

V E S T N I K

NOVOSIBIRSK STATE UNIVERSITY

Scientific Journal
Since 1999, November
In Russian

Series: Linguistics and Intercultural Communication

2026. Volume 24, № 2

CONTENTS

Theory of Language and Applied Linguistics

- Alypova S. A., Obolenskaya Y. L.* Functional-Semantic Analysis of Achievement Verbs as Part of the Progressive Periphrasis “Estar + Gerund” in Modern Peninsular Spanish 5
- Boginskaya O. A.* Research Gap Spotting Step: Evolution of the Genre Marker in the Era of Competition 15

Cognitive Studies and Intercultural Communication

- Samoylova A. V.* Axiological Markedness of English Words of the Year 29

Computer and Applied Linguistics

- Urikh A. E., Sherstinova T. Y.* Creation of a Biographical Dataset of Russian Prose Writers of the 20th Century Using ChatGPT-4o 42
- Belous A. A., Boikova A. P.* Application of Large Language Models for Qualitative Analysis of Loosely Structured Textual Data 55
- Grebenyuk A. V.* Cohesion and Coherence of Poetic Text in Neural Machine Translation 69
- Kolmogorova P. A., Artemenko E. D., Oreshina G. V., Koltsova E. Yu.* Towards Reliable Research: Development and Medical Verification of a Health Claims Corpus 83
- Popova T. I., Naumova V. V.* Computational Linguistics Methods for Identifying Language of Trauma Markers 97
- Khlusova Y. K.* “You Can’t Just Hop to the Hawaiian or Aeolian Islands: They’re Not Soviet Yet...”: Toponyms in Russian Short Stories of the 1930s 112

Linguoculturology and Sociolinguistics

- Grigoryeva A. S.* Field Research on the Language of Siberian Mennonites: Expedition Experience from the Villages of Neudachino and Apollonovka 124
- Krivopalova N. O.* The Lost Art of Communication: Interpersonal Communication Challenges in the Digital Society 136

<i>Nikolaeva A. V., Shorina L. V., Chizhikova S. N. Directive Speech Acts in Advertising Discourse (Based on English and Russian)</i>	146
---	-----

Translation and Translation Studies

<i>Chernykh A. S. Analysis of Translating Russian Participles into Malay</i>	160
--	-----

Instructions for Contributors	173
--------------------------------------	-----

Editor-in-Chief Dr. Sc. (Linguistics) Maria K. Timofeeva (Novosibirsk)
Deputy Editor-in-Chief Dr. Sc. (Linguistics), Prof. Michèle Debrenne (Novosibirsk)
Executive Secretary Cand. Sc. (Linguistics), Assoc. Prof. Anatoli F. Fefelov (Novosibirsk)
Secretary Darya A. Savostiyanova (Novosibirsk)

Foreign Text Editor

Cand. Sc. (Linguistics), Assoc. Prof. Olga A. Ryzhkina (Novosibirsk)

Members of the Editorial Board

Cand. Sc. (Linguistics), Assoc. Prof. Natalia V. Kozlova (Novosibirsk)
 Dr. Sc. (Philology), Prof. Alexey D. Palkin (Moscow)
 Cand. Sc. (Philology), Assoc. Prof. Ivan A. Remorov (Novosibirsk)
 Cand. Sc. (Linguistics), Assoc. Prof. Olga A. Ryzhkina (Novosibirsk)
 Cand. Sc. (Philology), Assoc. Prof. Anna L. Solomonovskaya
 PhD (Cross-Cultural Psychology), Assistant Prof. Snežana Stupar-Rutenfrans (Utrecht, Netherlands)
 PhD, Assoc. Prof. Maria V. Khokhlova (St. Petersburg)

Editorial Board of the Series

Academician of the Russian Academy of Sciences, Prof. A. E. Anikin (Novosibirsk)
 Dr. Sc. (Philology), Prof. V. E. Gorshkova (Irkutsk), Dr. Sc. (Linguistics), Senior researcher L. L. Iomdin (Moscow),
 Dr. Sc. (Linguistics), Prof. N. B. Koshkareva (Novosibirsk),
 Cand. Sc. (Linguistics), Doctorat National en sciences du langage A. M. Lavrentev (Lyon, France),
 Dr. Sc. (Linguistics), Prof. O. N. Aleshina (Taipei, The Republic of China, Taiwan),
 Dr. Sc. (Linguistics), Prof. D. B. Nikulicheva (Moscow), Dr. Sc. (Philology), Prof. L. G. Panin (Novosibirsk),
 Dr. Sc. (Linguistics), Prof. Z. G. Proshina (Moscow),
 Dr. Sc. (Philosophy, Neuroscience), Assoc. Prof. A. N. Savostyanov (Novosibirsk),
 Dr. Sc. (Linguistics), Prof. I. Ya. Selyutina (Novosibirsk), Dr. Sc. (Philology), Prof. I. V. Silantyev (Novosibirsk),
 Dr. Sc. (Linguistics), Prof. V. A. Stepanenko (Irkutsk), Dr. Sc. (Linguistics), Prof. Ye. F. Tarasov (Moscow),
 Dr. Sc. (Linguistics), Prof. N. V. Ufimtseva (Moscow), PhD (Philology) V. Obry (Mulhouse, France),
 PhD (Philology), Assoc. Prof. I. Chekhova (Sofia, Bulgaria),
 Dr. Sc. (Linguistics), Prof. I. V. Shaposhnikova (Novosibirsk)

*The journal is published quarterly in Russian since 1999
 by Novosibirsk State University Press*

The address for correspondence

Institute of Humanities, Novosibirsk State University
 1 Pirogov Street, Novosibirsk, 630090, Russian Federation

Tel. +7 (383) 363 42 23

E-mail address: lingua@vestnik.nsu.ru

On-line version: <http://elibrary.ru>

Научная статья

УДК 811.134.2

DOI 10.25205/1818-7935-2026-24-2-5-14

Функционально-семантический анализ глаголов-достижений в составе граммем прогрессивного аспекта в пиренейском варианте современного испанского языка

Светлана Алексеевна Алыпova¹

Юлия Леонардовна Оболенская²

¹ МГИМО МИД России,
Москва, Россия

² Московский государственный университет им. М. В. Ломоносова,
Москва, Россия

alypova_sveta@mail.ru
obolens7@yandex.ru

Аннотация

Объектом исследования является аспектуальная граммема прогрессива и, в частности, герундиальная перифраза «estar + герундий» как ее формальное проявление в испанском языке. Прогрессив представляет собой разновидность имперфектива, соответственно, изучаемая перифраза является имперфективирующей глагольной конструкцией. Помимо грамматического аспекта, рассматривается акциональность – семантическая характеристика глагола; в частности акциональный класс так называемых глаголов-достижений, или моментативов. Материалом для исследования послужили контексты из корпуса современного испанского языка RAE CORPES XXI, основное внимание уделено пиренейскому национальному варианту. Целью исследования является изучение свойств глаголов рассматриваемого акционального класса (на примере глагола «llegar» ‘прибыть, прибывать’), критический анализ и дополнение устоявшегося в испаноязычной традиции подхода к их характеристике и применяемой терминологии. В статье предпринята попытка уточнения классификации З. Вендлера для описания случаев использования имперфективных форм подобных глагольных предикатов. Основываясь на теории С. Г. Татевосова, был сделан вывод о необходимости рассматривать в испанском языке так называемые глаголы-достижения как совокупность значений основных перфективной и имперфективной форм в эпизодической интерпретации, поскольку это способствует полноценному пониманию их внутренней акциональной структуры. Данное наблюдение, в свою очередь, свидетельствует о не совсем корректном применении к ним ранее используемого в испанской грамматической традиции термина «достижения» («logros»). Также проведенный анализ контекстов позволил прийти к выводу об абсолютном превалировании использования с изучаемыми глаголами перифразы «estar + герундий» в имперфективных временах. Прогрессивные перфективные видовременные формы (прогрессивный аорист и прогрессивный перфект) практически не встречаются в пиренейском национальном варианте, притом что с глаголами других динамических акциональных классов они достаточно распространены. Тем не менее в немногочисленных контекстах с перфективными формами даже в составе имперфективирующей герундиальной конструкции отмечено активное проявление моментативной составляющей основного глагола (например, выявленные значения итеративности или дистрибутивности), что вызвано его акциональными характеристиками.

Ключевые слова

испанский язык, прогрессивный аспект, герундиальные перифразы, акциональность, глаголы-достижения, корпус RAE CORPES XXI

© Алыпova С. А., Оболенская Ю. Л., 2026

Для цитирования

Алыпova С. А., Оболенская Ю. Л. Функционально-семантический анализ глаголов-достижений в составе граммем прогрессивного аспекта в пиренейском варианте современного испанского языка // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 5–14. DOI 10.25205/1818-7935-2026-24-2-5-14

Functional-Semantic Analysis of Achievement Verbs as Part of the Progressive Periphrasis “Estar + Gerund” in Modern Peninsular Spanish

Svetlana A. Alypova¹, Yulia L. Obolenskaya²

¹ MGIMO University
Moscow, Russian Federation

² Lomonosov Moscow State University
Moscow, Russian Federation

alypova_sveta@mail.ru
obolens7@yandex.ru

Abstract

The current article focuses on the progressive aspect and, in particular, the gerundial periphrasis “estar + gerund” as its formal expression in the Spanish language. The progressive is often considered to be a subtype of the imperfective aspect. Consequently, according to numerous studies, the periphrasis “estar + gerund” is an imperfectivizing construction. In addition to the grammatical aspect, the research considers actionality as a semantic characteristic of the verb; in particular, the actional class of the so-called achievement verbs or momentatives. The material for the research are contexts from the corpus of modern Spanish RAE CORPES XXI, the main attention is paid to the Peninsular Spanish variety. The purpose of the article is to study the properties of the considered actional class verbs (using the example of the verb “llegar” ‘to arrive’), to critically analyze and supplement the approach to their characterization and terminology used in the Spanish tradition. The study attempts to clarify Z. Vendler’s classification to describe the cases of using these verb predicate imperfective forms. Based on S. G. Tatevosov’s theory, the research concludes that achievement verbs in Spanish should be viewed as a combination of meanings of both perfective and imperfective forms in episodic interpretation. This perspective facilitates a more comprehensive understanding of their internal actional structure, aiding in a deeper analysis of the progressive aspect’s role in conveying semantic nuances. This observation indicates a somewhat incorrect application of the term “achievements” (“logros”) from the Spanish grammatical tradition to the studied verbs. The context analysis leads to the conclusion that the periphrasis “estar + gerund” predominates in imperfective tenses with the verbs in question. Progressive perfective tense forms (progressive preterite and progressive perfect) are rarely encountered in the Peninsular Spanish variety, although they are quite common with the verbs of other dynamic actional classes. Nevertheless, in the few contexts featuring perfective forms within the imperfectivizing gerundial construction, an active manifestation of the momentaneous aspect of the main verb is noted, such as iterative or distributive meanings, caused by its actional characteristics.

Keywords

Spanish language, progressive aspect, gerundial periphrases, actionality, achievement verbs, RAE CORPES XXI

For citation

Alypova S. A., Obolenskaya Y. L. Functional-semantic analysis of achievement verbs as part of the progressive periphrasis “estar + gerund” in modern peninsular Spanish. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 5–14. DOI 10.25205/1818-7935-2026-24-2-5-14

Введение. Прогрессивный аспект

Объектом данного исследования являются герундиальные перифразы, рассматриваемые нами как формальное выражение граммемы прогрессивного аспекта в испанском языке.

Прежде всего представляется необходимым обосновать понятийный объем терминов, используемых в работе. Отечественный лингвист Ю. С. Маслов определяет общее категориаль-

ное значение прогрессива как «процессность, соединенную со специфической конкретностью, “сиюминутностью” действия, с его приуроченностью к определенному моменту или отрезку времени, четко выделяемому среди ряда других моментов или отрезков» [Маслов, 1984/2004. С. 42]. Исследовательница, занимающаяся изучением аспекта в испанском и в русском языках, Е. В. Горбова характеризует значение испанского прогрессива как «трансформацию значения конкретности в более эмфатическое значение актуализации действия» [Горбова, 2017. С. 182], опираясь на автора статьи о перифразах в «Дескриптивной грамматике испанского языка» А. Ильера [Yllera, 1999].

Герундиальные перифразы в испанском языке

Герундиальные перифразы представляют собой морфосинтаксические единства вспомогательного или полувспомогательного глагола, выражающего весь спектр грамматических значений (время, лицо, число, наклонение, залог), и полнозначного глагола, несущего основную лексико-семантическую нагрузку и стоящего в нефинитной форме герундия. Для определения герундиальных перифраз немаловажна грамматикализованность всей конструкции, необходимо отличать их от простого сочетания глагола и герундия, большинством лингвистов в качестве основного критерия рассматривается степень десемантизации финитного глагола. Также процитируем систему критериев «Новой грамматики испанского языка» («Nueva gramática de la lengua española») 2009 года, применяемую для определения вспомогательных глаголов. Согласно Грамматике [NGLE, 2009. С. 2117], вспомогательные глаголы: 1) допускают проклизу безударных местоимений; 2) не допускают собственное преобразование в пассивный залог, однако полнозначные глаголы таким образом могут быть преобразованы; 3) не ограничивают выбор субъекта действия, но согласуются с ним; 4) неспособны «выбирать» зависимые слова, относящиеся к глагольной группе; 5) неспособны «выбирать» полнозначный глагол. Однако последний критерий подвергается многочисленным сомнениям, оспаривается многими лингвистами, а также оспорен и в наших работах [Алыпova, 2023].

Помимо основного определения Ю. С. Маслова, касающегося понятия «прогрессив», необходимо отметить важную характеристику герундиальных перифраз, выделенную в «Новой грамматике испанского языка»: «Todas las perífrasis de gerundio muestran una acción, un proceso o un estado de cosas en su curso» [NGLE, 2009. С. 2185] ‘все герундиальные перифразы показывают действие, процесс или положение дел в их развертывании, в развитии’. В настоящем исследовании мы рассматриваем только самую распространенную в узусе перифразу «*estar* + герундий».

Акциональность и акциональные классификации

Следует определить понятийный объем и второго термина, составляющего семантическую характеристику глагольных предикатов – акциональность. В испанском языке «акциональность» синонимична термину «лексическая аспектуальность» («*aspecto léxico*»). Теоретическую значимость в этой связи представляет акциональная классификация американского ученого З. Вендлера [Vendler, 1967], предложившего распределить глаголы английского языка на четыре акциональных класса на основе трех оппозиций, определяющих базовую семантику предиката: стативность/динамичность; предельность/непредельность; пунктивность/дуративность. Вендлер делит глагольные предикаты на четыре группы: «*states*» («состояния»), «*activities*» («деятельности»), «*accomplishments*» («свершения») и «*achievements*» («достижения») (см. таблицу).

Акциональные классы по З. Вендлеру

Vendler's classification

	Динамичность	Предельность	Дуративность
States / Estados / Стативы	—	—	+
Activities / Actividades / Деятельности	+	—	+
Accomplishments / Realizaciones / Свершения	+	+	+
Achievements / Logros / Достижения	+	+	—

В переводе данных понятий на русский язык мы ориентируемся на работы Е. В. Падучевой [Падучева, 2009] и С. Г. Татевосова [Татевосов, 2015]. Также стоит обратить внимание на предложенный эквивалентный перевод терминов Вендлера В. П. Недялковым [Недялков, 1986]: «стативы», «агентивы», «терминативы» и «моментативы» (представлены в соответствующем порядке с вышеизложенными терминами).

Данная классификация претендовала на универсальность и была перенесена на материалы многих других языков (в том числе на русский и испанский). Однако мы разделяем мнение ряда российских и зарубежных исследователей о том, что данная классификация не может быть применима к любой лингвистической системе, поскольку она должна учитывать природу межъязыкового варьирования. Тем не менее важно подчеркнуть, что в «Дескриптивной грамматике испанского языка» [De Miguel, 1999] и в «Новой грамматике испанского языка» [NGLE, 2009] для характеристики испанских глаголов по критерию акциональности используется именно классификация Вендлера. В испаноязычных исследованиях перечисленные акциональные классы именуются соответственно: «estados», «actividades», «realizaciones o actuaciones» и «consecuciones o logros». Данной терминологией пользуются многие испаноязычные исследователи: в частности, такие крупные лингвисты, как И. Боске, Х. Гутьеррес-Рексач в труде «Основы формального синтаксиса» [Bosque & Gutiérrez-Rexach, 2009. С. 296–312], а также М. Х. Арче в статье о лексическом аспекте в авторитетном издании «Энциклопедия испанской лингвистики» [Arche, 2016]. Это обстоятельство послужило основой для осуществления нашего анализа.

В статье мы рассматриваем глаголы, относящиеся к классу достижений по Вендлеру («achievements» или «logros»). Глаголы этого класса динамичны, обладают временным пределом, но не обладают длительностью. Достижениями считаются точечные действия, обозначающие резкую смену двух ситуаций. Например, в испанском языке глаголами-достижениями являются, согласно «Дескриптивной грамматике испанского языка» [De Miguel, 1999. С. 3033], глаголы *llegar* ‘прибывать’, *romper* ‘разбивать, ломать’, *apagar* ‘выключать’, *encender* ‘включать’, *lograr* ‘достигать’, *alcanzar* ‘достигать’, *morir* ‘умирать’, *nacer* ‘рождаться’ и т. д. Крупнейший аспектолог С. Г. Татевосов, комментируя классификацию Вендлера, пишет о том, что глаголы-достижения «достигают кульминации, или предела, после чего возникает результирующее состояние. <...> С возникновением результирующего состояния ситуация с неизбежностью завершается» [Татевосов, 2015. С. 7–8]. Так, результирующим состоянием ситуации, обозначенной предикатом *alcanzar* ‘достигать’ в примере «Juan alcanzó la cima» ‘Хуан достиг вершины’, будет ситуация ‘Хуан находится на вершине’.

Проблема заключается в том, что упомянутые глаголы могут реализовывать разные значения в разных грамматических формах. Как известно, в отличие от английского, который лежит в основе классификации Вендлера, в испанском языке прошедшие времена имеют аспектуальное различие: существуют имперфективная и различные перфективные формы. Например,

форма простого перфекта «*llegó*» ‘прибыл’, которую можно было бы назвать «достижением» или «моментативом», отлична от формы имперфекта «*llegaba*» ‘прибывал’, которую интерпретировать как «достижение» сложно, она обозначает процесс, но никак не момент кульминации. Об этом делает общее замечание автор статьи в лексическом аспекте в «Дескриптивной грамматике» Э. де Мигель [De Miguel, 1999. С. 2988], комментируя так называемый «флексивный» аспект (то, что в общепринятой лингвистической терминологии именуется просто «аспектом») и свойство имперфекта описывать с точки зрения «неограниченности», «незавершенности» действия, обозначаемые лексически «ограниченными» (*delimitados*) глаголами. Однако де Мигель освещает таким образом лишь видовременную систему испанского глагола, акциональности это касается лишь опосредованно, термин «достижения» в работе не оспаривается. Тем не менее очевидно, что не во всех своих формах глагол, традиционно называющийся «достижением», может быть охарактеризован как «достижение».

В высказанных в работах С. Г. Татевосова теоретических положениях речь идет об акциональной вариативности, т. е. о возможности сосуществования в языке **разных** акциональных интерпретаций разных грамматических форм **одного и того же** глагола. Имеется в виду так называемая проблема непрямого доступа, т. е. невозможность рассматривать глагол как лексему изолированно, вне контекста, лишая его аспектуально-временных показателей. В связи с описываемым феноменом Татевосов предлагает рассматривать каждую аспектуальную форму как **сумму** акциональных значений, не ориентируясь на свойства **ненаблюдаемых лексем** и избегая классификации глаголов в инфинитивной форме. Таким образом в качестве акциональной характеристики глагола он предлагает рассматривать **сумму** всех возможных акциональных значений основной перфективной и основной имперфективной форм (в нашем случае это простой перфект и имперфект) в эпизодической интерпретации, так как, по его доказательному утверждению, значения остальных форм оказываются либо равны этим, либо предсказуемо выводятся из них. Мы разделяем данное утверждение ученого, опираясь на его классификацию пяти основных акциональных значений: состояния, (единичный) процесс, вхождение в состояние, вхождение в процесс и мультипликативный процесс. Вслед за Татевосовым будем характеризовать глаголы следующим образом:

llegar <вхождение в состояние; процесс>

Здесь до точки с запятой указаны все значения, принимаемые глаголом в основной перфективной форме, а после точки с запятой – в основной имперфективной форме (обе – в эпизодической интерпретации). Такие глаголы, как *llegar*, согласно терминологии Татевосова, называются сильными предельными глаголами. Исходя из вышесказанного, можно говорить о том, что относить такие глаголы только к классу глаголов-достижений (или моментативов) не совсем корректно.

Следует сделать вывод, что в контексте «встраивания» рассматриваемых глаголов в состав герундиальных перифраз на передний план выходят имперфективные значения, т. е. указание на процесс, а не на моментальное достижение результата. Именно за счет такой неоднозначной семантики испанских глаголов-достижений их сочетание с герундиальными перифразами позволяет реализовать значение, близкое к классу свершений («*accomplishments*», по Вендлеру), от которых они отличаются именно отсутствием семы длительности. Следовательно, в герундиальных конструкциях глаголы-достижения обозначают не только точечное действие, но также и поэтапный процесс «подготовительной фазы», приводящей в итоге к данному достижению, они зачастую приобретают проспективное значение.

Корпусное исследование

Высказанные наблюдения подтверждаются примерами из корпуса. Мы пользуемся инструментами корпуса Испанской Королевской Академии *RAE CORPES XXI*¹. Нами рассматриваются контексты, относящиеся только к пиренейскому национальному варианту, за исключением нескольких случаев, необходимых для иллюстрирования отмеченных особенностей. В качестве примера используется наиболее частотный в узусе, согласно данным корпуса, глагол-достижение *llegar* ‘прибывать’.

Отметим тот факт, что изначально Вендлер в своей теории подчеркивал невозможность употребления глаголов-достижений в прогрессивных формах. Тем не менее при анализе текстов, представленных в испанском корпусе, нами было обнаружено всего 563 случая употребления перифразы «*estar* + герундий», где в форме герундия стоит глагол *llegar*, что составляет около пяти случаев возникновения данного сочетания на каждый миллион слов, что подтверждает справедливость заключения о том, что такое сочетание возможно, хотя и не является частотным. В наших предыдущих исследованиях (например, [Алыпova, 2023]) уже был сделан вывод о важности роли акциональных характеристик основного глагола в составе герундиальных перифраз.

При рассмотрении случаев реализации процессного, а не моментативного значения глагола *llegar* в составе анализируемой перифразы, необходимо отметить несколько нюансов смысла, которые зависят от того, в каком времени стоит вспомогательный глагол *estar*. В тех случаях, когда глагол *estar* стоит в настоящем времени – таких контекстов большинство (399 из 563) – степень актуализации действия, свойственная всем герундиальным перифразам, максимальна. Контекстуальное значение свидетельствует о том, что описываемый процесс имеет начало, но пока не завершен (находится в развитии, в приближении к достижению), а данная перифраза приобретает проспективное значение, обозначает стадию, предшествующую достижению цели:

(1) El tren ya **está llegando** a la estación.²

Поезд является прибывающим (наст. время *estar* + герундий *llegar*) на станцию

‘Поезд **прибывает** на станцию’

(2) <...> es como si en el Titanic, a la gente que va en la última clase ya le **está llegando** el agua al cuerpo y tú estás en el vagón de primera en plan “madre mía, qué mal sujeta los cubiertos el de al lado”.³

<...> вода уже является достигающей (наст. время *estar* + герундий *llegar*) тел людей, путешествующих третьим классом <...>

‘< ... > это как на «Титанике»: вода уже **накрывает** (досл. **достигает**) людей, путешествующих третьим классом, а ты сидишь в первом классе, такой типа «мамочки, как же плохо сосед держит столовые приборы»’.

Значения, наблюдаемые в имперфекте (136 случаев из 563), позволяют сделать вывод о том, что действие, зафиксированное в плане прошедшего, характеризуется устремлением в будущее с точки зрения абсолютного времени. Здесь проспективность выражается только с помощью таксиса. Таким образом, данная перифраза в имперфекте имеет скорее описательную функцию:

¹ <https://www.rae.es/corpes/> – корпус текстов XXI в., представленных в различных жанрах: от интернет-форумов и периодики до художественных произведений и политических речей, содержит около 25 млн словоформ.

² Díaz Yanes, A. *Sin noticias de Dios*, 2001 (Здесь и далее в сносках приведены источники примеров так же, как они указаны в корпусе – прим. С. А.).

³ Jabois, M. «*ÍNIGO ERREJÓN. Diputado y líder de Más País*». El País, 2020.

(3) <...> pero lo que no dijo aquella noche era que entonces sentía una repentina tristeza por carecer ya de tiempo futuro, porque su vida **estaba llegando** al final.⁴

<...> его жизнь являлась подходящей (имперфект *estar* + герундий *llegar*) к финалу <...>

‘<...> но чего он не сказал той ночью, так это того, что тогда он чувствовал внезапную грусть: у него больше нет будущего, потому что его жизнь **подходила** к концу.’

Отметим, что случаи использования данной перифразы с глаголом *estar* в форме простого перфекта *pretérito perfecto simple* (испанский прогрессивный аорист⁵) не были обнаружены в пиренейском национальном варианте.

Однако для полноты проведенного лингвистического анализа были рассмотрены также латиноамериканские тексты, в которых было зафиксировано пять подобных контекстов, причем интересно, что все они зафиксированы только в 3-м лице (*estuvo, estuvieron*). В этих случаях можно говорить о приобретении перифразой типичного для глаголов-достижений в перфективном времени итеративного значения (значение многократности, повторяемости, сближающего даже со значением мультипликативного процесса): с точки зрения «внешнего» наблюдения, действие рассматривается как завершенное, поскольку время, в котором стоит вспомогательный глагол, четко указывает на достижение действием *правого предела*. Однако с точки зрения «внутреннего» развития герундиальная форма *llegando* указывает на пролонгацию действия, его динамичность и повторяемость, что, благодаря внешним лексическим единицам (например, наречиям *continuamente* ‘непрерывно’, *casi siempre* ‘почти всегда’), может обозначать даже некоторую регулярность, периодичность:

(4) San Marcos no se quedó atrás y **estuvo llegando** continuamente, pero la defensa somoteña sofocó cada amenaza.⁶

Сан Маркос не остался позади и являлся наступающим (аорист *estar* + герундий *llegar*) постоянно <...>

‘Сан-Маркос не отставал и непрерывно наступал, но защита Сомото подавляла каждую угрозу.’

(5) Durante meses, el marino **estuvo llegando** casi siempre a mi casa, aunque en ocasiones a alguna cafetería o a algún parque, a las horas más insospechadas.⁷

В течение нескольких месяцев моряк являлся приходящим (аорист *estar* + герундий *llegar*) почти всегда в мой дом <...>

‘В течение нескольких месяцев моряк почти всегда приходил ко мне домой, хотя иногда и в какое-нибудь кафе или в какой-нибудь парк, в самые неожиданные часы.’

Кроме того, в этом контексте мы наблюдаем и значение дистрибутивности, распределенности, или дисперсности во времени или пространстве, подкрепляемые второстепенными членами предложения (подчеркнуты в примерах):

(6) Durante meses, el marino **estuvo llegando** casi siempre a mi casa, aunque en ocasiones a alguna cafetería o a algún parque, a las horas más insospechadas.

‘В течение нескольких месяцев моряк почти всегда **приходил** ко мне домой, хотя иногда и в какое-нибудь кафе или в какой-нибудь парк, в самые неожиданные часы.’

⁴ García Morales, A. *El testamento de Regina*, 2001.

⁵ Терминология – см. работы Е. В. Горбовой, в частности [Горбова, 2017].

⁶ «*Listos los equipos para la Hexagonal*». La Prensa de Nicaragua, 2002.

⁷ Venegas Castaing, J. A. *Las aventuras de Liu Yuan, capitán de ultramar*, 2004 (Costa Rica).

(7) **Estuvieron llegando a diferentes horas de la mañana**, dependiendo de la distancia de sus comarcas y la topografía.⁸

Они являлись прибывающими (аорист *estar* + герудний *llegar*) в разные утренние часы <...>

‘Они **прибывали** в разное время утром, в зависимости от удаленности их районов и рельефа местности.’

Был обнаружен только один случай использования данной перифразы с глаголом *llegar* в перфекте *pretérito perfecto compuesto* (прогрессивный перфект) на территории Испании, а всего, с учетом латиноамериканских стран, таких контекстов двенадцать. Во всех выявленных случаях обнаруживаются те же значения, что и в случае *pretérito perfecto simple* – итеративность и дистрибутивность (в данном примере подчеркивается множественным числом подлежащего с определением *numerosos* ‘многочисленные’):

(8) De la misma forma, desde los países del Este **han estado llegando** a España numerosos inmigrantes que deberían haber sido detenidos en la primera frontera Schengen.⁹

<...> из стран Востока являлись прибывающими (перфект *estar* + герудний *llegar*) в Испанию многочисленные иммигранты <...>

‘Точно так же из восточных стран в Испанию **прибывали** многочисленные иммигранты, которые должны были быть задержаны на первой шенгенской границе.’

Также был зафиксирован один контекст с *pretérito pluscuamperfecto*, а всего шесть, учитывая все национальные варианты. Здесь так же, как и в других перфективных временах, можно отметить значение итеративности или дистрибутивности, отмеченное также лексически обстоятельством времени *día tras día* ‘изо дня в день’:

(9) Apreciaban vivir en el centro de la capital y habían rechazado hasta el momento todas las ofertas de compra que día tras día les **habían estado llegando**.¹⁰

<...> они отклонили к тому моменту все предложения о покупке, которые изо дня в день являлись поступающими (плюсquamперфект *estar* + герудний *llegar*) к ним.

