Т. 10. № 2. С. 125–140. DOI 10.24833/2410-2423-2024-2-39-125-140. EDN GDGMAO.
5. Сафина З. М. Переводческий анализ художественного текста на языке Python // Глобальный научный потенциал. 2024. Т. 1. № 11(164). С. 177–180. EDN RTJTGQ.
6. Красикова Е. А. Роль корпусного менеджера в анализе употребления имен собственных в текстах электронных СМИ (на примере англоязычного корпуса CNN) // Филологические науки в XXI веке: актуальность, многополярность, перспективы развития : Сборник научных трудов. Краснодар: КубГУ, 2024. С. 45–49. EDN JPRHAE.
7. Степанова Д. В. Презентация существительных семантического ряда «традиционные энергоносители» в современных СМИ ФРГ (на материале текстов газеты «Франкфуртер Альгемайне Цайтунг») // Филологические науки в XXI веке: актуальность, многополярность, перспективы развития : Сборник научных трудов. Краснодар: КубГУ, 2024. С. 136–140. EDN PUDELJ.
8. Горожанов А. И. Архитектура сбалансированного лингвистического корпуса, полученного автоматическим путем (опыт Московского государственного лингвистического университета) // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2024а. № 11 (892). С. 24–30. EDN BCSCXQ.
9. Глушак В. М. Отрицание немецких полярных слов и выражений в автоматизированном анализе тональности текста // Филологические науки. Вопросы теории и практики. 2023. Т. 16. № 10. С. 3287–3292. DOI 10.30853/phil20230510. EDN CWDXEU.
10. Горожанов А. И. Метод программного анализа контекста лексической единицы // Актуальные проблемы филологии и педагогической лингвистики. 2024б. № 3. С. 178–190. DOI 10.29025/2079-6021-2024-3-178-190. EDN QNFATA.
11. Горожанов А. И. Расширение стандартного сбалансированного лингвистического корпуса, построенного по правилам spaCy, коннотативными характеристиками // Филологические науки. Вопросы теории и практики. 2023. Т. 16. № 11. С. 3888–3893. DOI 10.30853/phil20230594. EDN FVUIUL.
12. Сытько А. В. Отрицательные деонтические высказывания: семантика и прагматика (на материале немецкого и русского языков) // Вестник Минского государственного лингвистического университета. Серия 1: Филология. 2023. № 2 (123). С. 40–49. EDN MZPNGY.

REFERENCES

1. Grashchenkov, P. V., Studenikina, K. A., Pasko, L. I. (2024). Coordinate structure constraint in the linguistic competence of large language models. *Vestnik of Saint Petersburg University. Language and Literature*, 21(3), 668–688. DOI 10.21638/spbu09.2024.309. (In Russ.)
2. Lagutina, K. V., Boychuk, E. I., Lagutina, N. S. (2023). Automatic classification of Russian-language Internet texts by genre. *Artificial Intelligence and Decision Making*, 4, 103–114. DOI 10.14357/20718594230410. (In Russ.)
3. Solnyshkina, M. I., McNamara, D. S., Zamaletdinov, R. R. (2022). Natural language processing and discourse complexity studies. *Russian Journal of Linguistics*, 26(2), 317–341. DOI 10.22363/2687-0088-30171. (In Russ.)
4. Gagarin, S. N. (2024). Getting a handle on a Hansard with Python and NLTK, or how to tame the linguistic picture of British politics with NLP. *Linguistics & Polyglot Studies*, 10(2), 125–140. DOI 10.24833/2410-2423-2024-2-39-125-140. (In Russ.)
5. Safina, Z. M. (2024). Translation analysis of a literary text in Python. *Global scientific potential*, 1-11(164), 177–180. (In Russ.)
6. Krasikova, E. A. (2024). The role of the corpus manager in analyzing the use of proper names in electronic media texts (on the material of the English-speaking CNN corpus). *Philological sciences in the 21st century: relevance, multipolarity, development prospects* (pp. 45–49): conference proceedings. Krasnodar. (In Russ.)
7. Stepanova, D. V. (2024). Presentation of the semantic row “traditional energy carriers” in modern media in Germany (based on the texts of “Frankfurter Allgemeine Zeitung”). *Philological sciences in the 21st century: relevance, multipolarity, development prospects* (pp. 136–140): conference proceedings. Krasnodar. (In Russ.)
8. Gorozhanov, A. I. (2024a). Architecture of a balanced linguistic corpus built automatically (experience of Moscow State Linguistic University). *Vestnik of Moscow State Linguistic University. Humanities*, 11(892), 24–30. (In Russ.)
9. Glushak, V. M. (2023). Negation of German polar words and expressions in automated analysis of text tonality. *Philology. Theory & Practice*, 16(10), 3287–3292. DOI 10.30853/phil20230510. (In Russ.)
10. Gorozhanov, A. I. (2024b). Programming analysis of the lexical unit context. *Current issues in philology and pedagogical linguistics*, 3, 178–190. DOI 10.29025/2079-6021-2024-3-178-190. (In Russ.)
11. Gorozhanov, A. I. (2023). Extension of a standard balanced linguistic corpus built according to spacy rules by connotative characteristics. *Philology. Theory & Practice*, 16(11), 3888–3893. DOI 10.30853/phil20230594. (In Russ.)
12. Sytko, A. V. (2023). Negative deontic statements: semantics and pragmatics (on the material of the German and Russian languages). *Minsk State Linguistic University Bulletin. Series 1. Philology*, 2(123), 40–49. (In Russ.)

ИНФОРМАЦИЯ ОБ АВТОРЕ

Горожанов Алексей Иванович

доктор филологических наук, доцент
 профессор кафедры грамматики и истории немецкого языка факультета немецкого языка
 Московского государственного лингвистического университета

INFORMATION ABOUT THE AUTHOR

Gorozhanov Alexey Ivanovich

Doctor of Philology, Associate Professor,
 Professor at the Department of German Language Grammar and History
 Faculty of German, Moscow State Linguistic University

Статья поступила в редакцию
 одобрена после рецензирования
 принята к публикации

13.01.2025
 27.01.2025
 29.01.2025

The article was submitted
 approved after reviewing
 accepted for publication