‘Им нравилось жить в центре столицы, и к тому моменту они отклонили все предложения о покупке, которые **поступали** к ним изо дня в день.’

Сходные результаты обнаруживаются и в контекстах с рядом глаголов того же акционального класса, однако их рассмотрение не входило в задачи данной статьи.

Выводы

Таким образом, можно сделать вывод о том, что для определения понятийного объема глагольной акциональности следует учитывать особые значения конкретных видовременных форм, а не абстрактных лексем в нефинитной форме. Кроме того, несмотря на сложившуюся под влиянием теории Вендлера традицию, следует рассматривать так называемые глаголы-достижения как совокупность значений перфективной и имперфективной форм, поскольку это может дать полноценную картину их внутренней акциональной структуры, что в свою очередь свидетельствует о не совсем корректном применении к ним изначального термина «достижения». Подобные глаголы обладают яркими процессными характеристиками в имперфективных формах, и поэтому они способны активно встраиваться в герундиальные перифразы. В соот-

⁸ Rodríguez de León, F. *Con flores a la reina*, 2002 (República Dominicana).

⁹ Dolado, J. J.; Vázquez, P., ed.: *Ensayos sobre los efectos económicos de la inmigración en España*, 2008.

¹⁰ March, J. A. *Espérame en la Habana*, 2008.

ветствии с подробно разработанной теорией С. Г. Татевосова, такие глаголы можно характеризовать как <вхождение в состояние; процесс> и именовать сильными предельными глаголами.

Проведенный анализ позволил прийти к выводу об абсолютном превалировании использования в испанском языке герундиальных перифраз с данными глаголами в имперфективных временах. Прогрессивный аорист или прогрессивный перфект практически не встречаются в пиренейском национальном варианте, при этом подчеркнем, что с глаголами других динамических акциональных классов упомянутые формы, согласно иным нашим исследованиям, достаточно распространены. В немногочисленных контекстах с перфективными формами мы отмечаем активное проявление моментативной составляющей основного глагола даже в составе имперфективирующей герундиальной конструкции, т. е. выявленное нами значение итеративности, по-видимому, вызвано именно акциональными характеристиками глагола, стоящего в форме герундия.

Список литературы

- Алыпova С. А. Акциональные классы глаголов в составе герундиальных перифраз: исследование на основе корпуса современного испанского языка // Современная наука: актуальные проблемы теории и практики. Серия «Гуманитарные науки». М., 2023. Вып. 5. С. 107–113.
- Горбова Е. В. Грамматическая категория аспекта и контекст: на материале испанского и русского языков. СПб.: Изд-во Санкт-Петербург. гос. ун-та, 2017. 418 с.
- Маслов Ю. С. Избранные труды. Аспектология. Общее языкознание. М.: Языки славянской культуры, 2004. 846 с.
- Недялков В. П. Основные типы начинательных глаголов: инхоативы, ингрессивы, инцептивы // Языковое общение и его единицы. Калинин: КГУ, 1986. С. 124–135.
- Падучева Е. В. Лексическая аспектуальность и классификация предикатов по Маслову – Вендлеру // Вопросы языкознания. М., 2009. № 6. С. 3–20.
- Татевосов С. Г. Акциональность в лексике и грамматике. М.: Языки славянской культуры, 2015. 368 с.
- Arche M. J. Aspecto léxico // Gutiérrez-Rexach J. (ed.) *Enciclopedia de Lingüística Hispánica*. Vol. 1. London & New York: Routledge, 2016. P. 405–415.
- Bosque I., Gutiérrez-Rexach J. *Fundamentos de sintaxis formal*. Madrid: Ediciones Akal, S.A., 2009. 800 p.
- De Miguel E. El aspecto léxico // Bosque I., Demonte V. (dirs.). *Gramática descriptiva de la lengua española*. Vol. 2. Madrid: Espasa-Calpe, 1999. P. 2977–3060.
- NGLE = Real Academia Española y Asociación de Academias de la Lengua Española: *Nueva Gramática de la Lengua Española. Morfología y sintaxis*. Madrid: Espasa, 2009. 4032 p.
- Vendler Z. *Verbs and times* // *Linguistics in philosophy*. Ithaca: Cornell University Press, 1967. P. 97–121.
- Yllera A. Las perífrasis verbales de gerundio y participio // Bosque I., Demonte V. (dirs.). *Gramática descriptiva de la lengua española*. Vol. 2. Madrid: Espasa-Calpe, 1999. P. 3391–3442.

References

- Alypova S. A. Aspectual Classes of Verbs in Verbal Periphrases with the Gerund: a Study Based on the Corpus of the Modern Spanish Language. In: *Modern Science: actual problems of theory and practice. Series Humanities*. Moscow, 2023, no. 5, pp. 107–113. (in Russ.)
- Arche M. J. Aspecto léxico. In: Gutiérrez-Rexach J. (ed.) *Enciclopedia de Lingüística Hispánica*. Vol. 1. London & New York: Routledge, 2016, pp. 405–415.

- Bosque I., Gutiérrez-Rexach J.** Fundamentos de sintaxis formal. Madrid: Ediciones Akal, S. A., 2009, 800 p.
- De Miguel E.** El aspecto léxico. In: Bosque, I., Demonte V. (dirs.). *Gramática descriptiva de la lengua española*. Vol. 2. Madrid: Espasa-Calpe, 1999, pp. 2977–3060.
- Gorbova E. V.** Grammatical category of aspect and context: based on the material of Spanish and Russian languages. St. Petersburg: SPbSU Publ., 2017, 418 p. (in Russ.)
- Maslov Iu. S.** Selected works. Aspectology. General linguistics. Moscow, LRC Publ., 2004. 846 p. (in Russ.)
- Nedialkov V. P.** The main types of initial verbs: inchoatives, ingressive, inceptive. In: *Linguistic communication and its units*. Kalinin: KSU publ., 1986, pp. 124–135. (in Russ.)
- NGLE** = Real Academia Española y Asociación de Academias de la Lengua Española: Nueva Gramática de la Lengua Española. Morfología y sintaxis. Madrid: Espasa, 2009, 4032 p.
- Paducheva E. V.** On lexical aspect and the Maslov - Vendler verbal taxonomy. In: *Voprosy Jazykoznanija*. Moscow, 2009, no. 6, pp. 3–20. (in Russ.)
- Tatevosov S. G.** Actionality in lexicon and grammar. Moscow, LRC Publ., 2015, 368 p. (in Russ.).
- Vendler Z.** Verbs and times. In: *Linguistics in philosophy*. Ithaca: Cornell University Press, 1967, pp. 97–121.
- Yllera A.** Las perífrasis verbales de gerundio y participio. In: Bosque I., Demonte V. (dirs.). *Gramática descriptiva de la lengua española*, vol. 2. Madrid: Espasa-Calpe, 1999, pp. 3391–3442.

Информация об авторах

Алыпova Светлана Алексеевна, кандидат филологических наук, преподаватель кафедры испанского языка МГИМО МИД России

Оболeнская Юлия Леонардовна, доктор филологических наук, профессор, заслуженный профессор МГУ им. М. В. Ломоносова, заведующая кафедрой иберо-романского языкознания филологического факультета МГУ им. М. В. Ломоносова

Information about the Authors

Svetlana A. Alypova, PhD in Philology, Lecturer, Department of Spanish Language, MGIMO University

Yulia L. Obolenskaya, PhD in Philology, Professor, Distinguished Professor of Lomonosov Moscow State University, Head of the Ibero-Romance Linguistics Department, Faculty of Philology, Lomonosov Moscow State University

Статья поступила в редакцию 27.01.2025;
одобрена после рецензирования 30.07.2025; принята к публикации 29.08.2025

The article was submitted 27.01.2025;
approved after reviewing 30.07.2025; accepted for publication 29.08.2025

Научная статья

УДК 81'42

DOI 10.25205/1818-7935-2026-24-2-15-28

Тактика указания на пробелы в исследованиях: эволюция жанрового маркера в эпоху конкуренции

Ольга Александровна Богинская

Иркутский национальный исследовательский технический университет
Иркутск, Россия

olgaa_boginskaya@mail.ru, <https://orcid.org/0000-0002-9738-8122>

Аннотация

На материале 200 статей, опубликованных в российских научных журналах по филологическим наукам, анализируется тактика указания на пробелы в исследованиях как структурный элемент жанра научной статьи, влияющий на композиционную организацию текста и формирование аргументативной линии. Ставится цель выявить изменения в использовании данной тактики за последние десять лет. В процессе исследования был применен комплексный подход, сочетающий анализ частотности разновидностей фокусной тактики и репрезентирующих ее лексических единиц, дискурс-анализ контекстов употребления и сравнительный анализ структурных моделей представления пробелов в разные периоды публикации. Было показано, что указание на пробелы в жанре научной статьи является динамичным процессом, и с 2020 года отмечается рост использования данной тактики (с 54 % в 2015–2019 гг. до 78 % в 2020–2024 гг.). Общая тенденция, прослеживаемая на основе полученных результатов, говорит об усилении критической направленности жанра научной статьи и осознании авторами необходимости завоевания пространства для продвижения новых исследований в условиях растущей конкуренции и влияния наукометрических систем. Было также установлено, что языковая репрезентация тактики указания на пробелы осуществляется помощью различных средств, включающих широкий диапазон лексем, принадлежащих разным частям речи, в которых было сосредоточено семантическое ядро лексико-синтаксических актуализаторов. Основными языковыми средствами актуализации тактики в исследуемом корпусе являются: лексико-синтаксические конструкции с семантическим ядром в глаголах *проводить*, *изучать*, *выступать* с отрицательной частицей *не*; лексико-синтаксические конструкции с семантическим ядром в глаголе с негативной коннотацией *отсутствовать*; лексико-синтаксические конструкции с семантическим ядром, сосредоточенном в количественных наречиях *мало*, *недостаточно*; лексико-синтаксические конструкции с семантическим ядром, сосредоточенном в прилагательных с негативной коннотацией *фрагментарный*, *размытый*; лексико-синтаксические конструкции с семантическим ядром, сосредоточенном в предикате *нет*. Было установлено, что во втором анализируемом периоде наблюдается расширение диапазона средств вербализации фокусной тактики.

Ключевые слова

академический дискурс, жанр научной статьи, пробел в исследованиях, тактика указания на пробелы

Для цитирования

Богинская О. А. Тактика указания на пробелы в исследованиях: эволюция жанрового маркера в эпоху конкуренции // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 15–28. DOI 10.25205/1818-7935-2026-24-2-15-28

Research Gap Spotting Step: Evolution of the Genre Marker in the Era of Competition

Olga A. Boginskaya

Irkutsk National Research Technical University
Irkutsk, Russian Federation

olgaa_boginskaya@mail.ru, <https://orcid.org/0000-0002-9738-8122>

Abstract

Analyzing 200 articles published in Russian philological journals over the past decade, this study examines the varieties and linguistic means of expressing the research gap spotting step as a structural element of the research article. This step influences the composition of the text and its argumentation. The purpose of the paper is to identify quantitative and qualitative changes in the use of this step, as well as the key lexemes constituting the semantic core of lexico-syntactic constructions expressing it. A mixed-method approach combining quantitative and interpretative analyses was applied to study the linguistic means implementing the step. The diachronic analysis indicates that the use of the step in research articles is a dynamic process. The explicit gap spotting step was found in 78% of articles published in 2020–2024 against 54 % of articles published in 2015–2019. The overall trend suggests enhanced critical orientation of the research articles and growing awareness of the need to conquer promotion space of new research. It was also shown that the step is verbalized using a wide range of lexemes belonging to different parts of speech, which form a semantic core of the lexico-syntactic means. The common linguistic means of actualizing the research gap spotting step in the corpus are lexico-syntactic constructions with a semantic core in the verbs *проводить*, *изучать*, *выступать* with the negative particle *не*; lexico-syntactic constructions with a semantic core in the verb *отсутствовать* with negative connotation; lexico-syntactic constructions with a semantic core in the quantitative adverbs *мало*, *недостаточно*; lexico-syntactic constructions with a semantic core in the adjectives with negative connotation *фрагментарный*, *размытый*; lexico-syntactic constructions with a semantic core in the predicate *нет*. It is concluded that in 2020–2024 there was an expansion of the lexical range and a trend to a more detailed description of the research gap, a shift towards more specific criticism in the methodology.

Keywords

academic discourse, genre of research article, research gap, gap statement tactics, academic discourse

For citation

Boginskaya O. A. Research gap spotting step: evolution of the genre marker in the era of competition. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 15–28. DOI 10.25205/1818-7935-2026-24-2-15-28

Введение

В эпоху наукометрии, когда оценка научной деятельности все больше ориентируется на формальные показатели цитируемости, публикационной активности и жесткой конкуренции, когда количество научных статей, публикуемых ежегодно, растет экспоненциально, эффективное обоснование значимости исследования, его новизны и вклада в научное знание становится критически важным. Эта тенденция, особенно актуальная в условиях информационного перенасыщения и борьбы за внимание в мировом научном сообществе, предъявляет повышенные требования к убедительности презентации результатов и обоснованию вклада каждой новой работы. Тактика указания на пробелы в предыдущих исследованиях, таким образом, превращается из риторического приема в необходимый инструмент для навигации в переполненном научном пространстве и эффективной демонстрации значимости своего вклада.

Научная статья как жанр академического дискурса служит для представления результатов исследований, формулирования новых идей и внесения вклада в развитие научной дисциплины. С этой целью авторы эксплицируют исследовательские ниши, указывают на методологические ограничения или противоречия ранее полученных результатов, другими словами, обозначают пробелы в существующем знании, чтобы позиционировать собственную работу как шаг вперед в развитии науки.

В модели жанрового анализа Дж. Суэйлза, получившей в научной литературе известность как модель CARS (Create-A-Research-Space¹), пробел в исследованиях описывается в рамках второго риторического хода ОБОЗНАЧЕНИЕ НИШИ, которому предшествует ЗАХВАТ ТЕРРИТОРИИ [Swales, 1990]. Исследование тактики указания на пробелы, начатое Дж. Суэйлзом, было продолжено в работах целого ряда зарубежных лингвистов, целью которых был преимущественно анализ способов вербализации пробелов в англоязычных научных статьях [Arsyad et al., 2023; Chen, Li, 2019; Lim, 2012; Shehzad, 2012; Suryani et al., 2015] либо сопоставление способов вербализации пробелов в статьях англоязычных и неанглоязычных авторов [Arianto et al., 2021; Arvay, Tanko, 2024; Fredrickson Swales, 1994; Taylor, Tingguan, 1991]. В отличие от неанглоязычного научного дискурса, где актуализация пробелов, как было установлено в данных исследованиях, не универсальна, в англоязычном дискурсе она является одной из частотных риторических тактик [Posteguillo, 1999].

Следует отметить, что в ходе обзора научной литературы в российском академическом контексте было найдено крайне мало статей, посвященных анализу риторических тактик, используемых в жанре научной статьи, в том числе тактике указания на пробелы в исследованиях. В работе А. А. Байбатыровой данная тактика рассматривается только на материале англоязычных статей [Байбатырова, 2020]. Ю. Н. Наumenко, исследовавший способы вербализации научных категорий исследования в англоязычных и русскоязычных медицинских статьях [Наumenко, 2022], рассмотрел фокусную тактику лишь в комплексе с другими риторическими тактиками данного жанра. О. А. Богинской был проведен сравнительный анализ использования тактики указания на пробелы исследователями в области технических и филологических наук. В работе высказано предположение, что более высокая конкуренция в инженерной сфере стимулирует авторов к более агрессивной критике предшествующих работ для обоснования необходимости разработки новых подходов [Богинская, 2025].

Учитывая недостаточную изученность тактики указания на пробелы в русскоязычном научном дискурсе, целью настоящего исследования является выявление и описание разновидностей данной тактики, определение динамики их использования на протяжении последних десяти лет, а также установление языковых средств ее реализации. Данная цель продиктована необходимостью восполнить существующий пробел в изучении способов обоснования новизны исследования и понять, как российские ученые позиционируют свои работы в глобальном научном пространстве. Для достижения поставленной цели решаются следующие задачи:

- (1) выявить разновидности тактики указания на пробелы;
- (2) проследить динамику использования данной тактики и изменения в ее языковом оформлении в 2015–2019 гг. и 2020–2024 гг. с целью выявления основных тенденций ее применения.

Вносится вклад в изучение меняющейся композиции жанра научной статьи, включающей речевые акты, стратегии и тактики в составе жанра [Дементьев, 2024]. Анализ тактики указания на пробелы и ее эволюции позволит углубить знания о внутренних механизмах жанра научной статьи и понять, каким образом российские ученые адаптируются к требованиям современной научной коммуникации и глобального наукометрического давления. Это исследование, таким образом, является важным шагом к пониманию стратегий выживания и успеха в современной российской науке.

Теоретическая основа исследования

В данном исследовании мы опираемся на модель CARS (Create-A-Research-Space), предложенную Дж. Суэйлзом и получившую широкое признание в исследованиях академического дискурса как эффективный инструмент анализа структуры научных текстов [Swales, 1990] (рис. 1).

¹ Создай исследовательское пространство.



Рис. 1. Модель CARS

Fig. 1. CARS model

Модель описывает типовую организацию введения в научную работу, акцентируя внимание на риторические приемы, используемые автором для обоснования значимости нового исследования. В частности, модель включает три основных хода (moves): 1) захват территории исследования (establishing a territory), который предполагает представление общей темы исследования, ее важности и актуальности в контексте текущего научного знания; 2) обозначение ниши (establishing a niche): центральный для настоящего исследования ход, описывающий способы, которыми автор обосновывает необходимость проведения исследования, и включающий тактику идентификации пробела (gap indication), тактику формулирования контраргумента (counter-claiming), тактику постановку вопросов исследования (question-raising) и тактику формулирования гипотезы (hypothesis elaboration); 3) заполнение ниши (occupying the niche), предполагающее описание целей и задач исследования, методов, используемых для заполнения обозначенного пробела, и предполагаемых результатов.

Важно подчеркнуть, что в модели Дж. Суэйлза риторический ход (move) понимается как дискурсивная единица, реализующая определенную коммуникативную функцию, направленную на достижение определенной цели автора. Тактика (step), в свою очередь, представляет собой конкретный текстовый элемент, способствующий реализации выбранной риторической стратегии. Таким образом, тактики являются инструментами, с помощью которых авторы реализуют риторические ходы.

Поскольку исследователи использовали модель CARS для анализа научных статей из разных областей знания, написанных не только на английском языке, были предложены различные модификации данной модели. Так, А. Арвей и Дж. Танко, сравнивая введения к теоретическим научным статьям на английском и венгерском языках, добавили два новых шага в модель CARS, поясняя, что, в отличие от эмпирических статей, преимущественно находившихся в фокусе внимания Дж. Суэйлза, в теоретических работах Ход 1 включает иллюстративный материал, а Ход 3 часто содержит аналитические детали текущего исследования [Arvay, Tanko, 2004]. Л. Энтони обнаружил, что Ход 1 в научно-технических статьях также включает в себя определения терминов и примеры, иллюстрирующие сложные концепции [Anthony, 1999]. Несмотря на различные модификации модели CARS, версия 1990 года остается по-прежнему актуальной и широко применяемой в исследованиях риторической структуры научных текстов. Она предоставляет четкую и проверенную основу для анализа, которая позволяет выявить основные риторические ходы во введениях к научным работам.

Следует отметить, что к настоящему времени был разработан целый ряд классификаций тактик указания на пробелы в исследованиях, среди которых таксономии Дж. Лима [Lim, 2012],

К. Чена и М. Ли [Chen, Li, 2019], Л. Суриани и др. [Suryani et al., 2015], М. Арианто и Я. Басфومي [Arianto, Basthomi, 2021] и М. Арианто, Я. Басфومي и А. Н. Вулиани [Arianto et al., 2021]. В качестве методологической базы настоящего исследования была использована таксономия разновидностей тактики указания на пробелы, разработанная Дж. Лимом и включающая тактику указания на отсутствие исследований, тактику указания на недостаточность исследований, тактику указания на методологические недостатки предыдущих исследований, тактику указания на противоречия в предыдущих исследованиях и тактику указания на слабость эмпирической базы [Lim, 2012]. Таксономия Дж. Лима включает в себя наиболее частотные тактики указания на пробелы в исследованиях. Эти категории предоставляют структурированную основу для анализа, позволяющую систематически идентифицировать и классифицировать различные способы, которыми авторы обосновывают необходимость проведения собственного исследования. Кроме того, применение часто используемой в аналогичных исследованиях таксономии позволит провести в дальнейшем сопоставительный анализ результатов исследований, использующих ту же таксономию. Это способствует кумулятивному накоплению знаний в данной области.

Материал и методы исследования

Материалом исследования послужили 200 научных статей, опубликованных в десяти российских журналах по филологическим наукам, входящих в «белый» список RSCI. 100 статей были опубликованы в период с 2015 по 2019 г. и 100 статей – в период с 2020 по 2024 г. Для исследования были выбраны статьи из журналов *Вестник Новосибирского государственного университета (Серия: История и филология)*, *Вестник Новосибирского государственного университета (Серия: Лингвистика и межкультурная коммуникация)*, *Жанры речи, Terra Linguistica*, *Вестник Волгоградского государственного университета (Серия: Языкознание)*, *Вестник Томского государственного университета (Филология)*, *Научный результат. Вопросы теоретической и прикладной лингвистики*, *Медиалингвистика*, *Вестник Санкт-Петербургского университета (Язык и литература)*, *Русистика* и *Сибирский филологический журнал*, которые являются ведущими российскими изданиями в области языкознания. Все журналы имеют высокий импакт-фактор, отличаются строгой системой рецензирования и имеют четкую структуру. Публикации в данных журналах отличаются оригинальностью и методологической строгостью, что делает их ценным источником для исследования используемых в научных статьях риторических тактик.

Для анализа языковых средств, реализующих тактику указания на пробелы в исследованиях, применен комплексный подход, сочетающий количественный и интерпретативный анализ. Интерпретативный анализ позволил уточнить дискурсивные функции выявленных языковых средств. Это позволило глубже понять, как авторы научных статей используют язык для обоснования актуальности своих исследований, и расширило наши знания о риторических стратегиях научного дискурса.

Анализ, представленный в настоящей статье, включает в себя поиск и отбор статей с использованием критериев «предметная область» (филология), «язык текста» (русский) и «дата публикации» (2015–2024 гг.); распределение статей по двум группам (дата публикации: 2015–2019 гг. и 2020–2024 гг.); анализ отобранных статей с целью выявления разновидностей тактики указания на пробелы с опорой на представленную выше таксономию; установление частотности обнаруженных разновидностей тактики указания на пробелы в обеих группах; поиск лексико-синтаксических конструкций, актуализирующих искомую тактику, и лексических единиц, составляющих ядро таких конструкций; интерпретация полученных результатов; сравнение частотности использования каждой из разновидностей искомой тактики в каждой группе.

Результаты

На рис. 2 показана частотность применения разновидностей тактики указания на пробелы в 2015–2019 и 2020–2024 гг.

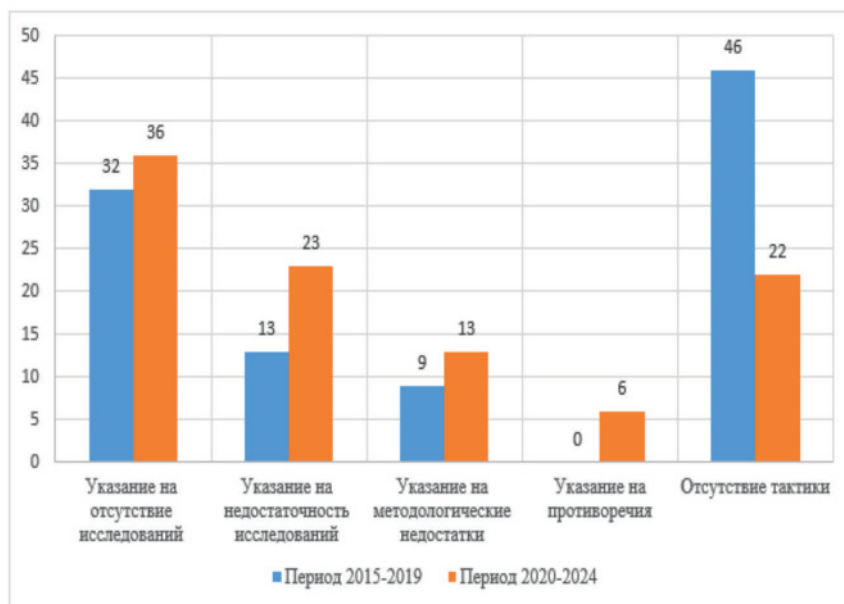


Рис. 2. Частотность разновидностей тактики указания на пробелы
Fig. 2. Frequency of the types of research gap spotting step

В ходе анализа эксплицитно выраженная тактика указания на пробелы в исследованиях была обнаружена в 78 % статей, опубликованных в 2020–2024 гг., и в 54 % статей, опубликованных в 2015–2019 гг. Обратим внимание на отсутствие в корпусе текстов указаний на слабость эмпирической базы предыдущих исследований. Согласно результатам диахронического анализа, наблюдается тенденция к более активному использованию данной тактики. Частотность использования тактики указания на отсутствие исследований и указания на методологические недостатки предыдущих исследований увеличилась незначительно. Тактика указания на недостаточность исследований и в первом и во втором периоде сохранила второе место по частотности употребления и показала тенденцию к более существенному увеличению (с 13 до 23 употреблений). Заметный рост указывает на то, что все больше авторов акцентируют внимание на неполноте существующих исследований, и означает, что в лингвистическом дискурсивном сообществе все больше внимания уделяется критической оценке текущих знаний и поиску областей для углубления исследований. Тактика указания на противоречия была обнаружена только в статьях, опубликованных в 2020–2024 гг., что может указывать на растущую тенденцию к явной критике предыдущих исследований, большую прозрачность научного процесса и желание авторов не скрывать, а наоборот, обсуждать противоречия в полученных результатах. Самым заметным изменением является значительное снижение количества статей, в которых не указаны пробелы. Это может свидетельствовать о растущей осведомленности и необходимости завоевания исследовательского пространства.

Рассмотрим каждую из представленных в таблице разновидностей тактики указания на пробелы на примерах из корпуса.

1. Указание на отсутствие исследований в предметной области.

Количественный анализ показал, что авторы в обеих группах чаще всего обосновывали необходимость проведения нового исследования отсутствием работ в предметной области.

Данная разновидность тактики преимущественно актуализировалась с помощью отрицательных глагольных конструкций:

*Новизна связана с анализом конкретных способов языкового выражения авторской позиции в публицистике Рунета, которые до сих пор специально **не изучались** в русистике.*

В примере необходимость проведения исследования обосновывается с помощью ссылки на отсутствие работ, посвященных способам описания языкового выражения авторской позиции в публицистике Рунета. Актуализатором выступает отрицательная конструкция с глаголом «изучать». Стоит также обратить внимание на использование автором неопределенно-личной конструкции, которая несет значение обобщенности, создавая впечатление всеобщности, того, что никто в русистике и жанроведении до сих пор не занимался этими вопросами и что это не просто недосмотр отдельных исследователей, а более широкая проблема, затрагивающая всю дисциплину.

В ходе анализа было обнаружено, что семантическое ядро конструкций, актуализирующих данную разновидность тактики, преимущественно было сосредоточено в глаголах «проводить», «изучать», «исследовать» с частицей «не». Кроме того, было обнаружено несколько случаев нахождения семантического ядра в предикате «нет»:

***Исследований** богатства лексики учебников и учебных пособий по русскому языку практически **нет**.*

В корпусе статей, опубликованных в 2020–2024 гг., были зафиксированы две конструкции с предлогом *вне*:

*Однако судебное экспертное заключение как конечный продукт деятельности эксперта, как текст, обладающий, несомненно, композиционно-речевыми особенностями, до недавнего времени оставалось **вне поля зрения** лингвистики.*

Предлог «вне» создает контекст для обоснования собственного исследования автора, подчеркивая отсутствие внимания со стороны академического сообщества к важной области исследования.

В статьях из второй группы также дважды встретилось указание на отсутствие исследований в предметной области с помощью конструкций, семантическое ядро которых было сосредоточено в наречии «впервые»:

*Однако вопросы оценки уровня сложности текста с помощью модели машинного обучения и представления результатов ее работы широкому кругу пользователей, а также проверки качества работы модели в реальной практике преподавания РКИ, насколько нам известно, **впервые ставятся** в данной работе.*

В следующем примере семантическое ядро лексико-синтаксической конструкции, актуализирующей тактику указания на отсутствие исследований в предметной области, сосредоточено в прилагательном «неизученный»:

*При несомненном интересе литературоведов к мифологическим мотивам и образам художественной картины мира И. А. Бунина с лингвистической точки зрения мифонимикон текстов нобелевского лауреата **представляет собой неизученный вопрос**.*

Конструкция с прилагательным встретила лишь один раз во второй группе.

2. Указание на недостаточность исследований в предметной области.

Пробел в предметной области может существовать и вследствие недостаточности работ и отсутствия в силу этого возможности всесторонне описать объект исследования и сделать объективные выводы:

*Однако, читая труды ведущих переводоведов, замечаешь, что они **редко обращают внимание** на письменный или устный перевод на неродной язык.*

*При огромном количестве работ, посвященных изучению языковой репрезентации оскорблений, проблеме диалогичности коммуникативной ситуации оскорбления (в дальнейшем – КСО) уделяется **мало внимания**.*

В данных высказываниях содержится указание на недостаточную освещенность проблемы перевода на неродной язык и коммуникативной ситуации оскорбления. Указывая на слабую изученность данных областей, авторы пытаются обосновать необходимость проведения дальнейших исследований. Семантическое ядро данных конструкций сосредоточено в лексемах *редко* и *мало*. Приведем еще один пример.

Исследуемая тема представляется актуальной, условные конструкции вызывают значительный интерес у современных ученых, однако НРУК до настоящего времени оставались малоизученными, а на материале каталанского языка они практически не рассматривались, чем и определяется новизна настоящего исследования.

Автор начинает с общего утверждения об актуальности темы, что сразу же позиционирует исследование в рамках важной и интересной для научного сообщества проблематики. Далее следует ключевая часть фрагмента, где указываются конкретные пробелы в существующем знании. Автор выделяет два важных аспекта: малоизученность НРУК и практическое отсутствие исследований на материале каталанского языка. С помощью первой тактики (указание на недостаточность исследований) автор создает впечатление, что существует общая нехватка исследований в данной области, а вторая тактика (указание на отсутствие исследований) значительно сужает область исследований и подчеркивает, что данная конкретная языковая группа практически не рассматривалась в контексте изучаемой темы. В первом случае актуализатором тактики выступает прилагательное *малоизученный*, во втором – глагол с отрицательной частицей *не* и наречием степени *практически*.

Интерес представляют конструкции с семантическим ядром, сосредоточенном в существительных, которые не содержат эксплицитного указания на недостаточность исследований, однако подчеркивают необходимость продолжения изучения того или иного явления или процесса:

Несмотря на многочисленные исследования в области лингводидактического тестирования, продолжается поиск эффективных методик подготовки к прохождению иностранными учащимися теста какого-либо уровня. В то же время наблюдения, сделанные исследователями на разных этапах развития постсоветских СМИ, требуют постоянного обновления в силу непрерывности динамики развития СМИ, реагирования языковых систем медиакоммуникации на меняющиеся условия производства новостей.

Фокусная тактика реализуется с помощью лексико-синтаксических конструкций, семантическое ядро которых сосредоточено в существительных *поиск* и *обновление*, которые имплицитно указывают на недостаточность исследований, посвященных подготовке к тестированию и языковым системам медиакоммуникации. В следующем примере семантическое ядро лексико-синтаксической конструкции находится в прилагательном *единичный*, имеющим негативную коннотацию:

Проблемы предметно-языкового интегрированного обучения иностранцев рассматриваются в единичных статьях российских методистов.

Следует отметить, что данные конструкции были достаточно редкими в исследуемых группах. Преимущественно тактика указания на недостаточность исследований актуализировалась с помощью конструкций с семантическим ядром, сосредоточенном в наречиях «мало», «недостаточно», «в меньшей степени».

3. Указание на методологические недостатки предыдущих исследований.

Для обоснования необходимости проведения нового исследования авторы также ссылались на методологические недостатки, выявленные в работах других авторов. Среди ограниченных указывались неверный выбор метода исследования, размытость целей и задач, прикладной или излишне теоретизированный характер исследования:

Тем не менее работы, посвященные исследованию аналитической статьи, дают весьма фрагментарное представление о данном явлении.

Таким образом, в настоящее время границы класса предлогов остаются размытыми.

Тем не менее необходимо отметить **не вполне сложившуюся** систему терминов, которыми пользуются лингвисты, занимающиеся теорией поликодового текста.

В первом высказывании автор критикует фрагментарность результатов, полученных коллегами в предыдущих исследованиях, указывая на необходимость более полного описания изучаемого явления. Тактика реализуется за счет использования лексико-синтаксической конструкции, включающей глагол в активном залоге, за которым следует именное словосочетание. Сематическое ядро конструкции сосредоточено в лексеме «фрагментарный» в значении «отрывочный, неполный»². В следующем примере конструкция «остаются размытыми» позволяет автору обосновать необходимость дальнейшего исследования класса предлогов. Семантическое ядро здесь находится в прилагательном «размытый» в значении «неясный, нечеткий»³. В примере автор указывает на такой недостаток, как отсутствие сложившейся системы терминов в предыдущих исследованиях. С целью актуализации тактики идентификации пробелов используется лексико-синтаксическая конструкция с ядром, сосредоточенным в наречии с частицей *не*, смягчающим иллокутивную силу критического высказывания.

4. Указание на расхождения в результатах, полученных в предыдущих исследованиях.

Примеры данной разновидности тактики являются самыми малочисленными в исследуемом корпусе и встречаются только в статьях, опубликованных в 2020–2024 гг. Шесть раз необходимость проведения нового исследования обосновывалась путем указания на отсутствие единого взгляда на предмет изучения, отсутствие общепринятой методики или разночтений в определениях тех иных понятий:

Несмотря на достаточно большое число публикаций на эту крайне актуальную для вузовской практики преподавания РКИ тему и разнообразие предлагаемых в литературе подходов, до сих пор не существует общепринятой эффективной методики обучения РКИ иностранных граждан с национальными дипломами бакалавра в создавшихся условиях.

Как видим, в современной русистике не существует единого взгляда на функции языковых афоризмов.

Единого мнения о составе и количестве базовых цветообозначений в казахском языке до сих пор нет.

Первые две конструкции являются отрицательными с семантическим ядром, сосредоточенном в прилагательных *общепринятый* и *единый*, выражающих свойства, отрицаемые с помощью частицы *не*. В третьей конструкции свойство «единство» отрицается с помощью предиката *нет*.

В таблице приведены наиболее частотные в каждой группе лексико-синтаксические конструкции, которые использовались авторами для актуализации пробелов в исследованиях.

Лексико-синтаксические конструкции, использующиеся для актуализации пробелов в исследованиях в 2015–2019 гг. и 2020–2024 гг.

Lexico-syntactic constructions used for gap spotting in 2015–2019 and 2020–2024

Разновидность тактики	Период	Лексико-синтаксические конструкции
1	2	3
Указание на отсутствие исследований	2015–2019	<i>Отсутствуют работы</i>
		<i>Исследований нет</i>
	2020–2024	<i>Рассмотрение / исследование не проводилось</i>
		<i>Отсутствуют работы/исследования</i>
		<i>Исследований/работ нет</i>
		<i>Не выступали объектом исследования/изучения</i>

² Толковый словарь русского языка / под ред. Д. Н. Ушакова. URL: <https://lexicography.online/explanatory/usakov/ф/фрагментарный> (дата обращения: 01.02.2024).

³ Ожегов С. И., Шведова Н. Ю. Толковый словарь русского языка. М.: Русский язык, 2007. 944 с.

Окончание табл.

1	2	3
Указание на недостаток исследований	2015–2019	<i>Является малоизученной</i>
		<i>Изучена недостаточно</i>
		<i>Мало обращают внимания</i>
	2020–2024	<i>Мало исследованы</i>
		<i>В меньшей степени изучена</i>
		<i>Изучена недостаточно полно</i>
Указание на методологические недостатки	2015–2019	<i>Исследования немногочисленны</i>
		<i>Рассматривают широко</i>
	2020–2024	<i>Нет четкого представления</i>
		<i>Давать фрагментарное представление</i>
		<i>Оставаться размытыми</i>
		<i>Носить сугубо описательный / прикладной характер</i>
Указание на противоречия	2020–2024	<i>Не существует общепринятой методики / единого взгляда</i>
		<i>По-разному определяются</i>
		<i>Единого мнения нет</i>

Заключение

Целью настоящего исследования является лингвистический анализ лексико-синтаксических конструкций, участвующих в реализации тактики указания на пробелы в жанре научной статьи. Было выявлено, что данная тактика использовалась большинством авторов статей, опубликованных в 2015–2019 гг. и 2020–2024 гг. Наиболее частотными в исследуемом корпусе оказались такие разновидности исследуемой тактики, как указание на отсутствие исследований и их недостаточность в предметной области. В 46 % статей, опубликованных в 2015–2019 гг. и 22 % статей, опубликованных в 2020–2024 гг., данная тактика не эксплицировалась.

Результаты анализа указывают на то, что использование тактики указания на пробелы в научных статьях является динамичным процессом, подверженным изменениям. Общая тенденция, наблюдаемая на основе полученных результатов, говорит об усилении критической направленности жанра научной статьи и осознании необходимости завоевания пространства для продвижения новых исследований. Среди причин изменения частотности употребления тактики указания на пробелы можно указать рост числа научных публикаций, который приводит к усилению конкуренции в научной среде и, как следствие, к более тщательному обоснованию новизны и значимости исследований, а также наблюдающуюся тенденцию к стандартизации структуры научной статьи вследствие процессов цифровизации, о которых говорит В. В. Дементьев, отмечая, что создание научных текстов в соответствии с требованиями наукометрических систем приводит к трансформации параметров жанра [Дементьев, 2020]. Данная трансформация подразумевает упорядочение структурных требований к статьям, что приводит к более четкому пониманию ожиданий, предъявляемых к авторам при обосновании актуальности их исследований. В контексте тактики указания на пробелы трансформация означает, что авторы стали более осознанно и последовательно использовать ее, чтобы соответствовать представлениям о том, как должна выглядеть статья, претендующая на индексацию в автори-

тетных наукометрических базах данных. Другими словами, акцентирование пробелов превращается в своего рода «жанровый маркер», позволяющий статье считаться соответствующей требованиям наукометрической валидности.

Языковая репрезентация тактики указания на пробелы осуществляется с помощью различных средств, включающих широкий диапазон лексем, принадлежащих разным частям речи, в которых было сосредоточено семантическое ядро лексико-синтаксических актуализаторов. Основными языковыми средствами актуализации тактики в исследуемом корпусе являются:

- а) лексико-синтаксические конструкции с семантическим ядром в глаголах *проводить*, *изучать*, *выступить* с отрицательной частицей *не*;
- б) лексико-синтаксические конструкции с семантическим ядром в глаголе с негативной коннотацией *отсутствовать*;
- в) лексико-синтаксические конструкции с семантическим ядром, сосредоточенном в количественных наречиях (*мало*, *недостаточно* и т. д.);
- г) лексико-синтаксические конструкции с семантическим ядром, сосредоточенном в прилагательных с негативной коннотацией *фрагментарный*, *размытый* – *иметь фрагментарный характер*, *оставаться размытыми* и т. д.;
- д) лексико-синтаксические конструкции с семантическим ядром, сосредоточенном в предикате *нет*.

Выявление ключевых лексем, формирующих семантическое ядро соответствующих конструкций, свидетельствует о том, что жанр научной статьи подразумевает определенный набор языковых ожиданий, связанных с обоснованием новизны исследования. Эти ожидания, формирующие аргументацию, во многом обусловлены необходимостью соответствия формальным критериям оценки, принятым в современной научной среде. Таким образом, исследование позволяет глубже понять, как стандартизация, связанная с эпохой наукометрии, посредством языка конструирует жанр научной статьи, и как, в свою очередь, эта стандартизация влияет на формирование аргументативной линии и обоснование значимости научного вклада. Актуализация пробелов, следовательно, является не только риторической тактикой, но и индикатором приверженности автора к доминирующим жанровым конвенциям, сложившимся под влиянием требований к публикации научных результатов.

В качестве перспективы исследования следует указать на сравнительное изучение данной тактики и ее языковой актуализации авторитетными учеными и начинающими исследователями, а также исследователями из различных областей знания. Более глубокое исследование диахронических изменений в использовании тактики указания на пробелы и языковых средств ее актуализации в жанре научной статьи также может представлять интерес. Будущие исследования могут быть направлены и на изучение влияния наукометрических требований на другие аспекты жанра научной статьи и на выявление новых риторических стратегий, используемых авторами для достижения успеха в современной научной среде.

Список литературы

- Байбатырова А. А. Дискурсивные тактики и тактики структурирования методологии исследования в англоязычных научных статьях // Вестник Томского государственного университета. 2020. № 458. С. 5–15. URL: <http://doi.org/10.17223/15617793/458/1>
- Богинская О. А. Риторический шаг «указание на пробелы в предыдущих исследованиях»: разновидности и способы вербализации // Вестник Омского государственного педагогического университета. Гуманитарные исследования. 2024. № 2 (43). С. 84–88. URL: <http://doi.org/10.36809/2309-9380-2024-43-84-88>.
- Дементьев В. В. К проблеме интегрального описания речевых жанров // Жанры речи. 2024. Т. 19, № 1 (41). С. 6–22. URL: <https://doi.org/10.18500/2311-0740-2024-19-1-41-6-22>.

- Дементьев В. В.** Трансформации научных жанров в контексте стратегий наукометрии // Коммуникативные исследования. 2020. Т. 7, № 2. С. 272–282. URL: [https://doi.org/10.24147/2413-6182.2020.7\(2\).272-282](https://doi.org/10.24147/2413-6182.2020.7(2).272-282)
- Науменко Ю. Г.** Способы вербализации научных категорий исследования в англоязычном и русскоязычном академических дискурсах: обобщенные результаты // Международный журнал гуманитарных и естественных наук. 2022. № 8-1 (71). С. 112–114.
- Anthony L.** Writing research article introductions in software engineering: How accurate is a standard model? // *IEEE Transactions on Professional Communication*. 1999. Vol. 42 (1). P. 38–46.
- Arianto M. A., Basthomi Y.** The authors' research gap strategies in ELT research article introductions: Does Scopus journal quartile matter? // *Journal of Language and Linguistic Studies*. 2021. Vol. 17 (4). P. 1743–1759. URL: <https://doi.org/10.52462/jlls.127>
- Arianto M. A., Basthomi Y., Wulyani A.** Previous studies have several limitations ...: Indonesian doctoral students', Indonesian academics', and international authors' research gap strategies in ELT research article abstracts and introductions // *Journal of Language & Education*. 2021. Vol. 7 (2). P. 25–44. URL: <https://doi.org/10.17323/jle.2021>
- Arsyad S. et al.** The rhetorical structure and research gap strategies in journal article abstracts in language-related fields published in high-impact international journals // *LEARN Journal: Language Education and Acquisition Research Network*. 2023. Vol. 16 (1). P. 703–725.
- Arvay A., Tanko G.** A contrastive analysis of English and Hungarian theoretical research article introductions // *IRAL*. 2004. Vol. 42. P. 71–100.
- Boginskaya O. A.** Disciplinary influences on research gap identification // *Research Result. Theoretical and Applied Linguistics*, 2025. Vol. 11 (1). P. 50–71. URL: <http://doi.org/10.18413/2313-8912-2025-11-1-0-3>
- Chen X., Li M.** Chinese learner writers' niche establishment in the literature review chapter of theses: A diachronic perspective // *Journal of English for Academic Purposes*. 2019. Vol. 39. P. 48–58. DOI: 10.1016/j.jeap.2019.03.006
- Fredrickson K. M., Swales J. M.** Competition and discourse community: Introductions from “Nysvenska Studier” / *Text and talk in professional contexts* ed. by B.-L. Gunnarsson, P. Linell, B. Nordberg. Uppsala: ASLA, 1994. P. 9–22.
- Lim J. M.** How do writers establish research niches? A genre-based investigation into management researchers' rhetorical steps and linguistic mechanisms // *Journal of English for Academic Purposes*. 2012. Vol. 11 (3). P. 229–245. DOI: 10.1016/j.jeap.2012.05.002
- Posteguillo S.** The schematic structure of computer science research articles // *English for specific purposes*. 1999. Vol. 18 (2). P. 139–160.
- Shehzad W.** Introduction of Computer Science Research Paper: Divergence from CARS // *Kashmir Journal of Language Research*. 2012. Vol. 15 (2). P. 19–39.
- Suryani I., Yacob A., Aziz N. H.** “Indicating a research gap” in Computer Science research article introductions by non-native English writers // *Asian Social Science*. 2015. Vol. 11 (28). P. 293–302. DOI: 10.5539/ass.v11n28p293
- Swales J.** *Genre Analysis: English in academic and research settings*. Cambridge University Press, 1990. 260 p.
- Taylor G., Tingguan C.** Linguistic, cultural and subcultural issues in contrastive discourse analysis: Anglo-American and Chinese scientific texts // *Applied Linguistics*. 1991. Vol. 12. P. 365–382.

References

- Anthony L.** Writing research article introductions in software engineering: How accurate is a standard model? *IEEE Transactions on Professional Communication*, 1999, no. 42(1), pp. 38–46.

- Arianto M. A., Basthomi Y.** The authors' research gap strategies in ELT research article introductions: Does Scopus journal quartile matter? *Journal of Language and Linguistic Studies*, 2021, no. 17(4), pp. 1743–1759. URL: <https://doi.org/10.52462/jlls.127>.
- Arianto M. A., Basthomi Y., Wulyani A.** Previous studies have several limitations ...: Indonesian doctoral students', Indonesian academics', and international authors' research gap strategies in ELT research article abstracts and introductions. *Journal of Language & Education*, 2021, no. 7(2), pp. 25–44. URL: <https://doi.org/10.17323/jle.2021>
- Arsyad S. et al.** The rhetorical structure and research gap strategies in journal article abstracts in language-related fields published in high-impact international journals. *LEARN Journal: Language Education and Acquisition Research Network*, 2023, no. 16(1), pp. 703–725.
- Arvay A., Tanko G.** A contrastive analysis of English and Hungarian theoretical research article introductions. *IRAL*, 2004, no. 42, pp. 71–100.
- Baibatyrova A. A.** Rhetorical Moves and Steps of Structuring Methodology in the English Language-Based Research Articles. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*, 2020, no. 458, pp. 5–15. URL: <http://doi.org/10.17223/15617793/458/1>. (in Russ.)
- Boginskaya O. A.** Disciplinary influences on research gap identification. Research Result. *Theoretical and Applied Linguistics*, 2025, no. 11(1), pp. 50–71. URL: <http://doi.org/10.18413/2313-8912-2025-11-1-0-3>
- Chen X., Li M.** Chinese learner writers' niche establishment in the literature review chapter of theses: A diachronic perspective. *Journal of English for Academic Purposes*, 2019, no. 39(1), pp. 48–58.
- Dementyev V. V.** On the problem of integral description of speech genres. *Speech Genres*, 2024, no. 19, 1(41), pp. 6–22. URL: <https://doi.org/10.18500/2311-0740-2024-19-1-41-6-22>. (in Russ.)
- Dementyev V. V.** Transformation of scientific genres in the context of scientometric strategies. *Communicative Research*, 2020, vol. 7, no. 2, pp. 272–282. URL: [https://doi.org/10.24147/2413-6182.2020.7\(2\).272-282](https://doi.org/10.24147/2413-6182.2020.7(2).272-282) (in Russ.).
- Fredrickson K., Swales J.** Competition and discourse community: Introductions from “Nysvenska Studier”. Text and talk in professional contexts. Uppsala: ASLA, 1994, pp. 9–22.
- Lim J.** How do writers establish research niches? A genre-based investigation into management researchers' rhetorical steps and linguistic mechanisms. *Journal of English for Academic Purposes*, 2012, no. 11(3), pp. 229–245.
- Naumenko Yu. N.** Verbalization of the research categories in English and Russian academic discourses: summary of the results. *International Journal of Humanities and Natural Sciences*, 2022, no. 8-1 (71), pp. 112–114. (in Russ.)
- Posteguillo S.** The schematic structure of computer science research articles. *English for specific purposes*, 1999, no. 18(2), pp. 139–160.
- Shehzad W.** Introduction of Computer Science Research Paper: Divergence from CARS. *Kashmir Journal of Language Research*, 2012, no. 15(2), pp. 19–39.
- Suryani I., Yacob A., Aziz N.** Indicating a research gap in Computer Science research article introductions by non-native English writers. *Asian Social Science*, 2015, no. 11(28), pp. 293–302.
- Swales J.** Genre Analysis: English in academic and research settings. Cambridge University Press, 1990.
- Taylor G., Tingguan C.** Linguistic, cultural and subcultural issues in contrastive discourse analysis: Anglo-American and Chinese scientific texts. *Applied Linguistics*, 1991, no. 12, pp. 365–382.

Информация об авторе

Богинская Ольга Александровна, доктор филологических наук, профессор кафедры иностранных языков Иркутского национального исследовательского технического университета

Information about the Author

Olga A. Boginskaya, Doctor of Philology, Professor at the Department of Foreign Languages, Irkutsk National Research Technical University

*Статья поступила в редакцию 07.04.2025;
одобрена после рецензирования 22.07.2025; принята к публикации 29.08.2025*

*The article was submitted 07.04.2025;
approved after reviewing 22.07.2025; accepted for publication 29.08.2025*

Научная статья

УДК 81'1:008

DOI 10.25205/1818-7935-2026-24-2-29-41

Аксиологическая маркированность английских слов года

Антонина Вадимовна Самойлова

Кубанский государственный университет
Краснодар, Россия

antina@inbox.ru, <https://orcid.org/0000-0002-9047-0772>

Аннотация

Статья посвящена лингвоаксиологическому анализу лексем и неолексем, являющихся словами 2023 года по версии авторитетных англоязычных лексикографических изданий. Научная новизна исследования состоит в комплексной оценке сущности англоязычной лексики года по трем аспектам: семантическому, аксиологическому и прагматическому с акцентом на ценностный компонент. Рассмотрение семантики ряда лексических единиц посредством компонентного анализа позволило понять содержательную специфику с опорой на парадигматические связи. Аксиологический подход дал возможность выделить ценностно-смысловые элементы – аксиологемы, являющиеся знаками лингвокультурных концептов, имплицитно интегрированных в семантике изучаемого лексического материала и формирующих целостную картину мира. Было выделено шесть аксиологем как ценностного, так и антиценностного характера: успех, ложь, несчастье, печаль, прогресс. Помимо этого, была определена ингерентно-оценочная или адгерентно-оценочная природа вычлененных ценностных компонентов анализируемых лексем. Прагматический подход позволил оценить смысловую значимость и импликацию изучаемого лексического материала, способствующих включению некоторых ценностей в аксиосферу языка. Было выявлено, что ингерентно-оценочный компонент значения характерен для лексем с положительной или отрицательной оценочностью в семантике, в то время как лексемы остальных категорий относятся к адгерентно-оценочной лексике и проявляют аксиологическую гибкость. Лексемы с флуктуирующей оценочностью и лексемы, подразумевающие тенденцию к негативизации в рамках оценочного значения, проявляют признаки окказионально-имплицитно-оценочной лексики, что выражается в контекстуальных изменениях вектора аксиологической маркированности. Также был сделан обобщенный вывод о функциональной нагрузке слов года в качестве маркеров усредненной коллективной системы ценностей с тенденцией к превалированию позитивности восприятия.

Ключевые слова

аксиологема, узуализация, ингерентно-оценочный компонент, адгерентно-оценочный компонент, слово года

Для цитирования

Самойлова А. В. Аксиологическая маркированность английских слов года // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 29–41. DOI 10.25205/1818-7935-2026-24-2-29-41

Axiological Markedness of English Words of the Year

Antonina V. Samoylova

Kuban State University
Krasnodar, Russian Federation

antina@inbox.ru, <https://orcid.org/0000-0002-9047-0772>

Abstract

The article is devoted to the linguo-axiological analysis of lexemes and neolexemes, which are words of the year 2023 according to the authoritative lexicographical sources. The scientific novelty of the study is a comprehensive assessment of the English lexis of the year regarding three aspects: semantic, axiological, and pragmatic with the focus on the value component. The componential analysis of some lexical units allowed us to understand their essential specificity at the paradigmatic level. The axiological approach made it possible to deduce valuable semantic elements – axiologemes, regarded as signs of linguo-cultural concepts implicitly integrated into the semantics of the studied lexical material. It is believed that they form a holistic world view. As a result, six axiologemes of both valuable and “anti-valuable” character have been identified: success, lie, misfortune, sadness, progress. In addition, we identified an inherent-evaluative and an adherent-evaluative nature of the deduced value components of the analyzed lexemes. The pragmatic approach allowed us to assess the semantic prominence and implicature of the studied lexical material resulting in the inclusion of some value components in the language axiosphere. It was found that the inherent-evaluative component of meaning is characteristic of lexemes with either positive or negative evaluation semantics, while the adherent-evaluative lexis demonstrates axiological flexibility. Lexemes with fluctuating evaluation and those tending to be negative, exhibit characteristics of occasional-implicit-evaluative lexis, which manifests itself in contextual changes of axiological markedness vector. Generally speaking, the functional load of the words of the year lies in their ability to be the markers of the averaged collective system of values with a tendency towards a positive way of perception.

Keywords

axiologeme, usualization, inherent-evaluative component, adherent-evaluative component, word of the year

For citation

Samoylova A. V. Axiological markedness of English words of the year. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 29–41. DOI 10.25205/1818-7935-2026-24-2-29-41

Введение

Процесс коммуникации невозможно представить оценочно-нейтральным. Выбор лингвистического материала для данного исследования позволяет произвести общую лингвоаксиологическую оценку приоритетов медийного и социального пространства в рамках прагматического подхода, а также определить ключевые тематические ориентиры языковой картины мира англоговорящего сообщества. Эмпирический материал исследования составляет 41 лексическая единица – английские слова 2023 года.

Цель данного исследования состояла в анализе векторов аксиологической маркированности англоязычных слов 2023 года в зависимости от их базовой лексикографической или вторичной коммуникативно-обусловленной специфики использования. С целью определения прагматического потенциала аксиологического компонента значения лексем года в исследовании была представлена типология на основании контекстуального анализа употребления данных лексем в текстовом массиве англоязычных онлайн-ресурсов: Business Insider¹, Chosing

¹ Business Insider. URL: <https://www.businessinsider.com> (accessed: 19.04.2024).

Therapy.com², Roast.dating³, Science of People⁴, Town and Country Magazine⁵, VerywellMind.com⁶, Vulture⁷.

Для достижения поставленной цели был определен ряд задач:

- 1) провести анализ теоретико-методологической базы лингвоаксиологического исследования;
- 2) определить траекторию вхождения аксиологом в коммуникативное медиапространство;
- 3) выделить спектр ключевых аксиологом, составляющих элементы ценностной картины мира англоговорящего социума;
- 4) сгруппировать анализируемые лексические единицы на основании способов определения оценочных смыслов;
- 5) проанализировать англоязычные слова 2023 года на предмет их аксиологической маркированности.

Актуальность исследования вызвана значительным интересом к лингвоаксиологии как некой ступени к пониманию процессов прагматического функционирования языка, в частности, к работе механизмов экспрессивно-оценочного и когнитивно-моделирующего характера посредством социально-востребованной лексики и неолексики.

Материал исследования был взят из электронных лексикографических источников: Cambridge Dictionary⁸, Collins Dictionary⁹, Dictionary.com¹⁰, Macquarie Dictionary¹¹, Merriam-Webster Dictionary¹², Oxford Dictionary¹³, Urban Dictionary¹⁴.

Методология исследования включала применение следующих общенаучных методов: анализ, синтез и обобщение научного опыта, гипотетико-индуктивный метод сбора языкового материала и его проверки на соответствие выдвигаемым гипотезам, графический метод визуализации общих аксиологических тенденций. Частные методы исследования включали метод компонентного анализа значения рассматриваемых лексем с целью выявления специфики их коннотативно-оценочного компонента значения; описательный и таксономический методы, используемые для пояснения семантики лексических единиц и их классифицирования. Примененный метод контекстуального анализа позволил рассмотреть оценочность и узуализацию лексем на аутентичном языковом материале. Сравнительно-сопоставительный метод позволил проанализировать слова 2022 и 2023 года на предмет общего вектора аксиологической маркированности коллективного сознания.

² ChoosingTherapy.com. URL: <https://www.choosingtherapy.com/> (accessed: 05.11.2024).

³ Roast.dating. URL: <https://roast.dating/> (accessed: 20.05.2024).

⁴ Science of People. URL: <https://www.scienceofpeople.com/> (accessed: 20.05.2024).

⁵ Town and Country Magazine. URL: <https://www.townandcountrymag.com/> (accessed: 06.05.2024).

⁶ VeryWellMind.com. URL: <https://www.verywellmind.com> (accessed: 01.05.2024).

⁷ Vulture. URL: <https://www.vulture.com/> (accessed: 27.05.2024).

⁸ Cambridge Dictionary. URL: <https://www.cam.ac.uk/research/news/cambridge-dictionary-names-hallucinate-word-of-the-year-2023> (accessed: 28.04.2024).

⁹ Collins Dictionary. URL: <https://www.collinsdictionary.com/woty?ysclid=lp8tw47kpf370498637> (accessed: 30.04.2024).

¹⁰ Dictionary.com. URL: <https://content.dictionary.com/word-of-the-year-2023/#lp-pom-text-60> (accessed: 05.05.2024).

¹¹ Macquarie Dictionary. URL: <https://www.macquariedictionary.com.au/blog/article/913/> (accessed: 06.05.2024).

¹² Merriam-Webster Dictionary. URL: <https://www.merriam-webster.com/wordplay/word-of-the-year> (accessed: 08.05.2024).

¹³ Oxford Dictionary. URL: <https://languages.oup.com/word-of-the-year/2023/> (accessed: 11.05.2024).

¹⁴ Urban Dictionary. URL: <https://www.urbandictionary.com> (accessed: 11.04.2024).

Основная часть

От оценки к аксиологеме: реализация ценностного потенциала

Выявление слов года является традиционной процедурой в рамках проведения комплексного исследования интересующих массмедийную аудиторию явлений и обозначающих их лексических единиц. В качестве опосредованной цели можно также считать сопутствующее аналитическое исследование ключевых векторов социального спроса, а также характеристику аксиосферы общедоступного информационного пространства через циркулирующую в его плоскости знаковую лексику.

Проблемами лингвоаксиологии занимались многие лингвисты, в частности, вопросам ценностного ориентирования посвящены работы Г. Ф. Гибатовой [Гибатова, 2011], Н. А. Купиной [Купина, 2021, 2022], Н. Г. Наумовой [Наумова, 2016]; работы В. В. Катерминой [Катермина, 2023], А. И. Лызлова [Лызлов, 2020], Kozhakhmetova [Kozhakhmetova et al., 2020], Wlodarski [Wlodarski, 2008] отражают изучение оценочного компонента в разных дискурсивных практиках; лексикологические исследования с акцентом на ценностную проблематику были представлены О. В. Казаченко [Казаченко, 2020], В. М. Мокиенко, Т. Г. Никитиной [Мокиенко, Никитина, 2022], А. В. Самойловой, А. В. Элмасын [Самойлова, Элмасын, 2023].

Несмотря на то что оценочная деятельность в значительной степени субъективна, именно «категория языковой оценки является основным способом отражения системы ценностей в языке» [Гибатова, 2011. С. 238]. Ценностная картина мира представляет собой сложную многоуровневую совокупность ценностных смыслов, верхним уровнем которой будут являться коллективные ценности. Система ценностей является динамической структурой; изменения в национальной психологии, культуре или социальной жизни неизменно влекут за собой ее реорганизацию [Schmitt, Rodgers, 2020]. Так, отмечаемый на современном этапе развития общества трансформационный переход от Homo sapiens (человека разумного) к Homo digitalis (человеку цифровому) [Черкасова, 2023] сопровождается постепенной эволюцией образа мышления и мутациями аксиосферы. Для «цифровизации» мышления характерна проекция аксиологем лингвокультуры в медиапространство с дальнейшей адаптацией оценочного смысла под запросы «цифрового социума». «Под аксиологемой понимается знак, репрезентирующий особо значимый культурный концепт лингвокультуры, составляющий элемент ценностного измерения картины мира или фрагмента картины мира» [Викулова и др., 2021. С. 22].

Как отмечают Л. Г. Викулова и др., семиотика аксиологемы предполагает управление интерпретационным процессом через активизацию ядерного яркого образа и ассоциируемой с ним совокупности признаков универсального для социума знания [Викулова и др., 2021]. Являясь элементом ценностного кода культуры, аксиологема задает код восприятия и осмысления контекста, в котором она фигурирует, а также модифицирует оценочный смысл, заложенный в когнитивной картине мира индивида, ее использующего. Выбор номинативной единицы для языковой актуализации аксиологемы остается за участником коммуникативного процесса. После ее многократного использования происходит процесс узуализации, т. е. вхождения данной номинативной единицы в узус, во всеобщее пользование. «Узуализация означает приобретение номинативной единицей свойства воспроизводимости» [Голев, 2001. С. 94]. Используемая в речевой деятельности узуальная номинация или спектр номинаций становятся лексическими понятиями, которые фиксируются в языковом сознании носителей языка и предопределяют их контекстуальное употребление. Узуализация аксиологемы посредством использования обозначающих ее лексических единиц достигает максимума при попадании в медийное пространство, параллельно оказывая манипулятивно-моделирующее воздействие на участников медиапространства и их взаимоотношения. Возвращаясь к тезису о субъективности оценочной деятельности, следует отметить, что вектор оценочной маркированности аксиологем так же, как и номинирующих их лексических единиц, может ситуативно варьироваться в зависимости от предопределяющих социальную реальность факторов.

Выбранный в рамках данного исследования лингвистический материал дает возможность рассмотреть актуализированные через знаковую лексику года оценочные векторы наряду с ключевыми аксиологемами, фигурирующими в медиапространстве. По итогам проведенного компонентного анализа семантики слов 2023 года нами были выделены следующие аксиологемы как ценностного, так и антиценностного характера, составляющие семные оценочные компоненты лингвистической оболочки слов и формирующие когнитивный ценностный код англоговорящего социума: успех (*coronation, EGOT, GOAT, nepo-baby, rizz, swiftie*), истинность (*authentic, covenant, canon event, wokeism*), ложь (обман/ошибка/ошибочность) (*deadname, de-banking, deepfake, de-influencing, Doppelgänger, Dystopian, greedflation, grifter, hallucinate, parasocial, situationship*), несчастье (беда/конфликт/насилие) (*bazball, blue-sky flood, implosion (implode), indict (indicted), kibbutz, strike, wildfire*), печаль (обреченность) (*beige flag, ennui*), прогресс (*AI, algospeak, elemental, generative AI, heat dome, prompt, semaglutide, ULEZ, X*).

Рассматривая возможности языковых единиц реализовывать аксиологический потенциал, В. П. Новиков различал следующие виды оценки:

- 1) узуальная или закрепленная за определенным лексико-семантическим вариантом языковой единицы в соответствующем коммуникативном контексте; окказиональная – приобретаемая в узконаправленном контексте;
- 2) системно-языковая оценка в форме либо закрепленной узуально-оценочной информации, либо в форме узуальной квазиоценки, т. е. не имеющей вербального выражения в системе языка в целом;
- 3) эксплицитная оценка, содержащаяся в лексикографической статье, описывающей семантику соответствующей лексической единицы; имплицитная оценка – подразумеваемая, вербально не выраженная, но контекстуально (окказионально) приобретаемая оценочная информация [Новиков, 1992. С. 6].

Выделяют два основных способа определения оценочных смыслов: ингерентно-оценочный и адгерентно-оценочный [Ретунская, 1996; Солдатова, 2019]. Ингерентно-оценочный компонент изначально интегрирован в семантическое значение анализируемой лексической единицы, что реализуется в стабильности ее оценочного значения вне зависимости от контекстуального использования. Адгерентно-оценочный компонент заключается в способности создания соответствующей оценки на тематически и коммуникативно-дифференцированном контекстуальном уровне. Адгерентно-оценочная лексика делится на узуально-имплицитно-оценочную лексику и окказионально-имплицитно-оценочную лексику [Солдатова, 2019. С. 70]. Первый тип предполагает сильный импликационал и частоту реализации аксиологического потенциала при контекстуальном использовании, в то время как второй тип воплощает аксиологический потенциал в менее частотных или узконаправленных контекстуальных условиях. Анализ англоязычных слов 2023 года в большей степени предполагает узуально-имплицитно-оценочную направленность, что объясняется самим выбором анализируемого лингвистического материала – его популяризацией, интересом к семантике отдельно взятых лексических единиц, а также обширным использованием в рамках медиапространства.

Аксиологическая типология лексики года

С целью анализа прагматической реализации аксиологического потенциала англоязычных слов 2023 года нами были выделены пять групп на основании анализа ингерентно- и адгерентно-представленной аксиологической маркированности. Оценка вектора аксиологической направленности производилась на основании контекстуального анализа узуально-имплицитно-оценочного компонента коннотативного значения рассматриваемых лексем в текстовом массиве англоязычных онлайн ресурсов:

- лексемы с *положительной оценочностью* в семантике (на основании базовой словарной дефиниции), коннотативный компонент значения которых предполагает наличие аксиологической маркированности в виде ингерентно-оценочного компонента, интегрированного в саму семантику данных единиц: *rizz* – сокр. харизма, аббревиатур *GOAT* (*Greatest of all times*), *EGOT* (*the Emmy + the Grammy + the Oscar + the Tony*). Отличительной чертой данной группы является ее малочисленность. Стоит отметить факт признания лексемы *rizz* в качестве слова 2023 года Oxford Dictionary в противовес подавляющему большинству негативно-окрашенной лексики, включенной в развернутые перечни слов года другими словарями. Неосленгизм *rizz* образован комбинацией двух типов словообразовательного усечения афerezиса и апокопы (усечение начала и конца слова *charisma*), а также удвоением буквы *z* по ономотопеическому принципу. *Rizz* – someone’s ability to attract another person through style, charm, or attractiveness [Oxford Dictionary]. Лексема также подлежит конверсии и может использоваться в глагольной форме *to rizz up* – to attract someone’s attention and make them like you in a romantic or sexual way, or to try to do this by behaving in a particular way [Cambridge Dictionary]. Рассмотрим некоторые случаи употребления данной лексической единицы:

When done correctly, rizzing up is all you need to impress your crush and unlock future romantic possibilities [Bailey, 2024].

No worries! Here’s how to rizz a girl or guy like a pro! [Hwang, 2024].

Положительная аксиологическая маркированность в комбинации с когнитивной релевантностью, простотой словоформы, а также неологической природой обусловили значительный интерес пользователей медиапространства к данной лексеме и высокую частотность ее употребления.

Аббревиатуры-акронимы *EGOT* и *GOAT* также могут быть рассмотрены как обособленные, зафиксированные в лексикографических изданиях лексические единицы [Арнольд, 1986. С. 143], обладающие планом содержания и планом выражения и выполняющие номинативную и атрибутивную функции. При наличии положительной аксиологической направленности данные лексические единицы тем не менее проявляют оценочность по-разному. Аббревиатура *GOAT* содержит эксплицитно выраженный ингерентно-оценочный компонент значения *greatest*, что расширяет узус ее применения вне зависимости от контекста, в то время как аббревиатура *EGOT* содержит узуальную имплицитную квазиоценку в рамках тематически очерченной дискурсивной области искусства и культуры; акроним может выполнять как номинативную, так и дискриптивную функции:

Elton John just became the next member of entertainment’s most exclusive club, the EGOT winners [Hubbard, 2024].

21 people who have won an EGOT, the most coveted award in Hollywood [Singh, Haasch, 2024];

- лексемы *нейтрально-оценочного характера*: *prompt*, *AI*, *semaglutide*, *covenant*, *elemental*, *cozzie livs*, *heat dome*, *blue-sky flood*, *algospeak*. Лексические единицы данной группы характеризуются отсутствием компонента оценочности в семантике и преобладанием ориентации на предметную составляющую или на смысловое наполнение [Самойлова, Шершнева, 2023]. Нейтральный оценочный статус расширяет узус использования и обеспечивает стабильность понимания универсального семантического значения в разных контекстуальных средах. В данную группу включены две лексические единицы, признанные главными словами года по версии Collins Dictionary – *AI* – artificial intelligence or artificial insemination: the modeling of human mental functions by computer programs, а также Macquarie Dictionary – *cozzie livs* – humorous play on cost of living;

- лексемы с *флуктуирующей оценочностью*: *swiftie*, *de-influencing*, *canon event*, *parasocial*, аббревиатура *ULEZ*, *authentic*, *coronation*, *X*, *generative AI*, *wokeism*. Рассмотрим некоторые из них более подробно. Так, лексическая единица *swiftie* – an enthusiastic fan of the singer Taylor Swift – была включена в список финалистов-слов года по версии Oxford Dictionary. В 2023 году внимание англоязычной аудитории было сосредоточено на американской певице Тейлор Свифт, ее концертном туре и предшествующей ему промокампании. Этимологически лексема представляет собой дериват, образованный на основе фамилии певицы и уменьшительно-ласкательного суффикса *-ie* с семантическим элементом, обозначающим привязанность, предрасположенность к объекту внимания. С позиции лингвоаксиологии лексема не содержит очевидных оценочных компонентов в семантике и по большей степени выполняет номинативную функцию. Тем не менее анализ ряда англоязычных информационных ресурсов показал двойственную природу оценки самой реалии, что сделало ее включение в группу лексем с флуктуирующей оценочностью обоснованным:

A swiftie is a person <...> who supports Taylor Swift in any way shape or form. <...> Swiftie is its own definition, because it's the best fan base in the world – положительная окказиональная оценочная импликация.

A kind of obsession and mental health problem. A swiftie always believes their idol is the best and no one would ever top her. They drag everyone but when it comes to themselves they act as innocent – отрицательная окказиональная оценочная импликация [Urban Dictionary].

Флуктуирующая аксиологическая маркированность характерна и для следующей лексической единицы, признанной лексикографами словаря Merriam-Webster Dictionary главным словом 2023 года. Исходя из наших наблюдений, лексическая единица *authentic* имеет тенденцию к изменению оценочного значения в зависимости от интенсивности и частоты ее использования. Семантическое значение с изначально положительной оценочностью – *authentic* – not false or imitation, a synonym of *real and actual*; true to one's own personality, spirit, or character [Merriam-Webster Dictionary], предполагающее оригинальность и неповторимость создаваемого продукта/материала, приобретает некоторую аксиологическую двоякость ввиду причин экстралингвистического характера. Развитие искусственного интеллекта, его возможности создания качественных дипфейков ставит под сомнение само понятие подлинности в самых разных областях применения: от дискурсивной области культуры и искусства до академического дискурса;

- лексемы *отрицательно-оценочного характера*: *hallucinate*, *implosion (implode)*, *ennui*, *grifter*, *nepo-baby*, *deepfake*, *Dystopian*, *deadname*, *bazball*, *de-banking*, *greedflation*, *indict (indicted)*, *strike*, *wildfire*. Отрицательная оценочность в семантике подавляющего большинства лексем данной группы выражена эксплицитно при помощи конкретных аксиологических маркеров, содержащихся в словарной дефиниции, и иллюстрирует ингерентно-оценочный способ проявления аксиологической маркированности:

to hallucinate – when an artificial intelligence hallucinates it produces *false* information [Cambridge Dictionary];

ennui – a feeling of being *bored and mentally tired* caused by having *nothing interesting or exciting* to do;

grifter – someone who gets money *dishonestly* by *tricking* people [Cambridge Dictionary];

deepfake – an image or recording that has been convincingly altered and *manipulated* to *misrepresent* someone as doing or saying something that was not actually done or said [Merriam-Webster Dictionary].

Лексема *nepo-baby* представляет собой пример единицы, непосредственное значение которой на первый взгляд не содержит эксплицитно выраженных семантических элементов оценочности, что характерно для адгерентно-оценочной маркированности: a person, esp. in the entertainment industry, whose career is believed to have been advanced by having a famous parent [Collins Dictionary]. Тем не менее аффикс *nepo-* является сокращением лексической единицы *nepotism* – the act of *using your power or influence to get ... unfair advantages* for members of your own family [Cambridge Dictionary], содержащей наглядные семантические маркеры отрицательной оценочности. При этом лексема может быть отнесена к узואльно-имплицитно-оценочной лексике на основании сильного импликационала, реализуемого в разных коммуникативно-речевых контекстах:

The celebrity-media complex allows Brooklyn Beckham ... to headline *Variety's* "Young Hollywood" issue without ever approaching anything you or I would recognize as a normal job ... a nepo baby is physical proof that meritocracy is a lie [Jones, 2024].

Исключение в данной группе составляет лексическая единица *ultra-processed*, чей способ определения оценочного смысла отличается от представленных выше примеров. *Ultra-processed* – (of food) prepared using complex industrial methods from multiple ingredients, often including ingredients with little or no nutritional value [Collins Dictionary]. Лексема имеет имплицативную, выведенную на основании дедуктивного подхода отрицательную оценку, напрямую не выраженную в словарной дефиниции, т. е. проявляет аксиологическую маркированность посредством адгерентно-оценочного способа в рамках узואльно-имплицитно-оценочного значения. На наш взгляд, ошибочно включать данную лексему в группу лексем нейтрально-оценочного характера ввиду того, что ее прагматическое значение определяется рефлексивно – лексема обозначает продукцию, содержащую минимальную питательную ценность;

- лексемы, подразумевающие *тенденцию к негативизации* в рамках оценочного значения.

Выделение данной группы в качестве отдельной категории видится целесообразным по причине того, что флуктуационные оценочно-семантические процессы в данной группе (в отличие от лексических единиц, определенных в группу с флуктуирующей оценочностью) носят константный односторонний характер, т. е. флуктуаций контекстуального использования лексем в направлении позитивизации отмечено не было. Под позитивизацией мы понимаем «создание позитивной эмоциональной окраски у информации, <...> превалирующую позитивность восприятия» [Комлев, 2021. С. 51]. Аксиологическая маркированность с тенденцией к негативизации характерна для следующих лексических единиц: *situationship*, *beige flag*, *Doppelgänger*, *kibbutz*, *deadname*. Рассмотрим одну из них более подробно.

Situationship – a romantic or sexual relationship that is not considered to be formal or established [Oxford Dictionary]. Поскольку термин был популяризирован англоязычными социальными медиа, он вошел в шорт-лист слов года, представленный словарем Oxford Dictionary. Этимологически лексическая единица представляет собой бленд лексем *situation* и *relationship*; ее тематическая принадлежность в некотором роде предопределяет присутствие оценочной маркированности в семантике. Тем не менее лексему можно отнести к категории адгерентно-оценочной лексики, реализующей свой аксиологический потенциал окказионально-имплицитно-оценочным способом, что доказывают следующие цитаты:

The biggest advantage of a situationship is that there is less responsibility. ... It can offer a fun, stress-free way to enjoy the benefits of a relationship without a major emotional commitment, provided both partners are on the same page [Tzall, 2024].

A situationship is a romantic relationship <...> However it can cause uncertainty, anxiety, and confusion about the future of relations [Tzall, 2024].

Компаративный анализ контекстуального материала, включающего перечень аргументов против вступления в отношения, обозначаемые номинативной лексической единицей *situationship* (uncertainty, anxiety, confusion), подтверждает обоснованность ее включения в группу слов года, подразумевающих тенденцию к негативизации в рамках оценочного значения.

Заключение

Прагматический и лингвоаксиологический анализ лексем позволил прийти к ряду выводов:

1) подавляющее большинство лексем с положительной и отрицательной оценочностью в семантике реализуют свой аксиологический потенциал в рамках ингерентно-оценочного компонента значения, в то время как лексемы нейтрально-оценочного характера, лексемы с флуктуирующей оценочностью либо подразумевающие тенденцию к негативизации, относятся к адгерентно-оценочной лексике и проявляют аксиологическую гибкость в дифференцированных контекстуальных условиях.

2) лексемы с флуктуирующей оценочностью, а также лексемы, подразумевающие тенденцию к негативизации в рамках оценочного значения, проявляют признаки окказионально-имплицитно-оценочной лексики, изменяя вектор аксиологической маркированности в зависимости от узконаправленного контекста.

3) большинство лексем социально рефлекторны, т. е. номинируют явления, понятия, объекты, связанные с жизнью общества и корреляциями его представителей, а также отражают усредненную коллективную аксиологическую характеристику. В целом лексемы года рассматривались в качестве источника выявления социокультурных взглядов и ценностных доминант как носителей языка, принадлежащих к англо-американскому обществу, так и коммуникантов иной этнической принадлежности, использующих английский язык в качестве средства общения. При этом стоит отметить, что крайне субъективная, флуктуирующая природа оценки может обуславливать вариативность ценностного потенциала той или иной лексемы в зависимости от национально-культурной аксиосферы.

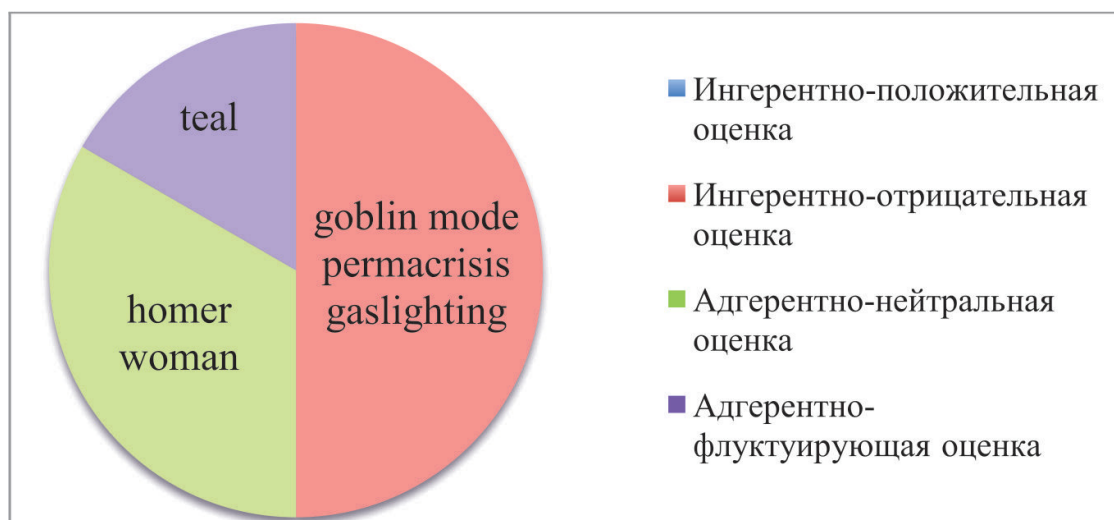


Рис. 1. Аксиологическая маркированность английских слов 2022 года

Fig. 1. Axiological markedness of the English words for the year 2022

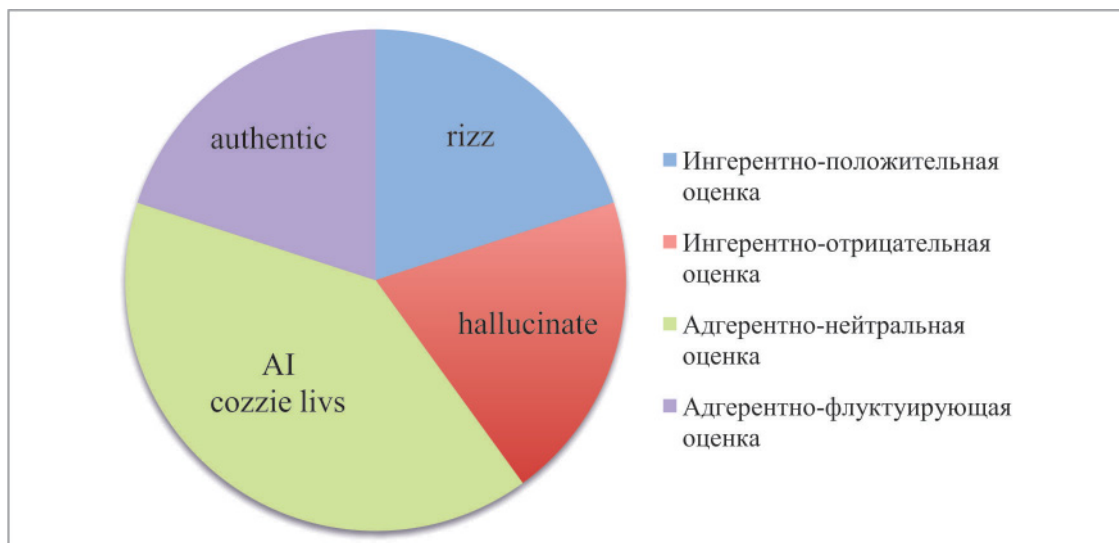


Рис. 2. Аксиологическая маркированность английских слов 2023 года

Fig. 2. Axiological markedness of the English words for the year 2023

4) сравнительно-сопоставительный анализ аксиологической маркированности слов 2022 [Самойлова, Шершнева, 2023] и 2023 годов выявил некоторую тенденцию превалирования позитивности восприятия. Ввиду напряженной общественно-политической обстановки в мире 2022 год запомнился преимущественно лексемами года с ингерентно-оценочным отрицательно-маркированным компонентом значений: *goblin mode* (Oxford Dictionary), *permacrisis* (Collins Dictionary), *gaslighting* (Merriam-Webster Dictionary) (рис. 1), в то время как перечень слов 2023 года включал лексемы с положительным ингерентно-оценочным компонентом: *rizz* (Oxford Dictionary); нейтральным адгерентно-оценочным компонентом: *AI* (Collins Dictionary), *cozzie livs* (Macquaire Dictionary); флуктуирующим адгерентно-оценочным компонентом: *authentic* (Merriam-Webster Dictionary) (рис. 2). Единственной ингерентно-отрицательно-маркированной лексемой стал глагол *to hallucinate* (Cambridge Dictionary, Dictionary.com).

В целом стоит признать статистическое преобладание лексем с отрицательной оценочностью в полных шорт-листах слов 2022 и 2023 годов. Однако факт некоторого сокращения количества отрицательно-маркированной лексики, как и само признание лексемы с положительной оценочностью *rizz* словом года по версии Oxford Dictionary и ее фигурирование в шорт-листах слов года Dictionary.com и Merriam-Webster Dictionary, на наш взгляд, отражает когнитивную гибкость коллективного сознания с тенденцией к позитивному или относительно аналитически нейтральному оцениванию явлений внеязыковой действительности.

Список литературы

- Арнольд И. В. Лексикология современного английского языка: 3-е изд., перераб. и доп. М.: Высш. шк., 1986. 295 с.
- Викуллова Л. Г., Серебренникова Е. Ф., Черкашина Е. И. Аксиологема *château* как код продвигающего текста // Маркетинговая лингвистика в цифровую эпоху: сб. науч. ст. М.: Языки народов мира, 2021. С. 22–30.
- Гибатова Г. Ф. Аксиология в языке // Вестник ОГУ. 2011. № 2 (121). С. 127–132.
- Голев Н. Д. Стихийная узуализация номинативных единиц // Известия УрГУ. 2001. № 21. С. 94–101.

- Казаченко О. В. Изучение слов, выражающих ценности в различных лингвистических дисциплинах // Психолингвистические аспекты изучения речевой деятельности. 2020. № 18. С. 168–175.
- Катермина В. В. Аксиологический потенциал англоязычных неологизмов в дискурсивных практиках. Краснодар, 2023. 192 с.
- Комлев И. В. Открытие силы суперпозиции жизни // Психологические науки. 2021. С. 47–63.
- Купина Н. А. Аксиология и аксиологический лексикон: к определению и интерпретации понятий // Аксиологические аспекты современных филологических исследований. Екатеринбург: Ажур, 2021. С. 50–52.
- Купина Н. А. Гуманитарные ценности текущего времени // Известия УрФУ. Сер. 1. Проблемы образования, науки и культуры. 2022. Т. 28, № 3. С. 86–96.
- Лызлов А. И. Аксиологическая динамика английской фразеологической системы: дис. ... д-ра филол. наук. М.: МГПУ, 2020.
- Мокиенко В. М., Никитина Т. Г. Аксиология неофразем и фразеотрансформаций в лексикографической интерпретации // Вопросы лексикографии. 2022. № 25. С. 81–107.
- Наумова Н. Г. Система ценностей человека как объект лингвистического исследования // Вестник Нижегородского университета им. Н. И. Лобачевского. 2016. № 1. С. 237–244.
- Новиков В. П. Оценочная лексика в языке английской газеты: автореф. дис. ... д-ра филол. наук. М.: РАН, Ин-т языкознания, 1992.
- Ретунская М. С. Аксиологическая лексика английского языка. Н. Новгород: изд-во Нижегородского гос. ун-та, 1996. 272 с.
- Самойлова А. В., Шершнева Н. Б. Аксиологический потенциал англоязычных слов 2022 года // Филологические науки. Вопросы теории и практики. 2023. Т. 6. Вып. 9. С. 3092–3098.
- Самойлова А. В., Элмасян А. В. Лингвоаксиологическая и прагматическая характеристика англоязычных слов 2023 года // Междисциплинарные аспекты лингвистических исследований. Краснодар, 2024. С. 131–140.
- Солдатов О. Б. Аксиологический потенциал разговорных слов mum, kid, dad // Филология и человек. 2019. № 4. С. 69–83.
- Черкасова М. Н., Тактарова В. В. Семантика и прагматика современных медиа в аспекте манипулятивного воздействия на поколение Z // Известия Южного федерального университета. Филологические науки. 2023. Т. 27, № 3. С. 30–42.
- Gebhardt L. The study of words. An Introduction. Oxon: Routledge, 2023. 202 p.
- Kozhakhmetova A., Ospanova Zh., Mussatayeva M., Bissenbayeva Zh. Axiological concepts of journalistic texts (lingua-stylistic analysis) // XLinguae. 2020. Vol. 13, iss. 2. P. 225–234.
- Kreidler C. W. Introducing English semantics: 2nd ed. New York: Routledge, 2023. 210 p.
- Schmitt N., Rodgers M. An introduction to applied linguistics. New York: Routledge, 2020. 405 p.
- Wlodarski T. Preliminary findings on the axiological potential of obituary headlines in the contemporary British press // Zeszyty Naukowe Uniwersytetu Rzeszowskiego. Ser. Filologiczna. 2008. No. 51. P. 154–167.

References

- Arnold I. V. The English word. Moscow, 1986, 295 p. (in Eng.)
- Cherkasova M. N., Taktarova V. V. Semantics and pragmatics of modern media from the perspective of manipulative effect on generation Z. *Proceedings of Southern Federal University. Philology*, 2023, vol. 27, no. 3, pp. 30–42. (in Russ.).
- Gebhardt L. The study of words. An Introduction. Oxon: Routledge, 2023. 202 p.
- Golev N. D. Spontaneous usual meaning of nominative units. *Izvestia. Ural Federal University Journal*, 2001, no. 21, pp. 94–101. (in Russ.)

- Katermina V. V.** Axiological potential of English neologisms in discursive practices. Krasnodar, 2023, 192 p. (in Russ.)
- Kazachenko O. V.** The analysis of words, denoting values in linguistic disciplines. *Psycholinguistic aspects of speech activity study*, 2020, no. 18, pp. 168–175. (in Russ.)
- Komlev I. V.** The discovery of the power of superposition in life. *Scientific Journal. Psychological sciences*, 2021, pp. 47–63. (in Russ.)
- Kozhakhmetova A., Ospanova Zh., Mussatayeva M., Bissenbayeva Zh.** Axiological concepts of journalistic texts (lingua-stylistic analysis). *XLinguae*. 2020, vol. 13, iss. 2, pp. 225–234.
- Kreidler C. W.** Introducing English semantics: 2nd ed. New York: Routledge, 2023. 210 p.
- Kupina N. A.** Axiology and axiological lexicon: to the definition and interpretation of terms. *Axiological aspects of modern philological studies*. Ekaterinburg, 2021, pp. 50–52. (in Russ.)
- Kupina N. A.** Humanitarian values of the current time. *Izvestia. Ural Federal University Journal. Ser. I. Issues in education, science and culture*, 2022, vol. 28, no. 3, pp. 86–96. (in Russ.)
- Lizlov A. I.** Axiological dynamics of the English phraseological system: dissertation. Moscow: MGPU, 2020. (in Russ.)
- Mokienko V. M., Nikitina T. G.** Axiology of neophrasemes and phraseotransformations in lexicographic interpretation. *Russian Journal of Lexicography*, 2022, no. 25, pp. 81–107. (in Russ.)
- Naumova N. G.** The system of human values as an object of linguistic research. *Vestnik of Lobachevsky University of Nizhni Novgorod*, 2016, no. 1, pp. 237–244. (in Russ.)
- Novikov V. P.** Evaluative lexis in the language of English press: author's abstract. Moscow: Russian Academy of Sciences, the Institute of Linguistics, 1992. (in Russ.)
- Retunskaya M. S.** Axiological lexis of the English language. N. Novgorod, 1996. 272 p. (in Russ.)
- Samoylova A. V., Shershneva N. B.** Axiological potential of the English words of the year 2022. *Philology. Theory and Practice*, 2023, vol. 6, iss. 9, pp. 3092–3098. (in Russ.)
- Samoylova A. V., Elmasyan A. V.** Linguoaxiological and pragmatic characteristics of the English words of the year 2023. *Interdisciplinary aspects of linguistic studies*. Krasnodar, 2024, pp. 131–140. (in Russ.)
- Schmitt N., Rodgers M.** An introduction to applied linguistics. New York: Routledge, 2020. 405 p.
- Soldatova O. B.** Axiological potential of colloquial words *mum, kid, dad*. *Philology and Human*, 2019, no. 4, pp. 69–83. (in Russ.)
- Vikulova L. G., Serebrennikova E. F., Cherkashina E. I.** Axiologeme château as the code of promoting text. Marketing linguistics in the digital epoch: ed. vol. Moscow, 2021, pp. 22–30. (in Russ.)
- Wlodarski T.** Preliminary findings on the axiological potential of obituary headlines in the contemporary British press. *Zeszyty Naukowe Uniwersytetu Rzeszowskiego. Ser. Filologiczna*. 2008, no. 51, pp. 154–167.

Список источников / List of Sources

- Bailey B.** How to rizz someone up: full guide (tested and approved). Roast.dating. URL: <https://roast.dating/blog/how-to-rizz-someone-up> (accessed: 05.11.2024).
- Cambridge Dictionary. URL: <https://www.cam.ac.uk/research/news/cambridge-dictionary-names-hallucinate-word-of-the-year-2023> (accessed: 28.04.2024).
- Collins Dictionary. URL: <https://www.collinsdictionary.com/woty?ysclid=lp8tw47kpf370498637> (accessed: 30.04.2024).
- Dictionary.com. URL: <https://content.dictionary.com/word-of-the-year-2023/#lp-pom-text-60> (accessed: 05.05.2024).
- Gupta S.** Situationship: how to cope when commotement is unclear. *VeryWellMind.com*. URL: <https://www.verywellmind.com/what-is-a-situationship-5216144> (accessed: 01.05.2024).

- Hubbard L.** All the EGOT winners in award history. *Town and Country Magazine*. URL: <https://www.townandcountrymag.com/leisure/arts-and-culture/g23064962/egot-winners-full-list/> (accessed: 05.11.2024).
- Hwang R.** 11 tips to rizz aguy or girl up (and make them fall for you). *Science of People*. URL: <https://www.scienceofpeople.com/how-to-rizz/> (accessed: 05.11.2024).
- Jones N.** How a Nepo Baby is born. *Vulture*. URL: <https://www.vulture.com/> (accessed: 27.05.2024).
- Macquarie Dictionary. URL: <https://www.macquariedictionary.com.au/blog/article/913/> (accessed: 06.05.2024).
- Merriam-Webster Dictionary. URL: <https://www.merriam-webster.com/wordplay/word-of-the-year> (accessed: 08.05.2024).
- Oxford Dictionary. URL: <https://languages.oup.com/word-of-the-year/2023/> (accessed: 11.05.2024).
- Singh O., Haasch P.** 21 people who have won an EGOT, the most coveted award in Hollywood. *Business Insider*. URL: <https://www.businessinsider.com> (accessed: 19.04.2024).
- Tzall D.** Situationship: what it means and signs you are in one. *Choosing therapy.com*. URL: <https://www.choosingtherapy.com/situationship/> (accessed: 05.11.2024).
- Urban Dictionary. URL: <https://www.urbandictionary.com> (accessed: 11.04.2024).

Информация об авторе

Самойлова Антонина Вадимовна, кандидат филологических наук, доцент кафедры английской филологии Кубанского государственного университета

Information about the Author

Antonina V. Samoylova, Candidate of Philology, Associate Professor at the English Philology Department, Kuban State University

*Статья поступила в редакцию 20.11.2024;
одобрена после рецензирования 22.01.2026; принята к публикации 30.01.2026*

*The article was submitted 20.11.2024;
approved after reviewing 22.01.2026; accepted for publication 30.01.2026*

Научная статья

УДК 004.8:821.161.1

DOI 10.25205/1818-7935-2026-24-2-42-54

Создание биографического датасета русских прозаиков XX века с помощью ChatGPT-4o

Алиса Евгеньевна Урих
Татьяна Юрьевна Шерстинова

Национальный исследовательский университет «Высшая школа экономики»
Санкт-Петербург, Россия

aeurikh@edu.hse.ru, <https://orcid.org/0009-0004-5778-2741>

tsherstinova@hse.ru, <https://orcid.org/0000-0002-9085-3378>

Аннотация

Цифровые технологии XXI века способствуют формированию новых методологических подходов в гуманитарных науках, включая литературоведение и биографическое исследование. Для автоматизации процессов сбора и аннотирования информации все чаще используют большие языковые модели (Large Language Models, LLM), включая ChatGPT. Развитие подобных инструментов стимулирует переход от традиционных методологических подходов, основанных на исключительно ручном аннотировании и экспертной верификации, к гибридным, сочетающим автоматическое извлечение информации с дальнейшей редакторской доработкой. Такая стратегия делает процесс сбора и обработки данных менее ресурсозатратным при сохранении высокого уровня точности фактов. Целью настоящей работы является создание биографического датасета русских прозаиков XX века с применением ChatGPT-4o, которая на момент проведения исследования являлась последней версией ChatGPT. Структура датасета, состоящая из 30 пунктов и описывающая особенности личной жизни и творчества авторов, была определена экспертным образом на первом этапе исследования. На вход модели подавались промпты для формулировки и последующей корректировки определяемых пользователем задач. Некоторые техники, использованные для создания эффективных запросов и ориентированные на улучшение качества и точности ответа, описаны в статье. Автоматически полученные данные были вручную перепроверены, все ошибки, допущенные моделью, исправлены. Тестирование модели показало, что ChatGPT-4o может эффективно использоваться для автоматизированной подготовки и первичного анализа больших массивов текстовых данных, а также выделило основные возможности и ограничения ChatGPT-4o при извлечении и структурировании биографических сведений. Подготовленный датасет представляет собой реляционную таблицу, содержащую биографические данные для 650 русских писателей, и стал одним из блоков биографического модуля Корпуса русского рассказа XX века. Структура датасета позволяет проводить с его помощью междисциплинарные исследования в области социологии литературы, моделирования национального литературного процесса, а также изучать корреляции между биографическими особенностями авторов и художественными, тематическими и лингвистическими характеристиками их литературных произведений.

Ключевые слова

биографии, русские прозаики, датасет, ChatGPT, большие языковые модели

Финансирование

Исследование выполнено в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ).

© Урих А. Е., Шерстинова Т. Ю., 2026

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2

Vestnik NSU. Series: Linguistics and Intercultural Communication, 2026, vol. 24, no. 2

Для цитирования

Урих А. Е., Шерстинова Т. Ю. Создание биографического датасета русских прозаиков XX века с помощью ChatGPT-4o // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 42–54. DOI 10.25205/1818-7935-2026-24-2-42-54

Creation of a Biographical Dataset of Russian Prose Writers of the 20th Century Using ChatGPT-4o

Alisa E. Urikh, Tatyana Y. Sherstinova

National Research University Higher School of Economics
Saint-Petersburg, Russian Federation

aeurikh@edu.hse.ru, <https://orcid.org/0009-0004-5778-2741>

tsherstinova@hse.ru, <https://orcid.org/0000-0002-9085-3378>

Abstract

In the 21st century, a rapid advancement of digital technologies has given rise to new methodological approaches within the humanities, including literary studies and biographical research. The automation of data processing and information annotation is increasingly supported by Large Language Models (LLMs), such as ChatGPT. The emergence of these tools has prompted a shift from traditional methodologies based exclusively on manual annotation and expert verification toward hybrid strategies that combine automated information extraction with subsequent human curation. Such an approach enhances the efficiency and sustainability of data collection and processing while preserving a high level of factual accuracy. LLMs demonstrate significant potential for automatic information extraction and classification, contributing to more systematic and scalable approaches to data-driven research in the humanities. The aim of this study is to create a biographical dataset of 20th-century Russian prose writers using ChatGPT-4o. The dataset structure, comprising 30 categories describing various aspects of the authors' personal lives and creative activities, was defined by experts at the initial stage of the study. The model received prompts designed to formulate and iteratively refine user-defined tasks. Several techniques for constructing effective prompts aimed at improving the quality and precision of model responses are described in the paper. Automatically generated data were manually verified, and all model errors corrected. The testing demonstrated that ChatGPT-4o can be effectively employed for the automated preparation and preliminary analysis of large volumes of textual data; it also revealed the model's main strengths and limitations in extracting and structuring biographical information. The resulting dataset, represented as a relational table containing biographical data for 650 Russian writers, became part of the biographical module of the Corpus of the 20th Century Russian Short Stories. Its structure enables interdisciplinary research in the fields of literary sociology and modeling of the national literary process, as well as the study of correlations between authors' biographical features and the artistic, thematic, and linguistic characteristics of their works.

Keywords

biographies, Russian prose writers, dataset, ChatGPT, large language models

Funding

This article was prepared as a result of research conducted within the framework of the Fundamental Research Program of the National Research University Higher School of Economics (HSE).

For citation

Urikh A. E., Sherstinova T. Y. Creation of a biographical dataset of Russian prose writers of the 20th century using ChatGPT-4o. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 42–54. DOI 10.25205/1818-7935-2026-24-2-42-54

Введение

Развитие современных цифровых технологий и методов искусственного интеллекта предоставляет новые возможности и перспективы для гуманитарных исследований. За последние несколько лет большие языковые модели привлекли значительное внимание исследователей и уже активно используются в образовательной, лингвистической, медицинской и других сферах, где некоторые процессы, связанные с обработкой данных, могут быть автоматизированы.

Применение больших языковых моделей позволяет не только ускорить сбор и последующую обработку данных, но и минимизировать ручной труд, что особенно актуально в области цифровой гуманитаристики (Digital Humanities). Внедрение генеративных моделей позволяет автоматизировать процессы анализа нарративных структур, аннотирования эмоций, извлечения метаданных и других аспектов работы с текстом. Несмотря на явные преимущества, использование подобных инструментов обязательно требует проверки точности и истинности генерируемых данных, исключения возможных ошибок, а также настройки и адаптации модели к специфике конкретной предметной области.

С момента появления ChatGPT, первая версия которой была выпущена компанией OpenAI в ноябре 2022 г., модель быстро получила широкое распространение благодаря высокой эффективности в решении задач, связанных с обработкой текста на естественном языке (NLP), включая генерацию, классификацию и извлечение информации, что подтверждается рядом прикладных исследований [Zhang et al., 2023; Shi et al., 2023; Wang et al., 2023]. Поэтому в настоящей работе использовался именно этот инструмент, а именно ChatGPT-4o, на момент проведения исследования являвшийся последней версией данной модели.

Цель описываемого эксперимента состояла в формировании датасета, содержащего биографические сведения о жизни и творчестве русских прозаиков XX в., с использованием технологий ChatGPT. Датасет предназначен для расширения Корпуса русского рассказа XX века [Шерстинова и др., 2025], который в свою очередь основывается на Корпусе 1900–1930 гг. [Мартыненко и др., 2018] (<https://russian-short-stories.ru/story>). Еще одной целью исследования являлось выявление возможностей и ограничений больших языковых моделей при их использовании для решения задач извлечения и классификации биографических сведений. Для достижения этих целей необходимо было решить ряд задач:

- 1) определить структуру формируемого датасета;
- 2) провести извлечение структурированных биографических данных в полуавтоматическом режиме с помощью ChatGPT-4o;
- 3) осуществить ручную проверку полученных данных;
- 4) проанализировать возможности и ограничения модели в области извлечения и структурирования биографических данных.

В отличие от существующих корпусов с биографическими данными, основанных преимущественно на ручном аннотировании, настоящее исследование рассматривает большие языковые модели как основной инструмент для извлечения и структурирования информации. Это позволит оценить потенциал и эффективность автоматизации подобных процессов для решения целого спектра задач, связанных с автоматизацией биографической информации в гуманитарных исследованиях.

1. Материал и методика исследования

Основой создаваемого структурированного биографического датасета была выборка, состоявшая из 1000 имен русских прозаиков XX в., произведения которых были отобраны для аннотированной части Корпуса русского рассказа – цифрового ресурса, предназначенного для изучения и популяризации русской малой прозы прошедшего столетия [Шерстинова, Кирина, Хлусова, 2025]. Создаваемый ресурс основан на системном подходе к литературе, сформулированном Ю. Н. Тыняновым [Тынянов, 1929], количественных методах Андрея Белого [Белый, 1934], а также лингвостатистических и стилистических принципах исследования художественного текста, разработанных Г. Я. Мартыненко [Мартыненко, 1988]. Такое сочетание методологических подходов предполагает включение в корпус произведений писателей различной степени известности – от признанных классиков до фактически забытых прозаиков, что обеспечивает объективность проводимых на его материале исследований.

Для формирования списка писателей, для которого создавался биографический датасет, были использованы данные словарей писателей, различных литературных энциклопедий, каталогов библиотек и публицистических изданий того временного периода. Принципы создания биографической базы данных для Корпуса русского рассказа подробно описаны в работах Т. Ю. Шерстиновой [Шерстинова, 2019] и А. Д. Узденовой [Уздёнова, 2021]. Датасет, созданный в рамках настоящего эксперимента, охватывает период с 1900 по 2000 г. Несмотря на то что выбор писателей «второго эшелона» осуществлялся случайным образом, их включение в итоговый список предусматривало наличие некоторых заранее определенных критериев. Первое условие подразумевало соответствие периода творческой деятельности авторов выбранным хронологическим рамкам. Второе предусматривало отсутствие повторяющихся прозаиков в итоговом списке. Третье заключалось в том, что писатели не должны находиться в эмиграции на момент выхода первой публикации. Четвертое касалось языка литературного дебюта: произведение должно было быть написано на русском языке. Однако проверка показала, что для значительной части отобранных авторов в открытых онлайн-источниках невозможно найти верифицированную биографическую информацию. Так, из 1000 русских прозаиков только 650 писателей имеют экспертно подтвержденные статьи в Википедии, среди них 549 авторов имеют статьи с краткими или фрагментарными сведениями и только 101 автор – статьи с развернутым и подробным жизнеописанием. Именно биографические данные для этих 650 имен и легли в основу предлагаемого исследования.

Структура создаваемого датасета была определена экспертным образом. Источниками для нее послужили электронные литературные энциклопедии (<https://www.concisel.ru/>; <https://feb-web.ru/>), а также структура биографической базы данных для Корпуса русского рассказа, созданная ранее [Шерстинова, 2019]. Содержание датасета должно было охватывать различные аспекты биографии прозаиков, включая как описание подробностей личной жизни авторов, так и их публикационной активности. Ниже представлены 30 параметров описания, которые определили финальную организацию датасета:

1. *ФИО писателя*
2. *Литературный псевдоним*
3. *Другие имена, под которыми публиковался автор (если есть)*
4. *Дата рождения (по старому стилю)*
5. *Дата рождения (по новому стилю)*
6. *Место рождения*
7. *Семья*
8. *Детство и юность*
9. *Образование*
10. *Личная жизнь*
11. *Дети*
12. *Социальное окружение*
13. *Эмиграция (если да, то куда и с какого года)*
14. *Основные места проживания*
15. *Основные события жизни*
16. *Первая публикация*
17. *Последняя публикация*
18. *Хронология издания важнейших произведений*
19. *Основные жанры произведений*
20. *Художественное направление*
21. *Основные темы произведений*
22. *Особенности творчества*
23. *Критика*
24. *Другие профессиональные роли*

25. *Награды и признание*
26. *Влияние и наследие*
27. *Причина смерти (если известна)*
28. *Дата смерти (по старому стилю)*
29. *Дата смерти (по новому стилю)*
30. *Место смерти*

Такая структура отличается довольно высоким уровнем детализации: она включает информацию о жизненном пути автора, личной жизни писателя, его социальном окружении, писательской деятельности и ее характерных чертах, критике произведений, других профессиональных ролях автора, наградах, а также обстоятельствах смерти.

2. Описание эксперимента

Для автоматического извлечения этих данных из биографических текстов в исследовании используется большая языковая модель ChatGPT4-o, которая на момент проведения исследования была самой последней версией ChatGPT, предоставляющей пользователям широкие возможности для загрузки текстовых, аудио- и видеофайлов [OpenAI, 2024]. Тестирование ее работы проводилось через веб-интерфейс, предоставляемый компанией-разработчиком OpenAI. В основе ChatGPT лежит глубокая нейронная архитектура трансформеров, которая на сегодняшний день считается одной из самых распространенных и эффективных технологий для задач обработки естественного языка [Капуста и др., 2024. С. 3]. Благодаря этому модель способна генерировать ответы на естественном языке на запросы пользователей и «понимать» большое количество тем [Welsby et al., 2023. С. 1047]. На вход ChatGPT подаются промпты – инструкции для больших языковых моделей, обеспечивающие соблюдение некоторых правил, касающихся объема и качества выходных данных [White et al., 2023]. Существует множество стратегий формулировки наиболее эффективного запроса, некоторые из которых были предложены специалистами платформы Открытого образования [Введение в промпт-инжиниринг, 2024]. Ниже представлены подходы, использованные при создании промпта в текущем исследовании:

1. Формулировка начальной установки. Согласно исследователям [Там же], модель предоставляет более содержательные и полезные данные в тех случаях, когда пользователь задает ей роль опытного и высококвалифицированного специалиста в определенной области. (Например, *Ты долгие годы работаешь биографом в литературном архиве. Ты – авторитетный специалист в этой сфере*).

2. Подробное описание задания. Для того чтобы улучшить результат выдачи, промпт должен включать понятную и однозначную постановку вопроса для повышения точности его интерпретации моделью. (*Тебе предложили составить биографию 650 русских прозаиков 1930–2000 гг. по заранее написанной и предложенной тебе структуре. Данные о биографии писателей нужно будет брать из Википедии, ссылки будут предоставляться тебе заранее. Нужно верно заполнить таблицу, согласно информации из Википедии, ничего не придумывая. Если данных недостаточно, то нужно просто оставить пропуск в ячейке таблицы напротив того пункта, который не упоминается в статье. Особенно тщательно проверяй все даты*).

3. Определение формата результата. Пользователь может уточнить, в каком виде должны быть представлены данные: сплошной текст, список или таблица (*Обязательно оформи данные в виде таблицы, состоящей из 2 столбцов: в первом должны быть представлены перечисленные мною выше пункты структуры, во втором – найденные сведения в статьях*).

4. Использование позитивных аффирмаций (do вместо don't). Исследователи утверждают [Там же], что языковые модели точнее обрабатывают промпты с позитивной лексикой, поэтому стоит исключить формулировки со словами «не разрешается», «запрещено», «нельзя» и т. п.,

используя вместо них утвердительные директивы (*Используй только правдивую информацию, это влияет на качество работы*).

5. В статье [Bsharat et al., 2023] дополнительно рекомендуется использовать эмоциональные промпты для повышения эффективности работы моделей. Для этого можно включить в запрос информацию о ее вознаграждении или наказании (*Тебе заплатят 10 000 долларов при условии, если весь датасет будет заполнен корректно*).

Первые результаты тестирования позволили сделать выводы об эффективности промпта и конкретизировать его для корректировки работы модели. Формулировка уточняющих вопросов – еще один метод оптимизации запроса. Большинство из внесенных в промпт изменений относится к формату написания отдельных пунктов (*В пункте «Награды и достижения» всегда указывай информацию из статьи в формате [Награда (год получения)]. В пункте «Влияние и наследие» всегда указывай информацию из статьи в виде связного текста*). Более того, модели было разрешено использовать другие авторитетные и проверенные источники, помимо Википедии, если в статье представлено недостаточно информации для извлечения данных (*Если информации из статьи в Википедии недостаточно, используй другие, но только проверенные и авторитетные источники. Обязательно приводи список этих ресурсов в ответе*).

Использование этой методики позволило получить структурированные биографические данные для всех 650 русских прозаиков, вошедших в финальную выборку. Каждый запрос формировался по единому шаблону, обеспечивающему воспроизводимость процедуры и сопоставимость результатов. В результате работы модели были автоматически сгенерированы таблицы, содержащие сведения по тридцати параметрам, отражающим ключевые аспекты жизни и творчества авторов. Полученные данные впоследствии прошли ручную верификацию и редактирование, что позволило устранить неточности, унифицировать формат представления информации и подготовить материал для дальнейшего корпусного и статистического анализа.

3. Полученные результаты и оценка работа модели

Результаты тестирования модели продемонстрировали, что она корректно соблюдает заданную структуру и формат представления данных. Ниже приведен пример типичного ответа ChatGPT-4o, иллюстрирующий способ извлечения и организации биографической информации (приводится фрагмент полной выдачи):

ФИО писателя	Анатолий Васильевич Кузнецов
1	2
Литературный псевдоним (если есть)	—
Другие имена, под которыми публиковался автор (если есть)	—
Дата рождения (по старому календарю)	—
Дата рождения (по новому календарю)	18 августа 1929 года
Место рождения	Киев, Украинская ССР, СССР

Продолжение табл.

1	2
Семья	Мать – Мария Фёдоровна Кузнецова (1904–1992), более сорока лет проработала учительницей русского языка младших классов средней школы. Отец – Василий Герасимович Кузнецов (1899–1964), в момент рождения сына сотрудник советской милиции, затем инженер, начальник цеха на Горьковском автомобильном заводе (ГАЗ). Состоял членом ВКП(б). Ушел из семьи и уехал в Нижний Новгород
Детство и юность	После ухода отца Анатолий воспитывался матерью, бабушкой и дедом. Жил в районе Куренёвка на окраине Киева. В годы оккупации тайно вел дневник, позже ставший основой романа «Бабий Яр». Учился в балетной студии при Киевском театре оперы и балета, затем в студии Театра русской драмы им. Леси Украинки. Был свидетелем нацистских преступлений в Киеве
Образование	Литературный институт имени А. М. Горького (окончил в 1960 году)
Личная жизнь	Первая жена – Ирина Нифонтовна Марченко (1935–1995), филолог. Вторая жена (фактическая) Надежда Цуркан (1945–2007), бывшая его литературным секретарем в Туле, впоследствии эмигрировала из СССР и скончалась в США. Третья жена – Иоланта Райт, англичанка польского происхождения
Дети	Сын от первого брака – Алексей Кузнецов (1960–2019), театральный критик, журналист, в 1995–2018 годах работал на Радио «Свобода». Сын от второго брака – Анатолий (род. в 1970 году). Дочь от третьего брака – Мария-Анатолия (род. в 1979 году)
Социальное окружение	Состоял в Союзе писателей СССР, был партийным работником в Тульском отделении СП, имел контакты с зарубежными издателями после эмиграции
Эмиграция (если да, то куда и с какого года)	Великобритания (с 1969 года)
Основные места проживания	Киев, Тула, Москва, Лондон
Основные события жизни	1952–1954 – работал на Каховской ГЭС, 1955 – поступление в Литературный институт, 1958 – публикация первой повести «Продолжение легенды», 1960–1969 – работал ответственным секретарем Тульского отделения Союза писателей, 1966 – выход романа «Бабий Яр», 1969 – эмиграция в Великобританию, с 1972 – работал в лондонском корреспондентском пункте Радио «Свобода», выступая с беседами в рамках еженедельной программы «Писатели у микрофона», 1978 – был экстренно госпитализирован с подозрением на инфаркт

С полной таблицей, состоящей из 650 биографий русских прозаиков XX в., можно ознакомиться по ссылке: <https://docs.google.com/spreadsheets/d/1G8ADV4XPdXW1Vy8AV3aGVp6lynyhjKgl/edit?usp=sharing&ouid=113486563260251450625&rtopof=true&sd=true>

Для оценки ответов ChatGPT-4o автоматические метрики не применялись, поскольку основная цель тестирования заключалась в проверке качественных характеристик выходных данных в условиях задач извлечения и структурирования биографических данных. Проверка всех 650 биографий из датасета проводилась вручную на основе сравнения с содержанием используемых моделью источников. Впоследствии данные были частично отредактированы, дополнены там, где ChatGPT предоставил неполную информацию, и приведены к единому формату: в сформированной таблице использовались одинаковые разделители для отдельных пунктов и унифицированные формулировки (например, дата рождения и смерти в формате «28 августа 1949 года»). Такой подход позволил найти основные преимущества и недостатки в работе модели, определить, какие структурные элементы затрудняют процесс автоматического извлечения данных, а также обеспечить стандартизированное представление информации.

В результате ручной проверки датасета удалось выявить основные возможности и ограничения работы ChatGPT-4o в извлечении и структурировании биографических данных:

1. В большинстве пунктов биографий ChatGPT смог сгенерировать таблицы с достоверными данными, руководствуясь информацией из авторитетных источников.

2. Модель всегда строго соблюдала структуру, требуемую в запросе, не пропуская ни одного пункта.

3. ChatGPT действительно использовал информацию из предоставленной в промпте статьи, а также другие авторитетные источники: литературные энциклопедии, сайты региональных музеев, верифицируемые веб-страницы, посвященные жизни и творчеству отдельного писателя (<https://www.armmuseum.ru/news-blog/vasily-nemirovich-danchenko-armenia?ysclid=ma3z3n0gni55032554>; <https://troepolskiy.ru/biograf/index.htm?ysclid=ma1bx0eqfo965092309>; https://old.bigenc.ru/domestic_history/text/3527139?ysclid=m9afgi21tt274417657), которые он приводил в списке источников в ответе.

4. Модель всегда учитывала корректировки в уточняющих запросах, частично редактируя выходные данные. Однако при последовательной генерации таблиц в рамках одного диалога наблюдалось снижение точности следования требованиям, изложенным в промпте. Модель постепенно начинала игнорировать инструкции, особенно относящиеся к формату описания пунктов.

5. Несмотря на отмеченные преимущества ChatGPT-4o, в его работе были выявлены многочисленные галлюцинации: он нередко использовал информацию, отсутствующую в предоставленных источниках и в Интернете в целом. Чаще всего ложные сведения встречались в пункте про личную жизнь и детей: модель либо игнорировала нужную информацию из статей, ставя прочерк, либо самостоятельно придумывала факты, генерируя недостоверные данные.

6. При упоминании о ключевых событиях жизни и составлении хронологии изданий важнейших произведений модель допускала фактические ошибки (чаще всего отклонения составляли один год, реже – несколько лет и десятилетий), а также орфографические ошибки при воспроизведении названий публикаций («Глаголяттяги...» вместо «Глаголюттяги») и грамматические ошибки в согласовании по роду («окончил» вместо «окончила»).

7. В ряде случаев при описании конкретного пункта модель извлекала только частичную информацию из исходной статьи. Большинство таких примеров зафиксировано при заполнении сведений об учебных заведениях, супругах и детях писателя.

8. ChatGPT был более точным при анализе коротких и структурированных статей. В тех случаях, когда логико-тематическая сегментация отсутствовала в тексте, модель часто пропускала нужные сведения. Более того, информация по одному и тому же пункту могла быть представлена сразу в нескольких местах статей, что еще больше усложняло работу ChatGPT.

Выявленные в ходе проверки возможности и ограничения работы модели демонстрируют потенциал ChatGPT-4o в задачах извлечения и структурирования биографических данных. Одновременно полученные результаты подчеркивают необходимость верификации ответов модели вручную с целью исключения допущенных ошибок и повышения качества данных в таблице.

4. Некоторые статистические данные биографического датасета

На материале сформированного биографического датасета русских прозаиков XX в. были получены статистические данные, отражающие ключевые биографические характеристики авторов и демонстрирующие распределение количества прозаиков по десятилетиям и местам их рождения и смерти, полу, уровню образования и образовательным учреждениям, странам эмиграции, основной профессии и причине смерти.

Основные статистические характеристики датасета следующие:

1. Отобранная выборка писателей демонстрирует выраженный гендерный дисбаланс: из 650 авторов, включенных в датасет, 89,4 % составляют мужчины, что почти в 8,5 раза превышает долю женщин (10,6 %). Такое распределение отражает не только исторические особенности отечественного литературного поля XX в., но и институциональные ограничения прошлого столетия.
2. Средняя продолжительность жизни русских прозаиков составила 67 лет, с самой высокой концентрацией в возрасте 70–80 лет, что несколько превышает среднюю продолжительность жизни населения России в XX в.
3. Динамика средней продолжительности жизни русских прозаиков демонстрирует устойчивый рост: от 51–52 лет в первой трети XX в. до более 80 лет в начале XXI в. Такая тенденция может быть связана как с общим повышением уровня жизни и развитием здравоохранения, так и со снижением масштабов политических репрессий, войн и эпидемий.
4. Географический анализ мест рождения и смерти говорит о том, что большая часть писателей (74 %) родилась в городах, в основном в Москве (19,7 %) и Санкт-Петербурге (14,5 %). При этом около четверти авторов родились в деревнях, селах и других малых населенных пунктах, что свидетельствует о многообразии социокультурных обстоятельств становления будущих прозаиков. В свою очередь, самые распространенные места смерти русских писателей демонстрируют ещё более выраженную их централизацию в Москве (54,4 % от числа умерших) и Санкт-Петербурге (13,2 %). 4,2 % смертей было зафиксировано также в зарубежных культурных центрах (Париж, Нью-Йорк, Ницца, Прага), что, очевидно, связано с эмиграцией авторов.
5. Анализ образовательного уровня писателей демонстрирует высокий показатель прозаиков с высшим образованием (более 70 %), что указывает на высокую академическую подготовку литераторов того времени. Чаще всего упоминаются такие учебные заведения, как Литературный институт им. А. М. Горького, Московский и Санкт-Петербургский государственные университеты.
6. В эмиграции находились 15,7 % из всех авторов, включенных в датасет. Наиболее распространенными странами оказались Франция (6,3 % от числа эмигрантов), Германия (5,1 %) и США (3,5 %) – ключевые направления русской эмиграции в XX в., особенно в периоды первой и второй волн. Помимо названных, в списке также присутствуют Чехия, Израиль, Великобритания и ряд других стран.
7. Почти все писатели исследованной выборки занимали также иные профессиональные должности, помимо литературной: работали редакторами, журналистами, сценаристами, переводчиками, преподавателями, занимались государственной и обществен-

ной деятельностью. Эта статистика указывает на разнообразие профессий авторов в XX в., а также демонстрирует тесную связь прозаиков с образовательной, творческой и культурной областями.

8. Данные, связанные с причинами смерти русских прозаиков XX в., содержащихся в датасете, позволяют увидеть корреляции с социально-историческим контекстом. Хотя большее количество смертей связано с естественными причинами, в выборке наблюдаются случаи неестественных (самоубийства, несчастные случаи) и насильственных причин смерти (казни, гибель на войне), что также может быть связано с периодом политических репрессий 1930-х гг. и военных конфликтов.
9. Большинство русских прозаиков XX в., представленных в датасете, опубликовали первое произведение в возрасте от 18 до 30 лет. С увеличением возраста показатель начал постепенно снижаться: исключительными примерами являются авторы, которые начали издавать произведения после 50 лет: Г. Н. Евневич (55 лет), В. Л. Кондратьев (59 лет). Самый ранний дебют состоялся в 14 лет у И. Л. Оршера, самый поздний – в 75 лет у Г. А. Воронской.
10. Среди многочисленных литературных жанров самыми популярными оказались рассказы, повести и романы, что указывает на предпочтение писателями традиционных малых и средних форм прозы. Наиболее часто встречающимися литературными направлениями в выборке являются социалистический реализм и реализм.

Заключение

Проведенное исследование позволило создать биографический датасет, содержащий основную информацию о жизненном пути и литературном творчестве 650 русских прозаиков XX в. Созданный датасет может использоваться для анализа соотношения биографических, исторических и литературных факторов, влияющих на формирование мировоззрения писателей, их жанровых предпочтений и тематического диапазона произведений. На его основе возможно выявление статистических закономерностей для междисциплинарных исследований в области цифровой гуманитаристики, литературоведения, социологии литературы и истории культуры.

Эксперимент продемонстрировал эффективность применения больших языковых моделей, в частности ChatGPT-4o, в задачах автоматического извлечения и структурирования биографических сведений. В ходе тестирования были выявлены основные функциональные возможности и ограничения в работе модели. В большинстве случаев модель смогла сгенерировать точные и репрезентативные данные, корректно выполняя требования запроса и обеспечивая высокий уровень качества информации. Однако полностью избежать ошибок в ответах, в том числе орфографических и грамматических, не удалось. Главное ограничение модели связано с генерацией недостоверной информации (так называемых «галлюцинаций»), что до сих пор остается серьезной проблемой, особенно в контексте гуманитарных исследований. При этом стоит учитывать, что автоматическое извлечение данных во многом связано с качеством исходных текстов и уровнем их структурной организации в авторитетных источниках. Необходимость ручной верификации не раз подчеркивалась исследователями и находит подтверждение в настоящей работе. Именно поэтому полученная информация была отредактирована и дополнена в тех пунктах биографий, которые недостаточно и некорректно освещались моделью. Такой подход обеспечил надежность и достоверность сведений как о личной жизни авторов, так и об особенностях их творческой деятельности, что принципиально важно при формировании биографического датасета. Модель испытывала затруднения при извлечении данных из объемных, слабо структурированных текстов, что обуславливает необходимость создания размеченных и формализованных источников для проведения исследований с применением технологий ChatGPT.

В дальнейшем возможны эксперименты с интеграцией различных больших языковых моделей с целью сравнения результатов работы нескольких инструментов. Также перспективным представляется масштабирование уже существующей выборки для расширения временных рамок и выявления биографических закономерностей для других литературных эпох.

Список литературы

- Белый А.** Мастерство Гоголя: Исследование. М.–Л.: ОГИЗ, 1934.
- Капуста К. Л.** и др. Аналитический обзор интегральных моделей и стратегий распознавания речи на основе архитектуры трансформер // Информационно-управляющие системы. 2024. № 5. С. 2–15.
- Мартыненко Г. Я.** Основы стилистики. Л.: ЛГУ, 1988.
- Тынянов Ю. Н.** Архаисты и новаторы. Л.: Прибой, 1929.
- Узденова А. Д.** Биографическая база данных русских писателей на основе Корпуса рассказа 1900–1930 гг.: выпускная квалификационная работа бакалавра. СПб.: НИУ ВШЭ, 2021.
- Шерстинова Т. Ю.** Биографическая база данных русских писателей (к созданию корпуса русского рассказа XX века) // Корпусная лингвистика – 2019. 2019. С. 439–447.
- Шерстинова Т. Ю., Кирина М. А., Хлусова Я. К.** Корпус русского рассказа XX века: текущее состояние и перспективы развития // Тр. междунар. конф. «Корпусная лингвистика – 2025». СПб.: СПбГУ (в печати).
- Bsharat S. M.** et al. Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4 // arXiv preprint arXiv:2312.16171. 2023.
- Shi Y.** et al. ChatGraph: Interpretable Text Classification by Converting ChatGPT Knowledge to Graphs // Proceedings of the IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 2023. P. 515–520.
- Wang A.** et al. An Extensive Benchmark Study on Biomedical Text Generation and Classification using ChatGPT // Bioinformatics. 2023.
- Welsby P., Cheung B. M. Y.** ChatGPT // Postgraduate Medical Journal. 2023. Vol. 99. No. 1176. P. 1047–1048.
- White J.** et al. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT // arXiv preprint arXiv:2302.11382. 2023.
- Zhang Y.** et al. ChatGPT and Finetuned BERT: A Comparative Study for Developing Intelligent Design Support Systems // Design, Construction and Building. 2023.

Список источников

- Армянский музей Москвы и культуры наций: [Электронный ресурс]. URL: <https://www.armmuseum.ru/news-blog/vasily-nemirovich-danchenko-armenia> (дата обращения: 12.05.2025).
- Большая российская энциклопедия: [Электронный ресурс]. URL: <https://bigenc.ru/p/about-project> (дата обращения: 13.05.2025).
- Введение в промпт-инжиниринг // Открытое образование. 2024. URL: https://apps.openedu.ru/learning/course/course-v1:hse+PROMPT_ENGINEERING+2024/home (дата обращения: 22.04.2025).
- Корпус русского рассказа 1900–1930: [Электронный ресурс]. URL: <https://russian-short-stories.ru/story> (дата обращения: 07.04.2025).
- Краткая литературная энциклопедия: [Электронный ресурс]. URL: <https://www.concisel.ru/> (дата обращения: 15.05.2025).
- О жизни и творчестве писателя Г. Н. Троепольского: [Электронный ресурс]. URL: <https://troepolskiy.ru/biograf/index.htm> (дата обращения: 12.05.2025).

Фундаментальная электронная библиотека «Русская литература и фольклор»: [Электронный ресурс]. URL: <https://feb-web.ru/feb/feb/about1.htm> (дата обращения: 15.05.2025).
Hello GPT-4o // OpenAI. 2024. URL: <https://openai.com/index/hello-gpt-4o/> (date of access: 07.05.2025).

References

- Bely A.** The Mastery of Gogol: A Study. Moscow–Leningrad: OGIZ, 1934. (in Russ.)
- Bsharat S. M.** et al. Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4. *arXiv preprint arXiv:2312.16171*, 2023.
- Kapusta K. L.** et al. Analytical Review of Integrated Models and Strategies for Speech Recognition Based on Transformer Architecture. *Information and Control Systems*, 2024, no. 5, pp. 2–15. (in Russ.)
- Martynenko G. Ya.** Fundamentals of Stylometry. Leningrad: Leningrad State University, 1988. (in Russ.)
- Shi Y.** et al. ChatGraph: Interpretable Text Classification by Converting ChatGPT Knowledge to Graphs. In: *Proceedings of the IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2023, pp. 515–520.
- Tynyanov Y. N.** Archaists and Innovators. Leningrad: Priboy, 1929. (in Russ.)
- Uzdenova A. D.** Biographical Database of Russian Writers Based on the Corpus of Short Stories 1900–1930. Bachelor's Thesis. Saint Petersburg: HSE University, 2021. (in Russ.)
- Wang A.** et al. An Extensive Benchmark Study on Biomedical Text Generation and Classification Using ChatGPT. *Bioinformatics*, 2023.
- Welsby P., Cheung B. M. Y.** ChatGPT. *Postgraduate Medical Journal*, 2023, vol. 99, no. 1176, pp. 1047–1048.
- White J.** et al. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *arXiv preprint arXiv:2302.11382*, 2023.
- Zhang Y.** et al. ChatGPT and Finetuned BERT: A Comparative Study for Developing Intelligent Design Support Systems. *Design, Construction and Building*, 2023.
- Sherstinova T. Yu.** Biographical Database of Russian Writers (Towards the Creation of the 20th Century Russian Short Story Corpus). In: *Corpus Linguistics – 2019*, 2019, pp. 439–447. (in Russ.)
- Sherstinova T. Yu., Kirina M. A., Khlusova Ya. K.** Russian Short Story Corpus of the 20th Century: Current State and Development Prospects. In: *Proceedings of the International Conference “Corpus Linguistics – 2025”*. Saint Petersburg: Saint Petersburg State University, in press. (in Russ.)

Sources

- Armenian Museum of Moscow and Culture of Nations. Vasily Nemirovich-Danchenko and Armenia. Available at: <https://www.armmuseum.ru/news-blog/vasily-nemirovich-danchenko-armenia> (accessed: 12 May 2025).
- Great Russian Encyclopedia. About the Project. Available at: <https://bigenc.ru/p/about-project> (accessed: 13 May 2025).
- Open Education Platform. Introduction to Prompt Engineering. 2024. Available at: https://apps.openedu.ru/learning/course/course-v1:hse+PROMPT_ENGINEERING+2024/home (accessed: 22 April 2025).
- Russian Short Stories Corpus 1900–1930. Available at: <https://russian-short-stories.ru/story> (accessed: 7 April 2025).
- Concise Literary Encyclopedia. Available at: <https://www.conciselit.ru/> (accessed: 15 May 2025).

On the Life and Work of Writer Gavriil Nikolaevich Troepolsky. Available at: <https://troepolskiy.ru/biograf/index.htm> (accessed: 12 May 2025).

Fundamental Electronic Library "Russian Literature and Folklore". Available at: <https://feb-web.ru/feb/feb/about1.htm> (accessed: 15 May 2025).

OpenAI. Hello GPT-4o. 2024. Available at: <https://openai.com/index/hello-gpt-4o/> (accessed: 7 May 2025).

Информация об авторах

Шерстинова Татьяна Юрьевна, кандидат филологических наук, доцент департамента филологии НИУ ВШЭ

Урих Алиса Евгеньевна, студентка НИУ ВШЭ

Information about the Authors

Tatyana Y. Sherstinova, Candidate of Philology, Associate Professor of the Department of Philology, Academic Director of the Philology Program, National Research University Higher School of Economics – St. Petersburg

Alisa E. Urikh, Student at the National Research University Higher School of Economics – St. Petersburg

*Статья поступила в редакцию 26.11.2025;
одобрена после рецензирования 16.04.2026; принята к публикации 17.04.2026*

*The article was submitted 26.11.2025;
approved after reviewing 16.04.2026; accepted for publication 17.04.2026*

Научная статья

УДК 81 + 81'33 + 811.161.1

DOI 10.25205/1818-7935-2026-24-2-55-68

Применение больших языковых моделей для качественного анализа слабоструктурированных текстовых данных

Анастасия Александровна Белоус

Анна Петровна Бойкова

Российский национальный исследовательский медицинский университет им. Н. И. Пирогова
Москва, Россия

Арктический научный центр
Москва, Россия

belous_aa@rsmu.ru, <https://orcid.org/0000-0002-9318-9222>
anny.boy@mail.ru, <https://orcid.org/0009-0003-2335-4522>

Аннотация

Цель настоящей работы – оценить возможность применения открытых больших языковых моделей для проведения качественного анализа слабоструктурированных текстовых данных. Примером таких данных являются результаты анкетирования, содержащего «открытые» вопросы. На нескольких общедоступных языковых моделях продемонстрирована методология обработки результатов анкетирования, предметом которого явилось определение связи между базовыми эмоциями и вызывающими их базовыми ароматами. Результаты обработки данных качественно согласуются между различными языковыми моделями и с выводами оригинальной методики. По результатам проведенной работы описаны выявленные возможности и ограничения больших языковых моделей, предложены перспективные направления их применения.

Ключевые слова

большие языковые модели, анализ текста, качественный анализ, открытые вопросы, автоматическая обработка естественного языка

Для цитирования

Белоус А. А., Бойкова А. П. Применение больших языковых моделей для качественного анализа слабоструктурированных текстовых данных // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 55–68. DOI 10.25205/1818-7935-2026-24-2-55-68

© Белоус А. А., Бойкова А. П., 2026

Application of Large Language Models for Qualitative Analysis of Loosely Structured Textual Data

Anastasiya A. Belous, Anna P. Boikova

Federal State Autonomous Educational Institution of Higher Education
«N. I. Pirogov Russian National Research Medical University» of the Ministry of Health Russian Federation
Moscow, Russian Federation

LLC «Arctic research center»
Moscow, Russian Federation

belous_aa@rsmu.ru, <https://orcid.org/0000-0002-9318-9222>
anny.boy@mail.ru, <https://orcid.org/0009-0003-2335-4522>

Abstract

The purpose of this paper is to evaluate the possibility of using publicly available large language models for qualitative analysis of loosely structured textual data. An example of such data are the results of a survey containing “open-ended” questions. Using several publicly available large language models, the processing methodology of the results of the questionnaire, the subject of which was to determine the relationship between basic emotions and odours that evoke them, is demonstrated. The results of the data processing are in qualitative agreement with different large language models and the conclusions of the original methodology. Based on the results of this work, the possibilities and limitations of large language models as well as promising directions of their application are presented.

Keywords

large language models, text analysis, qualitative analysis, open-ended questions, automatic natural language processing

For citation

Belous A. A., Boikova A. P. Application of large language models for qualitative analysis of loosely structured textual data. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 55–68. DOI 10.25205/1818-7935-2026-24-2-55-68

Введение

На данном этапе развития больших языковых моделей (далее – БЯМ) стремительно расширяются мыслимые варианты их применения. С помощью БЯМ успешно решаются такие сложные задачи, как реферирование и генерация текста [Голубинский и др., 2025; Паленова, Воронин, 2024], многоязычный перевод [Кочкаров, 2024] и интеллектуальный поиск. Вместе с тем основные прогнозируемые приложения БЯМ сосредоточены вокруг новой возможности, касающейся качественного анализа слабоструктурированных текстовых данных [Nazi, Peng, 2024]. При анализе данных на качественном уровне исследователи разрабатывают концепции, категории и выводы на основе паттернов в данных. В отличие от количественных исследований, сила которых заключается в способности обобщать и воспроизводить данные при наличии соответствующей теории, изначальный процесс формирования гипотез и новых теоретических представлений относится скорее к категории качественных методов [Taylor et al., 2015]. При этом сам процесс анализа данных считается одной из сложных частей качественного исследования [Castleberry, Nolen, 2018].

Цель данной работы – оценить возможность применения БЯМ для проведения качественного анализа слабоструктурированных текстовых данных на русском языке. Примером таких данных являются результаты анкетирования, содержащего «открытые» вопросы. Специфика обработки такой информации состоит в том, что на практике для исследователя при кодировании невозможно одновременно добиться надежности и простоты обработки без существенных потерь информации [Оберемко, 2018]. Более того, обработка большого количества текстовых данных является трудоемкой, а интерпретация полученных результатов не всегда однозначна, поскольку доля субъективности в высказываниях респондентов очень высока.

Материалы и методы

Обработка результатов опроса/анкетирования с помощью больших языковых моделей

В данной работе проводится апробация нескольких БЯМ в качестве инструмента для интерпретации результатов анкетирования. Перед выполнением основной задачи предлагается осуществить предварительную проверку моделей «на адекватность» путем оценки качества ответов на тестовые вопросы, аналогично [Nezhurina et al., 2024].

Выбор предмета анкетирования: связь запахов и эмоций

Исследователи давно отмечают наличие взаимосвязи между обонянием и лимбической системой, отвечающей в свою очередь за генерацию человеческих эмоций [Zald et al., 1997; Buchanan et al., 2003]. Пациенты, инфицированные вирусом SARS-COV-2 (ранее 2019-nCoV, вызывающий коронавирусную инфекцию), примерно в половине случаев теряют обоняние, при этом наличие нарушений в обонянии более чем в два раза увеличивает вероятность последующих психиатрических расстройств (42 % против 19 %) [Metelkina-Fernandez et al., 2024]. Существуют также свидетельства того, что миндалевидное тело играет роль в формировании врожденного поведения, обусловленного запахами, что, по мнению ряда авторов, не исключает его участия в выученном обонятельном поведении [Root et al., 2014; Wang et al., 2024].

Одним из следствий такой взаимосвязи считается наличие эмоциональной составляющей в восприятии современным человеком запахов, так как информация, поступившая по каналу обоняния, менее подвержена интеллектуализации, чем визуальные, слуховые и другие образы.

Последние десятилетия формируются базы данных запахов, такие как FlavorNet [Acree, Arn, 2024], включающие в себя описание физико-химических свойств одорантов и их восприятие человеком и млекопитающими [Hamel, 2024], а также проводятся исследования по выявлению взаимосвязей между строением молекул и дескрипторами запахов в естественном языке [Harada, 2024]. Особое внимание уделяется изучению роли запахов в межкультурной коммуникации, поскольку обонятельный канал является древнейшим способом дистантной связи: как часть семиотического поля той или иной культуры, запахи могут как содействовать, так и препятствовать общению, определять глубинные культурные установки ее носителей [Пак Сон, 2009]. Самостоятельное направление образуют исследования, изучающие репрезентацию запахов в художественных произведениях и ольфакторную образность как средство передачи философских, символических и эмоциональных смыслов [Юрченко, 2021а; 2021б], а также трансформацию ольфакторного кода при переводе [Олицкая, Черткова, 2020].

Для выявления взаимосвязи между обонянием и лимбической системой относительно недавно была предложена методика анкетирования [Березина, 2011], которая взята за основу в данной работе. Методика включает восприятие анкетирруемыми набора базовых запахов и формирование анкетирруемыми ответов на перечень «открытых» вопросов, с последующей интерпретацией результатов анкетирования с целью выявления базовых эмоций, вызываемых каждым запахом.

Используемые базовые запахи

Согласно стереохимической теории обоняния, базовые запахи отличаются не химическим составом, а наличием в конфигурации молекулы, воздействующей на рецептор, структур определенной геометрии и поляризации [Amoore, 1963].

В настоящей работе была применена методика эксперимента, описанная в работах проф. Т. Н. Березиной, с небольшими модификациями в формировании линейки базовых арома-

тов: вместо семи представительных веществ для формирования линейки запахов было взято десять различных представительных веществ (табл. 1). Таким образом, часть запахов была продублирована. Дублирование осуществлялось при формировании набора базовых запахов в случае, если, по оценке авторов, у различных производителей представленные на рынке образцы запахов существенно различались между собой. В основном дублирование коснулось образцов запахов, представленных в форме эфирных масел. Стимульные материалы помещались в пробирки типа эппендорф объемом 1,5 мл в количестве 0,2–0,4 мл, в промежутках между проведениями тестирования материалы хранились в холодильнике при температуре 10 °С. Образцы готовились заново после четырех дней хранения или после использования 6–7 испытуемыми.

Таблица 1

Описание базовых запахов и представительных образцов,
используемых в качестве стимульных материалов.
Некоторые образцы представлены ароматическими маслами с соответствующим запахом

Table 1

Description of the basic odors and representative samples used as stimulus materials.
A few samples are presented as aromatic oils with corresponding odours

Название запаха (согласно [Березина, 2011])	Образцы, используемые в данной работе
1. Камфарный	1. Масло эвкалипта
2. Мятный	2. Мята полевая
3. Фруктовый/ эфирный	3. Масло груши 4. Масло апельсина
4. Цветочный	5. Масло лаванды 6. Масло жасмина 7. Масло розы
5. Гнилостный	8. Тухлые яйца/гнилой банан
6. Едкий	9. Уксус столовый
7. Muskusный	10. Струя бобра

Выявляемые базовые эмоции

Зависимые переменные – семь базовых эмоций, определяемых по классификации Экмана [Ekman, Cordaro, 2011], модифицированной согласно [Березина, 2011]:

Положительные эмоции:

- радость (англ. happiness/joy),
- удовольствие (англ. pleasure/joy),
- удивление (англ. surprise).

Отрицательные эмоции:

- страх (англ. fear),
- гнев (англ. anger),
- отвращение (англ. disgust),
- печаль (англ. grief/sadness).

Методика ассоциативного анкетирования

Методика изучения ассоциаций состоит в последовательном восприятии запаха и ответе на вопросы специально разработанной анкеты, приведенной в [Березина, 2011]. Суть анкеты, содержащей восемь вопросов, сводится к следующему:

- 1) описание в свободной форме эмоций, ассоциируемых с запахом;
- 2) последующее раскрытие описания через ассоциативные образы, вкусовые и телесные ощущения, звуки, а также соответствующие запаху жизненные ситуации.

В первичном варианте методики проф. Т. Н. Березиной для облегчения обработки выполнялась формализация процесса классификации ответов в базисе эмоций по количеству появления ключевых слов или словосочетаний. Позднее, в силу сложности интерпретации анкет, автор отказался от обработки результатов анкетирования [Berezina, 2014], и каждому испытуемому уже было предложено самостоятельно проставить количественные величины взаимосвязей между базовыми запахами и базовыми эмоциями. Таким образом, в приведенном примере [Berezina, 2014] выбор алгоритма обработки результатов анкетирования остался открытым, что, в свою очередь, делает его привлекательным материалом для исследования возможностей языковых моделей.

Для целей данной работы была сформирована выборка из 20 анкет испытуемых, каждая из которых состояла из 80 ячеек с ответами на «открытые» вопросы, внесенными после восприятия запаха, идентично первоначальному варианту методики [Березина, 2011]. В отличие от оригинальной методики, дополнительно анкетироваемым предлагалось в ответе с помощью условных знаков «-1», «0», «+1» пояснить, является ли ассоциация, вызываемая запахом, отрицательной, нейтральной или положительной.

Методика обработки анкет с помощью больших языковых моделей

В ходе обработки результатов для каждой анкеты, идентично исходной методике, решалась задача построения матрицы «запахи – эмоции». Для построения данной матрицы на основе анкеты формировался текст запроса (так называемого промпта, от англ. prompt) к соответствующей языковой модели, структура которого состояла из четырех смысловых блоков:

- Блок 1: описание структуры таблицы анкеты.
- Блок 2: текст результата анкетирования – заполненная таблица анкеты в формате csv.
- Блок 3: описание задачи соотнесения запахов с базовыми эмоциями; при этом использовались два варианта описания эмоций: (а) простой перечень названий эмоций и (б) уточнение определения выявляемых эмоций, идентично взятой за основу методике [Березина, 2011].
- Блок 4: запрос на формирование матрицы «запахи – эмоции», в ячейках которой содержатся числа 0, 1, 2 в соответствии с выраженностью эмоции, вызываемой образцом запаха.

Введение дифференцированной градации по выраженности эмоции в данной работе продиктовано тем, что интенсивность эмоции принято рассматривать как ее неотъемлемую составляющую [Reisenzein, Junge, 2024].

Для каждого нового запроса формировалась новая сессия, т. е. результаты предыдущих запросов не влияли на результаты последующих.

Результирующие матрицы «запахи (строки) – эмоции (колонки)» содержали в каждой ячейке значение от 0 до 2, отражающее значимость взаимосвязи в контексте запаха:

- 2 – означает наиболее выраженную эмоцию (далее – «выраженная»);
- 1 – означает менее выраженную эмоцию (далее – «умеренно выраженная»);
- 0 – означает отсутствие или очень слабую выраженность эмоции.

Для оценки величины различия/схожести двух матриц A и B «запахи – эмоции» для каждого i -го запаха была введена метрика E_i , показывающая величину различия в соотношении данного запаха с эмоциями в рассматриваемой паре матриц (0 – идентичные, 1 – отсутствие совпадений). Данная метрика вычисляется как

$$E_i = \frac{\left(\sum_{j=1}^7 |A_{ij} - B_{ij}| \right)}{\left(\sum_{j=1}^7 (A_{ij} + B_{ij}) \right)}, \quad (1)$$

где A_{ij} , B_{ij} – ячейки сравниваемых матриц, содержащие величины значимости взаимосвязи между запахом $i \in [1...10]$ и базовой эмоцией $j \in [1...7]$, суммирование в числителе и знаменателе выполняется по индексу базовых эмоций.

Как будет показано далее, для каждого запаха в среднем идентифицируется около одной выраженной и одной умеренно выраженной эмоций. С учетом заполненности матриц для заданного запаха i значение величины:

- $E_i = 0$ соответствует полному совпадению в идентификации эмоций;
- $E_i = 0,33$ приблизительно соответствует ситуации совпадения в идентификации выраженной и различия в умеренно выраженной эмоции;
- $E_i = 0,67$ приблизительно соответствует ситуации совпадения в идентификации умеренно выраженной и различия в выраженной эмоции;
- $E_i = 0,76$ представляет собой среднее отклонение для случайного соотношения запаха с эмоциями;
- $E_i = 1$ соответствует ситуации отсутствия пересечений в выявленных эмоциях.

Для сопоставления двух матриц результаты построчного сравнения E_i усреднялись по индексу запахов i , и полученная величина, принимающая значения от 0 до 1, характеризовала степень различия двух матриц.

Результаты

Перед решением основной задачи была осуществлена первичная проверка БЯМ путем оценки качества ответов на тестовые вопросы, аналогично методике [Белоус, Бойкова, 2025]. Использованы следующие формулировки тестовых вопросов:

1. «Задача Алисы» [Nezhurina et al., 2024] с модифицированными числовыми значениями.
2. «Почему зимой холодно, а летом тепло?» – для оценки способности излагать факты без искажения.
3. «Если наблюдаем у человека одновременно слезы и смех, он испытывает положительную или отрицательную эмоцию?» – для оценки способности формулировать наиболее вероятное решение/гипотезу при нескольких возможных. Запрос был дополнен в рамках одной сессии: «Оцени вероятность возникшей эмоции: отрицательная или положительная?».
4. «Что обычно означает ответ «Да нет, наверное?»» – оценивалась как правильность интерпретации («отрицательный ответ»), так и указание на исходный язык данного выражения (русский).

Выбор ансамбля языковых моделей

Был сформирован предварительный список известных языковых моделей (на момент проведения эксперимента, ноябрь 2024 г.), находящихся в открытом доступе:

1. ChatGPT-4-turbo, <https://chatgpt.com/>

2. Perplexity, <https://www.perplexity.ai/>
 3. GigaChat, <https://giga.chat/>
 4. YandexGPT 4 Pro, <https://ya.ru/ai/gpt-3>
 5. Llama 3.1-405B, <https://huggingface.co/spaces/llamameta/llama3.1-405B>
 6. Claude 3.5 (sonnet), <https://www.anthropic.com/>
 7. Gemma 2, <https://huggingface.co/spaces/AIDC-AI/Ovis1.6-Gemma2-9B>
 8. Qwen 2.5-72B-instruct, <https://huggingface.co/spaces/Qwen/Qwen2.5>
- Результаты тестирования БЯМ представлены в табл. 2.

Таблица 2

Результаты проверки языковых моделей на качество обработки тестовых задач

Table 2

Results of testing language models for the quality of processing test tasks

Тестируемая языковая модель	1. «Зима – лето»	2. «Задача Алисы»	3. «Смех и слезы»		4. «Да нет, наверное»	
			выбор эмоции	вероятность положительной эмоции, %	ответ	указание на язык
ChatGPT	+/-	–	–	60–70	+	–
Perplexity	+	+	+	70–80	–	–
GigaChat	+	–	-/+	60	–	–
YandexGPT	-/+	–	–	–	–	–
Llama	–	+	–	60	+/-	+
Claude	+	+	–	65	+	+
Gemma	–	–	–	–	–	–
Qwen	+/-	–	–	50–60	–	–

Примечание: «+» – полностью корректный ответ; «–» – неправильный ответ; «+/-», «-/+» – частично правильный ответ, в большей или меньшей степени; 60 % – пример оценки вероятности положительной эмоции в уточняющем запросе в тестовой задаче № 3.

По результатам проверки для дальнейшего проведения эксперимента были выбраны следующие языковые модели: GigaChat, ChatGPT, Claude, Llama, Perplexity.

Следует отметить, что данная работа носит методический характер, а проведение эксперимента и обработка результатов представленными языковыми моделями соответствуют концу 2024 года. Позже было показано, что на интервале нескольких месяцев существенно расширился список общедоступных языковых моделей, при этом качество ряда языковых моделей, таких как YandexGPT, GigaChat, DeepSeek, за столь короткий временной интервал выросло на порядок, и в настоящее время в данном рейтинге они занимают лидирующие позиции [Белоус, Бойкова, 2025].

Результаты обработки анкет с помощью больших языковых моделей

По результатам обработки анкет представленными языковыми моделями получены достаточно разреженные матрицы «запахи – эмоции», в ячейках которых преобладают значения «0». В среднем в одной строке ненулевые значения встречались:

- для «1» ~ 1,1 раза;
- для «2» ~ 0,8 раза.

Полученные величины означают, что при обработке анкет для каждого запаха в среднем выявляется около одной ярко выраженной и одной умеренно выраженной эмоции.

С целью оценить чувствительность (вариативность) вычисляемой матрицы к формулировкам запроса для каждой языковой модели выполнены следующие тесты:

- Сравнение результатов идентичных запросов, выполненных при настройках языковых моделей «по умолчанию».
- Сравнение результатов запросов, содержащих два синонимичных варианта описания эмоций: 1) перечень названий эмоций и 2) поясняющее уточнение определений эмоций, сформулированное идентично тому, как это делается во взятой за основу методике [Березина, 2011].

Для оценки изменчивости (вариативности) вычисляемой матрицы «запахи – эмоции» в зависимости от языковой модели проведено попарное сравнение результатов обработки анкет, выполненное разными языковыми моделями. Результаты представлены в табл. 3.

Таблица 3

Оценка величины вариативности вычисляемой матрицы «запахи – эмоции» между моделями. Диагональные элементы соответствуют вариативности идентичных и синонимичных запросов

Table 3

Estimation of the variability of the calculated “smells-emotions” matrix between the models. Diagonal elements correspond to the variability of identical and synonymous queries

	GigaChat	ChatGPT	Claude	Llama	Perplexity
GigaChat	0,37+/-0,04 0,46+/-0,04	0,38+/-0,04	0,38+/-0,03	0,35+/-0,04	0,41+/-0,04
ChatGPT	—	0,13+/-0,03 0,26+/-0,04	0,29+/-0,03	0,36+/-0,02	0,31+/-0,03
Claude	—	—	0,13+/-0,04 0,27+/-0,04	0,30+/-0,03	0,26+/-0,05
Llama	—	—	—	0,09+/-0,04 0,24+/-0,03	0,28+/-0,04
Perplexity	—	—	—	—	0,10+/-0,03 0,30+/-0,04

Наблюдаемая изменчивость (неопределенность) в результатах запроса характерна для БЯМ и иногда может регулироваться такими настроечными параметрами, как «температура». Также есть неопределенности, связанные:

- с произвольностью формулировки задачи (в нашем примере – описание определения эмоций в запросе или его отсутствие);
- с выбором языковой модели для обработки анкет.

Итоговая вариативность в результате обработки анкеты составила порядка 0,34+/-0,05. Для сравнения, величина различия между вычисляемой матрицей «запахи – эмоции» в данной работе с использованием выбранных моделей и матрицей, представленной в [Березина, 2011], составила 0,44+/-0,15.

Способ идентификации эмоций

Для всех моделей после получения ответа на запрос обработки анкет давался дополнительный запрос на пояснение, как были определены эмоции. Данный вопрос должен был по-

мочь восстановить «схему» используемых в языковых моделях алгоритмов качественного анализа, если такие прослеживались. Далее представлено краткое обобщение, которое базируется на пояснениях, представленных самими моделями.

Все модели упомянули при описании алгоритма, используемого при построении матрицы «запахи – эмоции», схожие подходы:

- использование подсчета ключевых слов и фраз, аналогично оригинальной работе [Березина, 2011];
- декомпозиция ответов в базисе характеристик: вкусовые и телесные ощущения, образы, ситуационный контекст и вторичные – цветовые ассоциации, что отчасти продиктовано структурой анкеты; последующее соотнесение результата декомпозиции с целевым списком эмоций;
- ранжирование по явному и неявному упоминанию эмоции.

При этом ряд моделей указал на использование дополнительного функционала в идентификации эмоций, например:

- Claude сформулировала несколько правил – алгоритмов качественного анализа, используемых при решении задачи: пищевые образы в лексике и ассоциациях, как правило, позволяют отнести эмоцию к удовольствию даже при нейтральной основной оценке; подробные описания позволяли отнести эмоцию к положительной, краткие – к отрицательной. Также модель ссылалась на использование психологического портрета анкетированного.
- Llama дополнительно нашла в сети Интернет необходимый инструментарий: оригинальную методику и таблицы группировки эмоций (<https://trendingclub.by/tabлица-chuvstv/>). Также модель выдвинула дополнительные гипотезы по обработке данных, например, что запах одновременно может вызывать эмоции различных знаков.
- Perplexity указала на возможность идентифицировать принадлежность респондентов к определенной (в широком смысле) культуре как на фактор, способный изменять оценку эмоции вплоть до противоположной (указав, например, что запах лаванды может ассоциироваться с расслаблением в одной культуре, и с медицинскими процедурами – в другой). Примечательно, что в недавнем исследовании [Sorokowski, 2024] было показано, что на обонятельные предпочтения человека влияет доступность запаха в экологических и культурных нишах. При уточняющих вопросах модель Perplexity идентифицировала отвечавших на анкеты преимущественно как географически принадлежащих к русской / (восточно-)европейской культурам.

Обсуждение

Анализ результатов анкетирования

Как показало обсуждение анкет с отдельными интервьюируемыми, также проводившееся в рамках данного исследования, задача идентификации эмоций является далеко не однозначной и сильно зависит от жизненного опыта испытуемого, в том числе принадлежности к определенной культуре. Например, в средней полосе, где зима и межсезонье длятся более половины года, наличие любых запахов в целом может оцениваться более позитивно, чем в менее контрастном климате. Но даже в этом случае ассоциации могут быть не очевидными: например, образ лета иногда подразумевал отрицательную эмоцию (комары), а образ больницы – положительную (выздоровление).

В целом, результат обработки анкет с помощью языковых моделей на качественном уровне согласуется с результатом оригинальной работы [Березина, 2011] с учетом различия выборок и инструментария обработки анкет.

Уровень вариативности в вычисляемых матрицах «запахи – эмоции» между рассмотренными языковыми моделями соответствует, например, ситуации совпадения выявляемой выраженной эмоции, но расхождения в определении умеренно выраженной эмоции.

Инструментарии анализа качественных данных

В настоящее время разработано множество стандартизированных инструментов психологического тестирования, для интерпретации которых достаточно простой арифметики, и как таковых методов искусственного интеллекта не требуется [Gomez, 2024]. Вместе с тем для анализа результатов опросов, содержащих «открытые» вопросы, необходимо привлечение более универсальных инструментов. Основными решаемыми задачами при обработке результатов подобных опросов являются следующие: категоризация ответов в открытом тексте и анализ результатов [O’Kane, 2021]. Технически при этом возникает необходимость обработки большого количества данных (как правило, текстовых), которые требуется проанализировать, не имея готового алгоритма. Соответственно, усилия в качественном анализе сосредоточены на разработке методов, позволяющих анализировать и выявлять закономерности в текстовых данных. Некоторые из подобных методов направлены на выявление, анализ и представление так называемых тем в данных, например: обоснованная теория, нарративный анализ [Corbin, Strauss, 1990] или тематический анализ [Braun, Clarke, 2006].

Несмотря на то что для анализа качественных данных на протяжении более 30 лет разрабатывается специализированное программное обеспечение, такое как NVivo, Atlas.Ti, MAXQDA (https://en.wikipedia.org/wiki/Computer-assisted_qualitative_data_analysis_software), в некотором смысле подобные решения относятся к предыдущему поколению инструментов.

Появившиеся недавно БЯМ позволяют работать с данными и знаниями без необходимости разработки узкоспециализированного инструмента. Современные же версии программного обеспечения анализа качественных данных либо пытаются интегрировать элементы искусственного интеллекта, либо используют программный вызов готовых языковых моделей из своего интерфейса.

Заключение

В работе выполнена оценка возможностей перспективного направления в применении БЯМ, а именно обработки результатов анкетирования, содержащего «открытые вопросы». Для этого выполнены тестовые примеры построения матрицы связи базовых ароматов и вызывающих их эмоций методом анкетирования испытуемых с применением больших языковых моделей, отвечающих заданным бенчмаркам [Белоус, Бойкова, 2025]. Результаты обработки качественно согласуются как между различными языковыми моделями, так и с выводами оригинальной методики. Что позволяет сформулировать гипотезу о сходимости результатов обработки анкет большими языковыми моделями при условии достижения моделями определенного порога качества. Полученные результаты также могут быть использованы при пополнении базы данных проекта по изучению запахов [Hamel, 2024].

При описании алгоритма расшифровки эмоциональных реакций в ответах языковые модели сформулировали ряд новых правил качественного анализа, а также указали на способность идентифицировать и учитывать индивидуальные особенности анкетиремого для более точной обработки результатов, что имеет практическое значение для учета специфики ольфакторного культурного кода, в том числе в сфере межкультурной коммуникации.

Полученные в работе результаты указывают на перспективность предложенной методологии применения больших языковых моделей для решения задач интерпретации свободных формулировок, включая обработку результатов анкетирования, содержащего «открытые во-

просы», при наличии достаточно объемного контекста. С появлением новых версий больших языковых моделей эффективность данного подхода, по всей видимости, будет возрастать.

Список литературы

- Белоус А. А., Бойкова А. П.** Оценка качества больших языковых моделей: лингвистический подход // Глобальный научный потенциал. 2025. № 3-2 (168). С. 207–211.
- Березина Т. Н.** Возникновение позитивных и негативных базовых эмоций под влиянием базовых запахов // Вестник МГГУ им. М. А. Шолохова. Серия «Педагогика и психология». 2011. № 3. С. 59–69.
- Голубинский А. Н., Толстых А. А., Толстых М. Ю.** Автоматическая генерация аннотаций научных статей на основе больших языковых моделей // Информатика и автоматизация. 2025. Т. 24, № 1. С. 275–301.
- Кочкаров Р. А., Пойкалайнен А. М.** Применение больших языковых моделей и технологий машинного перевода в здравоохранении // Нейрокомпьютеры: разработка, применение. 2024. Т. 26, № 6. С. 31–38.
- Оберемко О. А.** К типологии открытых вопросов // Мониторинг общественного мнения: экономические и социальные перемены. 2018. № 4 (146). С. 97–108.
- Олицкая Д. А., Черткова В. В.** Ольфакторный код в оригинале и переводах рассказов А. П. Чехова на английский язык // Филологические науки. Вопросы теории и практики. 2020. Т. 13, № 5. С. 289–294.
- Пак Сон Г.** Запахи в различных языках и культурах. Попытка сопоставительного анализа (на материале корейских, японских и русских фразеологизмов) // Язык, сознание, коммуникация: сб. науч. ст. / под ред. В. В. Красных, А. И. Изотова, В. Г. Кульпиной. М., 2009. Вып. 37. С. 30–40.
- Паленова В. В., Воронин А. Н.** Применение больших языковых моделей в создании и адаптации текстов: потенциал и вызовы ChatGPT // Тенденции развития науки и образования. 2024. № 115-3. С. 30–34.
- Юрченко Т. Г.** Ольфакторное пространство русской литературы: проблемы изучения. Часть 1. Поэзия // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 7: Литературоведение. 2021а. № 3. С. 56–68.
- Юрченко Т. Г.** Ольфакторное пространство русской литературы: проблемы изучения. Часть 2. Проза и драматургия // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 7: Литературоведение. 2021б. № 4. С. 7–21.
- Acree T., Arn H.** Flavornet and human odor space. 2024. URL: <https://www.flavornet.org/> (дата обращения: 20.01.2025).
- Amoore J.** Stereochemical Theory of Olfaction // Nature. 1963. Vol. 198. P. 271–272.
- Berezina T. N.** Research of the Psychological Interconnection between the Basic Odors and the Emotional State of a Human Being // Asian Social Science. 2014. Vol. 10, No. 21. P. 159–167.
- Braun V., Clarke V.** Using thematic analysis in psychology // Qualitative research in psychology. 2006. Vol. 3, No. 2. P. 77–101.
- Buchanan T. W., Tranel D., Adolphs R.** A specific role for the human amygdala in olfactory memory // Learn. Mem. 2003. Vol. 10, No. 5. P. 319–325.
- Castleberry A., Nolen A.** Thematic Analysis of Qualitative Research Data: Is It as Easy as It Sounds? // Currents in Pharmacy Teaching and Learning. 2018. Vol. 10. P. 807–815.
- Corbin J. M., Strauss A.** Grounded theory research: Procedures, canons, and evaluative criteria // Qualitative sociology. 1990. Vol. 13, No. 1. P. 3–21.
- Ekman P., Cordaro D.** What is Meant by Calling Emotions Basic // Emotion Review. 2011. Vol. 3, No. 4. P. 364–370.

- Gomez R., Karimi L., Hein K., Papapetrou S., Stavropoulos V.** Symptom Checklist-90-Revised: Factor structure and relevance as a measure for the Hierarchical Taxonomy of Psychopathology model // *International Journal of Mental Health*. 2024. P. 1–24.
- Hamel E. A., Castro J. B., Gould T. J., Pellegrino R., Liang Z., Coleman L. A., Patel F., Wallace D. S., Bhatnagar T., Mainland J. D., Gerkin R. C.** Pyrfume: A window to the world's olfactory data // *Sci Data*. 2024. Vol. 11. P. 1220.
- Harada Y., Maeda S., Shen J., Misonou T., Hori H., Nakamura S.** Regression Study of Odorant Chemical Space, Molecular Structural Diversity, and Natural Language Description // *ACS Omega*. 2024. Vol. 9, No. 23. P. 25054–25062.
- Nazi Z. A., Peng W.** Large Language Models in Healthcare and Medical Domain: A Review // *Informatics*. 2024. Vol. 11. P. 57.
- Nezhurina M., Cipolina-Kun L., Cherti M., Jitsev J.** Alice in wonderland: Simple tasks showing complete reasoning breakdown in state-of-the-art large language models. 2024. arXiv: 2406.02061.
- Metelkina-Fernandez V., Dumas L.-E., Vandersteen C., Chirio D., Gros A., Fernandez A., Askenazy F., Manera V.** Psychiatric outcomes in outpatients affected by long COVID: A link between mental health and persistence of olfactory complaint // *World Journal of Psychiatry*. 2024. Vol. 14. P. 507–512.
- O'Kane P., Smith A., Lerman M. P.** Building transparency and trustworthiness in inductive research through computer-aided qualitative data analysis software // *Organizational Research Methods*. 2021. Vol. 24, No. 1 P. 104–139.
- Reisenzein R., Junge M.** Measuring the intensity of emotions // *Frontiers in Psychology*. 2024. Vol. 15. Article 1437843 : URL: <https://doi.org/10.3389/fpsyg.2024.1437843> (дата обращения: 04.08.2025).
- Root C. M., Denny C. A., Hen R., Axel R.** The participation of cortical amygdala in innate, odour-driven behaviour // *Nature*. 2014. Vol. 515. P. 269–273.
- Sorokowski P., Misiak M., Roberts S. C., Kowal M., Butovskaya M., Omar-Fauzee M. S., Huanca T., Sorokowska A.** Is the perception of odour pleasantness shared across cultures and ecological conditions? Evidence from Amazonia, East Africa, New Guinea, Malaysia and Poland // *Biol. Lett.* 2024. Vol. 20. P. 20240120.
- Taylor S. J., Bogdan R., DeVault M.** Introduction to qualitative research methods: A guidebook and resource. John Wiley & Sons, 2015. 413p.
- Wang H., Wang Q., Cui L., Feng X., Dong P., Tan L., Lin L., Lian H., Cao S., Huang H., Cao P., Li X. M.** A molecularly defined amygdala-independent tetra-synaptic forebrain-to-hindbrain pathway for odor-driven innate fear and anxiety // *Nat. Neurosci.* 2024. Vol. 27. P. 514–526.
- Zald D. H., Pardo J. V.** Emotion, olfaction, and the human amygdala: amygdala activation during aversive olfactory stimulation // *Proc. Natl. Acad. Sci. U. S. A.* 1997. Vol. 94, No. 8. P. 4119–4124.

References

- Acree T., Arn H.** Flavornet and human odor space. 2024. URL: <https://www.flavornet.org/>
- Amoore J.** Stereochemical Theory of Olfaction. *Nature*, 1963, vol. 198, pp. 271–272.
- Belous A. A., Boikova A. P.** Evaluation of Large Language Models Quality: Linguistic Approach. *Global Scientific Potential*, 2025, no. 3-2(168), pp. 207–211. (in Russ.)
- Berezina T. N.** Emergence of positive and negative basic emotions under the influence of basic odors. *Bulletin of Sholokhov Moscow State University for Humanities. Series: Pedagogy and Psychology*, 2011, no 3, pp. 59–69. (in Russ.)
- Berezina T. N.** Research of the Psychological Interconnection between the Basic Odors and the Emotional State of a Human Being. *Asian Social Science*, 2014, vol. 10, no. 21, pp. 159–167.

- Braun V., Clarke V.** Using thematic analysis in psychology. *Qualitative research in psychology*, 2006, vol. 3, no. 2, pp. 77–101.
- Buchanan T. W., Tranel D., Adolphs R.** A specific role for the human amygdala in olfactory memory. *Learn. Mem.*, 2003, vol. 10, no. 5, pp. 319–325.
- Castleberry A., Nolen A.** Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning*, 2018, vol. 10, pp. 807–815.
- Corbin J. M., Strauss A.** Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, 1990, vol. 13, no. 1, pp. 3–21.
- Ekman P., Cordaro D.** What is meant by calling emotions basic. *Emotion Review*, 2011, vol. 3, no. 4, pp. 364–370.
- Golubinskiy, A. N., Tolstykh, A. A., Tolstykh, M. Yu.** Automatic generation of scientific articles abstracts based on large language models. *Informatics and Automation*, 2025, vol. 24, no 1, pp. 275–301. (in Russ.)
- Gomez R., Karimi L., Hein K., Papapetrou S., Stavropoulos V.** Symptom Checklist-90-Revised: Factor structure and relevance as a measure for the Hierarchical Taxonomy of Psychopathology model. *International Journal of Mental Health*, 2024, pp. 1–24.
- Hamel E. A., Castro J. B., Gould T. J., Pellegrino R., Liang Z., Coleman L. A., Patel F., Wallace D. S., Bhatnagar T., Mainland J. D., Gerkin R. C.** Pyrfume: A window to the world's olfactory data. *Sci Data*, 2024, vol. 11, p. 1220.
- Harada Y., Maeda S., Shen J., Misonou T., Hori H., Nakamura S.** Regression Study of Odorant Chemical Space, Molecular Structural Diversity, and Natural Language Description. *ACS Omega*, 2024, vol. 9, no. 23, pp. 25054–25062.
- Kochkarov R. A., Poykalaynen A. M.** Application of large language models and machine translation technologies in healthcare. *Neurocomputers: development, application*, 2024, vol. 26, no 6, pp. 31–38. (in Russ.)
- Metelkina-Fernandez V., Dumas L.-E., Vandersteen C., Chirio D., Gros A., Fernandez A., Askenazy F., Manera V.** Psychiatric outcomes in outpatients affected by long COVID: A link between mental health and persistence of olfactory complaint. *World Journal of Psychiatry*, 2024, vol. 14, pp. 507–512.
- Nazi Z. A., Peng W.** Large Language Models in Healthcare and Medical Domain: A Review. *Informatics*, 2024, vol. 11, p. 57.
- Nezhurina M., Cipolina-Kun L., Cherti M., Jitsev J.** Alice in wonderland: Simple tasks showing complete reasoning breakdown in state-of-the-art large language models. 2024. arXiv: 2406.02061.
- Oberemko O. A.** On typology of open-ended questions. *Monitoring of public opinion: economic and social changes*, 2018, no 4 (146), pp. 97–108. (in Russ.)
- O'Kane P., Smith A., Lerman M. P.** Building transparency and trustworthiness in inductive research through computer-aided qualitative data analysis software. *Organizational Research Methods*, 2021, vol. 24, no. 1, pp. 104–139.
- Olitskaya D. A., Chertkova V. V.** Olfactory code in the Russian and English versions of A. P. Chekhov's stories. *Philological Sciences. Theory and Practice*, 2020, vol. 13, no 5, pp. 289–294. (in Russ.)
- Pak Son G.** Odors in different languages and cultures. An attempt of comparative analysis (on the material of Korean, Japanese and Russian idioms). *Language, consciousness, communication: Collection of scientific articles*. Edited by V. V. Krasnykh, A. I. Izotov, V. G. Kulpina. Moscow, 2009, Issue 37, pp. 30–40. (in Russ.)
- Palenova V. V., Voronin A. N.** Application of large language models in creation and adaptation of texts: potential and challenges of ChatGPT. *Tendencies of Science and Education Development*, 2024, no 115-3, pp. 30–34. (in Russ.)

- Reisenzein R., Junge M.** Measuring the intensity of emotions. *Frontiers in Psychology*, 2024, vol. 15, Article 1437843. URL: <https://doi.org/10.3389/fpsyg.2024.1437843>.
- Root C. M., Denny C. A., Hen R., Axel R.** The participation of cortical amygdala in innate, odour-driven behaviour. *Nature*, 2014, vol. 515, pp. 269–273.
- Sorokowski P., Misiak M., Roberts S. C., Kowal M., Butovskaya M., Omar-Fauzee M. S., Huanca T., Sorokowska A.** Is the perception of odour pleasantness shared across cultures and ecological conditions? Evidence from Amazonia, East Africa, New Guinea, Malaysia and Poland. *Biol. Lett.*, 2024, vol. 20, p. 20240120.
- Taylor S. J., Bogdan R., DeVault M.** Introduction to qualitative research methods: A guidebook and resource. John Wiley & Sons, 2015. 413 p.
- Wang H., Wang Q., Cui L., Feng X., Dong P., Tan L., Lin L., Lian H., Cao S., Huang H., Cao P., Li X.M.** A molecularly defined amygdala-independent tetra-synaptic forebrain-to-hindbrain pathway for odor-driven innate fear and anxiety. *Nat. Neurosci.*, 2024, vol. 27, pp. 514–526.
- Yurchenko T. G.** The olfactory space of Russian literature: problems of research. Part 1. Poetry. *Social Sciences and Humanities. Domestic and Foreign Literature. Series 7: Literary Studies*, 2021a, no 3, pp. 56–68. (in Russ.)
- Yurchenko T. G.** The olfactory space of Russian literature as a research problem. Part 2. Prose and drama. *Social Sciences and Humanities. Domestic and Foreign Literature. Series 7: Literary Studies*, 2021b, no 4, pp. 7–21. (in Russ.)
- Zald D. H., Pardo J. V.** Emotion, olfaction, and the human amygdala: amygdala activation during aversive olfactory stimulation. *Proc. Natl. Acad. Sci. U. S. A.*, 1997, vol. 94, no. 8, pp. 4119–4124.

Информация об авторах

Белоус Анастасия Александровна, кандидат филологических наук, доцент кафедры русского языка Института мировой медицины ФГАОУ ВО РНИМУ им. Н. И. Пирогова Минздрава России

Бойкова Анна Петровна, специалист ООО «Арктический научный центр»

Information about the Authors

Anastasiya A. Belous, Candidate of Philological Sciences, Associate Professor at the Department of Russian Language, Institute of Global Medicine Russian National Research Medical University (Pirogov Medical University)

Anna P. Boikova, Specialist, Arctic Research Centre LLC

Статья поступила в редакцию 23.02.2025;
одобрена после рецензирования 13.09.2025; принята к публикации 26.09.2025

The article was submitted 23.02.2025;
approved after reviewing 13.09.2025; accepted for publication 26.09.2025

Научная статья

УДК 81'33, 82-1: 81'322.4

DOI 10.25205/1818-7935-2026-24-2-69-82

Когерентность и когезия поэтического текста при машинном нейросетевом переводе

Андрей Владимирович Гребенюк

Санкт-Петербургский государственный университет
Санкт-Петербург, Россия

yax-light@yandex.ru, <https://orcid.org/0009-0008-8855-2657>

Аннотация

В статье рассматриваются особенности средств связности и цельности поэтических русскоязычных текстов и связанная с этой спецификой проблема сохранения связности и цельности данных текстов при их машинном нейросетевом переводе на английский язык. Актуальность работы обусловлена спецификой ошибок, совершаемых системами автоматизированного перевода по сравнению с ошибками, возникающими при традиционном, «человеческом», переводе, а также новизной и, следовательно, малой изученностью машинного перевода поэзии в целом. В теоретическом разделе обсуждаются основные категории текста – связность (когезия) и цельность (когерентность), а также особенности их реализации в случае поэтических текстов. Делается вывод о более низкой степени связности поэтических текстов, наличии специфической информационной структуры, преобладании сочинительной связи внутри предложения, об исключительной важности лексического повтора и сквозного использования той или иной стилистической фигуры для поддержания цельности таких текстов. В практическом разделе выполнен машинный нейросетевой перевод шести поэтических произведений Серебряного века с помощью сервиса «Яндекс Переводчик» с целью создания подстрочного перевода. Далее построчно рассматриваются задействованные средства связности в исходных текстах и те из них, которые отсутствуют в текстах перевода. Была продемонстрирована справедливость основных положений теоретического раздела о специфике средств связности поэтического текста, а также установлено, что при машинном переводе наиболее частотными являются именно те ошибки, которые в наибольшей степени ответственны за снижение степени связности и цельности текста, – ошибки, связанные с передачей повторяющихся лексем (а также отдельных сем и морфем) и стилистических фигур. Указывается на специфичность ошибок этого типа в первую очередь для автоматического перевода, а также обсуждается их значение в контексте прикладной задачи поэтического перевода на базе автоматически сгенерированного подстрочника.

Ключевые слова

цельность, связность, когерентность, когезия, поэтический текст, машинный перевод

Для цитирования

Гребенюк А. В. Когерентность и когезия поэтического текста при машинном нейросетевом переводе // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 69–82. DOI 10.25205/1818-7935-2026-24-2-69-82

© Гребенюк А. В., 2026

Cohesion and Coherence of Poetic Text in Neural Machine Translation

Andrey V. Grebenyuk

University of St Petersburg
Saint-Petersburg, Russian Federation

yax-light@yandex.ru, <https://orcid.org/0009-0008-8855-2657>

Abstract

The presented article examines the features of coherence and cohesion means in Russian poetry and the related issue of preserving coherence and cohesion of these texts in their neural machine translation into English. The relevance of the work is determined by the specifics of the errors made by automatic translation systems in comparison with the errors occurring in traditional 'human' translation as well as by the novelty and insufficient study of machine translation of poetry. The theoretical section introduces the main categories of the text – cohesion and coherence – and discusses their specifics in poetry. The conclusions made are: a lower extent of coherence and cohesion in poetic texts, specific information structure, and the predominance of coordinating conjunctions within a sentence. Also demonstrated is the exceptional importance of such leitmotifs as lexical and synonymous repetitions and the cross-cutting use of one or more stylistic devices for preserving the coherence of poetic texts. In the second section, a neural network machine translation of six poetic works belonging to the Silver Age was done using the 'Yandex Translate' service. Next, the means of coherence and cohesion employed in the original texts and those that are absent in the translation texts were considered line by line. Thus, the validity of the main conclusions on the specifics of coherence and cohesion means in poetic texts is demonstrated. It was also found that in machine translation, the most frequent errors are those mainly responsible for reducing coherence and cohesion of the text: errors associated with the transmission of various leitmotifs (repetitive lexemes as well as individual semantic components, morphemes and stylistic devices). The specificity of these errors for automatic translation and their significance in the context of poetic translation based on automatically generated interlinear translation were discussed.

Keywords

cohesion, coherence, discourse analysis, poetic text, machine translation

For citation

Grebenyuk A. V. Cohesion and coherence of poetic text in neural machine translation. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 69–82. DOI 10.25205/1818-7935-2026-24-2-69-82

Введение

Под термином «текст» понимают последовательность языковых знаков, обладающую свойствами связности, целостности и завершенности: «единицу коммуникации, которой свойственна относительная смысловая завершенность» [Мурзин, 1991. С. 6–10].

Художественный текст в то же время представляет собой репрезентацию не какой-то реальной коммуникативной ситуации и связанного с ней дискурса, а действительности, преломленной через сознание и мировосприятие автора текста [Купина, 2009. С. 259]. И тем не менее автор использует те же языковые средства, что и для описания реальности, благодаря чему читатель устанавливает аналогии с известной ему действительностью. Художественный текст, являясь одновременно текстом в общелингвистическом смысле и результатом вымысла, отражает взаимосвязь языка как инструмента передачи информации в реальных коммуникативных ситуациях, с одной стороны, и художественной литературы – с другой.

Категории и уровни художественного текста

Основные категории текста должны отражать тесную взаимосвязь всех его компонентов, отличать текст от простого набора фраз. Такими категориями обычно признаются связность (когезия) и цельность (когерентность) [Леонтьев, 1974. С. 168–172].

Когезия представляет собой совокупность связей, обеспечивающих логическую состоятельность текста, а также взаимосвязь действий, высказываний во времени и пространстве.

М. Хэллидей и Р. Хасан утверждали, что «отношениями когезии являются отношения между двумя и более элементами в тексте, независимо от их структурной оформленности» [Halliday, 1976. С. 109]. В качестве основных средств когезии ими были выделены следующие.

1. Референция – самый простой и наиболее универсальный способ обеспечения связности текста за счет замены определённых слов детерминантами, в первую очередь местоимениями и артиклями.

2. Субституция является заменой местоимениями целых словосочетаний и предложений.

3. Эллипсис как средство обеспечения связности текста, наиболее характерен для разговорной речи и представляет собой намеренное опущение слова или словосочетания, которое при этом легко может быть восстановлено адресатом из контекста.

4. Использование коннекторов – соединение частей текста с помощью сочинительных и подчинительных союзов и вводных конструкций. Данное средство связности является наиболее лексически разнообразным.

5. Лексическая когезия заключается в повторе лексических единиц. Это достаточно обширная группа средств включает в себя синонимический, антонимический повтор, повтор однокоренных слов, слов, связанных родовидовыми отношениями гипонимии и гиперонимии, соотнесение части и целого и ассоциативный повтор.

Достаточно близкую классификацию предлагают Р. Богранд и В. Дресслер [Тичер, 2017. С. 39]: повторение, анафора и катафора, эллипсис и конъюнкция. Также авторы различали когезию (локальную связность) и когерентность (глобальную связность или цельность), которая сохраняет текст как смысловое целое и, в значительной степени, достигается сквозным повторением определенных элементов: сем, морфем, лексем, стилистических фигур. В тексте, не являющимся цельным, невозможно установить логические связи между отдельными его фрагментами и понять его прагматическую функцию. Приведем цитату А. А. Леонтьева о соотношении понятий связности и цельности: «...связность обычно является условием цельности, но цельность не может полностью определяться через связность. С другой стороны, связный текст не всегда обладает характеристикой цельности» [Леонтьев, 1974. С. 54].

Особенности поэтических текстов

Поэзия и проза являются двумя различными видами художественной речи. При этом существуют различные подходы к их разграничению. В работе, посвященной синтаксической связности поэтического текста, Б. Ю. Норман отмечает принципиальную важность такого структурного аспекта, как строкоделение: «...в поэтическом произведении может не быть косвенных падежей, может не быть предлогов и союзов, может не быть знаков препинания и других стандартных средств структурной организации, но либо ритм и рифма, либо, на худой конец, деление на строки должно быть!» [Норман, 2019; см. также Скулачева, 2020]. Говоря же о лексико-семантическом аспекте поэтической речи, автор указывает, что более низкая степень синтаксической связности поэтического текста (измеренная как доля служебных слов в тексте) компенсируется другими, неграмматическими, средствами достижения связности и целостности: фонетической организацией текста (отсюда следует важность деления на строки) и лексической (в первую очередь за счет большей актуализации ассоциативных связей лексем): «...мы наблюдаем действие компенсаторных механизмов многоуровневой природы языка, которые в конечном счете и обеспечивают тексту должную цельность» [Норман, 2019; см. также Rifqi, 2009].

Вопросы связности поэтического текста исследовались в работе В. И. Заики и Г. Н. Гиржева. На материале стихотворения В. Набокова «В неволе я, в неволе я, в неволе» делают вывод о ключевой роли повтора (в первую очередь лексического) и анжамбемана (переноса) в обеспечении связности поэтического текста: первый прием повышает связность, второй, являясь зна-

чительным выразительным средством, наоборот, разрушает ее [Заика, 2023]. Семантико-стилистическим средствам связности посвящена статья Л. А. Моря [Моря, 2020], где на материале стихотворения Б. Пастернака «Январь 1919 года» демонстрируется осуществление связности поэтического текста за счет сквозных стилистических фигур (олицетворение и антитеза). В работе [Дрейзис, 2016] показывается, как в условиях «запрета» на лексический повтор основным механизмом, обеспечивающим связность в китайской поэзии, является параллелизм.

В статье Т. В. Скулачевой и А. Э. Костюк указывается на преобладание сочинительной связи, что затрудняет логический анализ текста [Скулачева, 2020]. Также авторы отмечают намеренную многозначность на уровне лексической семантики (и даже намеренные логические ошибки, например взаимоисключающая предикация) и отличное от прозаического тема-рематическое членение. Другими особенностями поэтического текста могут являться намеренная синтаксическая усложненность и неоднозначность, устранение и неконвенциональное использование знаков препинания и заглавных букв (и даже специальных символов), а также использование таких графических средств, как деление текста на строфы и квазистрофы [Бочавер, 2016].

Приведенная краткая характеристика поэтической речи показывает, что поэзия обладает специфической организацией на всех уровнях языковой системы (фонетическом, грамматическом и лексико-семантическом): выраженной ритмической структурой, богатством непрямых значений и намеренной многозначностью, малоупотребительной и архаичной лексикой, окказионализмами, аграмматизмами, а также специфическими синтаксическими конструкциями и средствами когезии. При этом особенности данной организации, с одной стороны, способствуют активизации образного мышления, с другой – затрудняют логический анализ поэзии [Скулачева, 2020].

Анализ машинного перевода поэтических текстов

Описанные выше особенности поэзии означают, что поэтический текст представляет собой серьезный вызов для современных систем машинного перевода. Тем не менее уже существует ряд работ, исследующих нейросетевой перевод поэзии [Dunder, 2020; Humblé, 2019; Ghazvininejad, 2018]. В частности, в работах [Гребенюк, 2024; Гребенюк, 2025] изучались характерные ошибки, возникающие при данном виде перевода: была предложена классификация подобных ошибок, где отдельной категорией выделялись ошибки, связанные с лексическим повтором. В настоящей публикации мы подробнее остановимся на вопросе сохранения лексической связности при машинном переводе.

Также отметим, что, помимо теоретического интереса, связанного с исследованием границы возможного для нейросетей, лежащих в основе современных систем автоматического перевода, машинный перевод поэзии представляет и практический интерес – автоматическая генерация подстрочника для последующего поэтического перевода человеком [Гребенюк, 2025]. Укажем, что речь идет именно о так называемом информационном переводе: передача, например, звуковой формы поэтической речи пока остается за рамками возможностей существующих систем перевода.

В качестве материала исследования были выбраны шесть поэтических произведений: «А если нет» и «Поэза странностей жизни» И. Северянина, «Нежнее нежного» и «Ленинград» О. Мандельштама, а также «Последний тост» и «Я не знаю, ты жив или умер...» А. Ахматовой. В качестве системы перевода был выбран «Яндекс Переводчик» как наиболее популярный сервис в России в паре русский/английский [Гребенюк, 2025]. Методология исследования связности поэтического текста воспроизводит предложенную в [Rifqi, 2009]: построчное указание вида средства связности (по приведенной выше классификации М. Халлидея) и строк, к которым данное средство относится. При анализе переводного текста указываются только строки,

в которых отсутствуют соответствующие средства связности и тем самым происходит нарушение связности относительно оригинального текста. В табл. 1 представлен оригинальный текст первого стихотворения и его машинный перевод.

Таблица 1

Перевод стихотворения И. Северянина «А если нет»

Table 1

The translation of 'And what if not...' by I. Severyanin

L1	А если нет?.. А если ты ушла,	And if not?.. What if you left
L2	Чтоб не прийти ко мне на панихиды?	so you wouldn't come to my memorial service?
L3	Кто даст ответ?	Who will give the answer?
L4	Одна лишь ты могла,	You alone could,
L5	Но ты полна обиды...	but you're full of resentment....
L6	А если нет?	And if not?
L7	Какая грусть... Как мраморна печаль...	What a sadness... How marble is sadness...
L8	Как высока, свята и вдохновенна!	How high, holy and inspired!
L9	Но пусть, но пусть.	But let it be, but let it be.
L10	Разбитая скрижаль —	A broken tablet
L11	Осколком драгоценна...	is precious...
L12	Какая грусть!	What a sadness!

В исходном русскоязычном тексте используются следующие средства связности:

- L1. Лексическая связность: повтор словосочетания *а если*;
- L2. Использование коннектора: подчинительный союз *чтобы*, вводящий придаточное цели к первой строке; синтаксический параллелизм: повтор вопросительного предложения;
- L3. Синтаксический параллелизм: повтор вопросительного предложения; лексическая когезия: лексема *ответ* отсылает к предыдущим вопросительным конструкциям;
- L4. Референция: повтор местоимения *ты*, отсылающего к L1 и эллипсис, отсылающий к L3;
- L5. Использование коннектора: сочинительная связь (конъюнкция) с L4 союзом *но*; повтор местоимения *ты*;
- L6. Лексическая когезия: повтор предложения *А если нет?*, отсылающего к L1;
- L7. Лексическая когезия: повтор морфемы *как*, синонимический повтор *грусть*, *печаль*; синтаксический параллелизм: повтор назывного предложения;
- L8. Лексическая когезия: повтор морфемы *как*, отсылающий к L7;
- L9. Использование коннектора: противопоставление L7 союзом *но*; лексический повтор *пусть*;
- L10–L11. Лексическая когезия: повтор семы «разрушение» общей для *разбитая* и *осколком*;
- L12. Лексическая когезия: повтор морфемы *как*; синтаксический параллелизм: повтор назывного предложения; отсылают к L7 и L8;

В переводном тексте отсутствуют следующие средства, связанные с лексическим повтором:

- L1. Повторяется не вся конструкция, а только слово *if*;
- L7. Отсутствует повтор морфемы (*какая/как* → *what a/how*), при этом имеет место следующее искажение: синонимический повтор передан как полноценный лексический повтор лексемы *sadness*;

L11. При переводе не передана лексема *осколок* и, следовательно, отсутствует повтор соответствующей семы.

В качестве второго примера была выбрана «Поэза странностей жизни» И. Северянина (табл. 2).

Таблица 2

Перевод стихотворения И. Северянина «Поэза странностей жизни»

Table 2

The translation of 'Poetry of the oddities of life' by I. Severyanin

L1	Встречаются, чтоб разлучаться...	They meet to separate...
L2	Влюбляются, чтобы разлюбить...	they fall in love to fall out of love...
L3	Мне хочется расхохотаться,	I want to burst out laughing,
L4	И разрыдаться – и не жить!	and burst into tears, and not live!
L5	Клянутся, чтоб нарушить клятвы...	They swear to break their vows...
L6	Мечтают, чтоб проклянуть мечты...	They dream to curse their dreams...
L7	О, скорбь тому, кому понятны	Oh, sorrow to those who understand
L8	Все наслаждения тщетны!..	that all pleasures are in vain!..
L9	В деревне хочется столицы...	The village wants a capital...
L10	В столице хочется глуши...	In the capital, you want the wilderness...
L11	И всюду чело́вечьи лица	And everywhere there are human faces
L12	Без человеческой души...	without a human soul...
L13	Как часто красота уродна	How often is beauty ugly,
L14	И есть в уродстве красота...	and there is beauty in ugliness...
L15	Как часто низость благородна	How often baseness is noble
L16	И злы невинные уста.	and innocent lips are evil.
L17	Так как же не расхохотаться,	So how can you not burst out laughing,
L18	Не разрыдаться, как же жить,	how can you not burst into tears, how can you
L19	Когда возможно расставаться,	live
L20	Когда возможно разлюбить?!	when it is possible to break up,
		when it is possible to stop loving?!

L1–L2. Синтаксический параллелизм и лексическая когезия: повтор союза *чтобы* и морфемы *раз-*;

L3–L4. Использование коннектора: сочинительная связь (конъюнкция) с помощью союза *и*; повтор корня *люб* и морфемы *раз-*, отсылающей также к L1 и L2;

L5–L6. Синтаксический параллелизм и лексическая когезия: повтор родственных корней *клят/кля* и корня *мечт*;

L7–L8. Референция: местоимение *тому* указывает на агенс в L5–L6;

L9–L10. Синтаксический параллелизм и лексическая когезия: повтор слоговой формы *хочется* и лексемы *столица*;

L11–L12. Референция: наречие *всюду* указывает на обстоятельства места в L9–L10; повтор корня *человеч*;

L13–L14. Конъюнкция: сочинительная связь с помощью союза *и*; повтор корня *урод* и слоговой формы *красота*; повтор стилистической фигуры оксюморон;

- L15–L16. Синтаксический параллелизм с L13; сочинительная связь с помощью союза *и*; повтор стилистической фигуры оксюморон (так же отсылает к L13–L14);
- L17–L18. Частичный синтаксический параллелизм и повтор частицы *не*, двукратный повтор морфемы *раз-*, отсылающей к L1–L4;
- L19–L20. Синтаксический хиазм и повтор словосочетания *когда возможно*; повтор морфемы *раз-*, отсылающей также к L17–L18 и L1–L4 и создающий кольцевую структуру в данном произведении.

Также отметим, что, помимо указанных средств связности, цельность произведения обеспечивается сквозным повтором в каждом катрене стилистических фигур синтаксического параллелизма и алогизма/оксюморона.

При переводе степень связности и цельности текста уменьшается:

- L1–L2, L17–L20. Отсутствует повтор единой морфемы (повтор сохраняется лишь частично *fall in/fall out* в L1–L2 и *burst into/burst out* в L3–L4 и L17–L18) и, как следствие, нарушение кольцевой композиции исходного текста;
- L5–L6. Отсутствие повтора корневой морфемы: *клянутся*, *клятвы* и *клянуть* переданы тремя разнокоренными лексемами *swear*, *vow*, *curse*;
- L9–L10. Ошибка анализа синтаксических связей (В деревне хочется столицы... → The village wants a capital...) и, как следствие, отсутствие синтаксического параллелизма;
- L13–L15. Частичное нарушение синтаксического параллелизма: наличие инверсии в L13 и отсутствие в L15.

В табл. 3 представлен оригинальный текст и машинный перевод стихотворения «Нежнее нежного» О. Мандельштама:

Таблица 3

Перевод стихотворения О. Мандельштама «Нежнее нежного»

Table 3

The translation of «More tender than tenderness» by O. Mandelstam

L1	Нежнее нежного	Your face is more tender
L2	Лицо твоё,	than a gentle one,
L3	Белее белого	your hand is
L4	Твоя рука,	whiter than white,
L5	От мира целого	you are far
L6	Ты далека,	from the whole world,
L7	И все твоё –	and everything is yours
L8	От неизбежного.	from the inevitable.
L9	От неизбежного	From the inevitable,
L10	Твоя печаль,	your sadness,
L11	И пальцы рук	and the fingers
L12	Неостывающих,	of your unceasing hands,
L13	И тихий звук	and the quiet sound of
L14	Неунывающих	cheerful
L15	Речей,	speeches,
L16	И даль	and the distance
L17	Твоих очей.	of your eyes.

В оригинальном тексте присутствуют следующие средства связности:

- L1–L2. Лексическая связность: повтор лексемы *нежный*;

- L3–L4. Лексическая связность: повтор лексемы *белый*; синтаксический параллелизм: *нежнее нежного / белее белого*; референция: притяжательное местоимение *твоя*, отсылающее к L2;
- L5–L10. Референция и лексический повтор: местоимения *ты, твое, твоя*, отсылающие к L2, повтор лексемы *неизбежный*; синтаксический параллелизм: трехкратный повтор предлога *от* в обстоятельстве причины;
- L11–L17. Синтаксический параллелизм: трехкратный повтор союза *и*; референция: *твоих*. В переводном тексте отсутствуют следующие лексические средства когезии:
- L1. Отсутствует лексический повтор (*нежнее нежного* → *more tender than gentle*);
- L3. Отсутствует синтаксический параллелизм с L1: сравнительные степени прилагательных переданы в одном случае с помощью аналитической формы, в другом – синтетической.

Так же к нарушению связности текста приводит ошибка распознавания синтаксических связей в L7: определитель в *все твое* распознан как именная часть сложного сказуемого: *everything is yours*.

В табл. 4 представлены исходный и переводной тексты стихотворения «Ленинград» О. Мандельштама.

Таблица 4

Перевод стихотворения О. Мандельштама «Ленинград»

Table 4

The translation of «Leningrad» by O. Mandelstam

L1	Я вернулся в мой город, знакомый до слез,	I returned to my city, familiar to the point of tears,
L2	До прожилок, до детских припухлых желез.	to the veins, to the swollen glands of children.
L3	Ты вернулся сюда, – так глотай же скорей	You came back here, so swallow the fish oil
L4	Рыбий жир ленинградских речных фонарей.	of the Leningrad river lanterns as soon as possible.
L5	Узнавай же скорее декабрьский денек,	Find out soon the December day
L6	Где к зловещему дегтю подмешан желток.	where the yolk is mixed with the ominous tar.
L7	Петербург, я еще не хочу умирать:	Petersburg, I don't want to die yet:
L8	У тебя телефонов моих номера.	you have my phone numbers.
L9	Петербург, у меня еще есть адреса,	Petersburg, I still have addresses
L10	По которым найду мертвецов голоса.	where I can find dead voices.
L11	Я на лестнице черной живу, и в висок	I live on a black staircase, and a bell
L12	Ударяет мне вырванный с мясом звонок.	ripped out with meat hits me in the temple.
L13	И всю ночь напролет жду гостей дорогих,	And all night long I wait for dear guests,
L14	Шевеля кандалами цепочек дверных.	moving the shackles of the door chains.

В исходном русскоязычном тексте используются следующие средства связности:

- L1–L2. Синтаксический параллелизм: трехкратное повторение предлога *до* в обстоятельстве степени; референция: местоимения *я*, *мой*;
- L3–L4. Лексическая когезия: *глотай* и *рыбий жир* объединены общей семой «еда», прилагательное *ленинградский* отсылает к лексеме *город* в L1, лексический повтор *вернулся*;
- L5–L6. Референция (эллипсис): глагол в повелительном наклонении единственного числа отсылает к L3; конъюнкция: союз *где* вводит придаточное места к L5;
- L7–L8. Лексическая когезия: *Петербург* отсылает к *ленинградский* и *город* в L1 и L4 соответственно; референция: *тебя* и *моих* в L8 указывает на *Петербург*, и *я* в L7, местоимение *я* указывает на лирического героя и отсылает к L1;
- L9–L10. Лексическая когезия: повтор *Петербург* и *еще*, отсылающие к L1, L4 и L7; *адреса* имеет общую сему «место жительства» с *номера* (L8); конъюнкция: *по которым* в L10 вводит придаточное предложение к L9; синтаксический параллелизм: обращение в начале строфы;
- L11–L12. Референция: местоимения *я* и *мне*; лексическая когезия: *лестнице черной* и *звонок* также принадлежат к семантическому полю «место жительства»;
- L13–L14. Использование коннектора, конъюнкция: сочинительная связь с L11–L12 посредством союза *и*; лексическая когезия: *гостей* и *цепочек дверных* также содержат сквозной для данного стихотворения компонент семантики;

В тексте машинного перевода отсутствует:

- L3. Лексический повтор (*вернулся/вернулся* → *returned/came back*);
- L5. Частично сохранен лексический повтор (*глотай же скорее/узнавай же скорей* → *swallow ... as soon as possible/find out soon*);
- L9. Лексический повтор (*еще/еще* → *yet/still*).

Также к нарушению лексической связности приводит буквальная (и, соответственно, неэквивалентная) передача фразеологических единиц: *лестнице черной*, *вырванный с мясом*.

Явление нарушения цельности и связности поэтического текста при неверной передаче синтаксического параллелизма и лексического повтора также хорошо иллюстрируется примером перевода стихотворения А. Ахматовой «Последний тост» (табл. 5).

Таблица 5

Перевод стихотворения А. Ахматовой «Последний тост»

Table 5

The translation of «The last toast» by A. Akhmatova

L1	Я пью за разоренный дом,	I drink to my ruined house,
L2	За злую жизнь мою,	to my evil life,
L3	За одиночество вдвоем,	to the loneliness of the two of us,
L4	И за тебя я пью, –	and to you I drink,
L5	За ложь меня предавших губ,	for the lies of my betrayed lips,
L6	За мертвый холод глаз,	for the dead cold of my eyes,
L7	За то, что мир жесток и груб,	for the fact that the world is cruel and rude, for the
L8	За то, что Бог не спас.	fact that God did not save.

Связность данного произведения обеспечивается синтаксическим параллелизмом (и хиазмом в L4 отсылающим к L1) и повторением предлога *за* в каждой из восьми строк, который отсылает к глаголу *пью* в L1 и L4. Нетрудно видеть, что текст перевода фактически распадается на две части: там, где предлог *за* переведен как *to* (L1–L4), и там, где он передан как *for* (L5–L6). Как следствие, нарушается не только узус английского языка (*to drink to something*),

но и локальная связность между L4 и L5, и цельность всего текста. Так же отметим другую часто возникающую ошибку при переводе поэтических текстов: в отличие от русского в английском языке существительные, как правило, требуют детерминатива, например артикля или притяжательного местоимения, и система автоматического перевода зачастую приписывает неверный детерминатив. Так, в L1, L5 и L6 в качестве детерминатива выбрано местоимение *ты*, и если в L1 один такой выбор еще может быть оправдан, то в L5 и L6 существительные *губы* и *глаза*, очевидно, относятся к адресату лирического героя (в L5 имеет место комплексная ошибка анализа синтаксических связей).

Вторым анализируемым произведением А. Ахматовой является стихотворение «Я не знаю, ты жив или умер...» (табл. 6).

Таблица 6

Перевод стихотворения А. Ахматовой «Я не знаю, ты жив или умер...»

Table 6

The translation of «I do not know if you are alive or dead...» by A. Akhmatova

L1	Я не знаю, ты жив или умер,	I do not know if you are dead or alive,
L2	На земле тебя можно искать	if you can be searched for on earth
L3	Или только в вечерней думе	or if you can only grieve
L4	По усопшем светло горевать.	for the deceased in the evening.
L5	Все тебе: и молитва дневная,	Everything is for you: the daily prayer,
L6	И бессонницы млеющий жар,	the burning heat of insomnia,
L7	И стихов моих белая стая,	the white flock of my poems,
L8	И очей моих синий пожар.	and the blue fire in my eyes.
L9	Мне никто сокровенней не был,	No one was more intimate to me,
L10	Так меня никто не томил,	no one tormented me so much,
L11	Даже тот, кто на муку предал,	not even the one who betrayed me to torment,
L12	Даже тот, кто ласкал и забыл.	even the one who caressed and forgot.

Здесь можно указать на следующие средства связности:

L1–L2. Референция и лексическая когезия: двукратный повтор местоимения *ты*;

L3–L4. Использование коннектора: сочинительный союз *или* (дизъюнктивная связь); лексическая когезия: лексема *усопший* связана с *умер* в L1 общей семой «смерть», *светло* антонимически противопоставляется *вечерней*;

L5–L8. Синтаксический параллелизм; конъюнкция и лексическая когезия: четырехкратный повтор союза *и* (стилистическая фигура полисиндетон), повтор корня *жар*, двукратный повтор местоимения *моих*, осуществляющий связность L7 и L8, и референциально отсылающий к *я* в L1; *тебе* отсылает к *ты* и *тебя* в L1–L2;

L9–L10. Референция и лексическая когезия: *мне* и *меня* реализуют связность внутри L9–L10 и отсылают к L1 и L7–L8, повтор местоимения *никто* и частицы *не*;

L11–L12. Синтаксический параллелизм и лексическая когезия: повтор *даже тот, кто*.

При переводе возникают следующие нарушения связности:

L4. Не переведена единица *светло*, противопоставляемая *вечерней* в L3. Также здесь неверно восстановлен агентс безличного предложения (*you*, т. е. вместо лирического героя его адресат);

L5–L8. Отсутствует общая корневая морфема *жар* (*жар* → *heat*, *пожар* → *fire*);

нарушен синтаксический параллелизм: союз *and* присутствует только в L8, в L6–L7 и L8 использованы различные предлоги, *of* и *in* соответственно;

- L11–L12. Нарушен синтаксический параллелизм: *not even the one who/ even the one who*. Отметим также, что здесь возникает отсутствующий в исходном тексте лексический повтор: *томил/муку* → *tormented/toment*.

Обсуждение результатов

Приведенный выше анализ исходных поэтических текстов подтверждает основные выводы, сделанные на этапе литературного обзора: набор средств, обеспечивающих связность и цельность поэтического текста, имеет свою специфику. Преобладание сочинительной связи, нестандартная информационная структура предложений, несовпадение границ строк и предложений/синтагм (явление анжамбемана) приводят к более низкой локальной связности текста. Сохранение же связности (как локальной, так и глобальной) обеспечивается в первую очередь лексическими средствами когезии: лексическим повтором и присутствием общих сем в несовпадающих лексических единицах. Также значимую роль играет сквозное использование определенных стилистических приемов.

Действительно, оставляя вне рассмотрения вопросы, связанные со звуковой организацией текста, на материале выбранных произведений мы видели, что основными видами связности являются:

1. Референция: в первую очередь личные и притяжательные местоимения, указывающие на лирического героя и его адресата;
2. Использование коннекторов: преимущественно сочинительная связь с помощью союза *и* (реже: *или, но*);
3. Стилистическая фигура синтаксического параллелизма;
4. Лексическая когезия: повтор стилистических фигур, словосочетаний, лексем или словоформ, а также отдельных морфем и сем.

Именно последние два вида средств связности являются наиболее частотными и, что самое важное, они зачастую охватывают более чем две строки, являясь тем самым сквозным средством обеспечения не только локальной связности, но и цельности всего текста. Но именно эти виды связности чаще всего и нарушаются при машинном переводе: синонимы передаются одной лексемой, одна лексема, наоборот, синонимами; различными способами передаются синтаксически сходные конструкции и т. д. Также необходимо указать на специфичность данных ошибок в первую очередь именно для машинного перевода: для человека, занимающегося поэтическим переводом, необходимость сохранения повторов в тексте перевода представляется очевидной и в большинстве случаев достаточно легко реализуемой (например, с использованием переводческого приема компенсации [Яковлева, 2008]).

Важно отметить, что при машинном переводе возникают и другие ошибки, в первую очередь проистекающие из лексико-грамматических особенностей поэтического текста и чаще всего связанные с неправильным выбором детерминатива и анализом синтаксических связей [Гребенюк, 2025]. Разумеется, данные ошибки также понижают локальную связность текста (вплоть до полной невозможности восстановить смысл исходного предложения). Вместе с тем они, именно в силу своей способности приводить к полному коммуникационному сбою, являются более очевидными для реципиента переводного текста и связаны с неверным анализом одного или нескольких соседних предложений, что в свою очередь позволяет надеяться на снижение частотности подобных ошибок по мере развития систем машинного перевода. Ошибки же, связанные с непередачей того или иного повтора, сохраняющие грамматичность текста и узус переводящего языка, с большой вероятностью останутся незаметными для реципиента и, например, могут не позволить воспроизвести соответствующий повтор при обработке перевода-подстрочника. Тот факт, что повторяющиеся элементы могут быть разделены значительным расстоянием в тексте (в том числе буквально всем прочим текстом произведения,

например, при кольцевой композиции), означает, что решение данной проблемы представляет серьезный вызов для существующих систем машинного перевода.

Таким образом, можно дать следующую методическую рекомендацию: при автоматической генерации подстрочника следует дополнительно проверить текст оригинала как минимум на наличие повторяющихся лексем и словосочетаний. Как правило, это оказывается возможным даже в случае недостаточного владения языком исходного текста. Возможным более фундаментальным решением описанной проблемы на данный момент представляется дообучение существующих больших языковых моделей на материале параллельных корпусов поэтических текстов.

Заключение

В настоящей работе была дана краткая характеристика ключевых категорий текста – связности и цельности, а также особенностей их реализации в поэтической речи. Ключевым выводом здесь является принципиально более низкая степень локальной связности (когезии) поэтического текста по сравнению с прозаическим и сохранение глобальной связности (цельности, когерентности) за счет сквозного повтора определенных лексем (или даже отдельных морфем), сем и стилистических фигур.

В практическом разделе были проанализированы средства связности и цельности исходных текстов и текстов, переведенных системой «Яндекс Переводчик». Было показано, что при машинном переводе в значительном количестве случаев нарушается повтор упомянутых выше элементов текста, что, в свою очередь, приводит к дальнейшему понижению связности и, еще более, цельности поэтического текста. Применительно к задаче автоматической генерации подстрочного перевода это означает необходимость «ручного» контроля наличия повторяющихся элементов в исходном тексте. Возможность дообучения существующих моделей с целью повышения качества поэтического перевода требует дальнейшего исследования.

Список литературы

- Бочавер С. Ю.** Неконвенциональная связность современного русского поэтического текста // Тр. Ин-та русского языка им. В. В. Виноградова. 2016. № 1 (7). С. 338–347.
- Гребенюк А. В.** Проблемы типологии ошибок машинного перевода: стремление к универсальности vs таргетированность // Вестник Московского университета. Серия 22: Теория перевода. 2024. № 2. С. 7–25.
- Гребенюк А. В., Митренина О. В.** Анализ распределения «прозаических» ошибок в машинном переводе поэтических текстов // Вестник СПбГУ. Язык и литература. 2025. № 3. С. 648–663.
- Дрейзис Ю. А.** Параллелизм классического типа как механизм связности в современной китайской поэзии // Культура и цивилизация. 2016. № 4. С. 192–202.
- Купина Н. А.** Массовая литература сегодня. М.: Наука, 2009. 423 с.
- Леонтьев А. А.** Признаки связности и цельности текста // Лингвистика текста. Материалы науч. конф. Ч. 1. М.: Изд-во МГУ, 1974. 230 с.
- Моря Л. А.** Семантико-стилистическая связность поэтического текста (на материале стихотворения Б. Пастернака «Январь 1919 года») // Гуманитарная парадигма. 2020. № 4 (15). С. 35–44.
- Мурзин Л. М.** Текст и его восприятие. Свердловск: Урал, 1991. 171 с.
- Норман Б. Ю.** Степень синтаксической связности поэтического текста как переменная величина // Экология языка и коммуникативная практика. 2019. № 4. С. 83–91.

- Скулачева Т. В., Костюк А. Э.** Лингвистические особенности стиха и их функции // Вестник Волгоградского государственного университета. Серия 2: Языкознание. 2020. № 3. С. 155–168.
- Тичер С., Мейер М., Водак Р.** Методы анализа текста и дискурса / пер. с англ. А. А. Киселевой. Х.: Гуманитарный центр, 2017. 356 с.
- Яковлева М. А.** К вопросу о классификации различных видов компенсации // Вопросы филологических наук. 2008. № 4(33). С. 46–51.
- Dunder S. Seljan, Pavlovski M.** Automatic Machine Translation of Poetry and a Low-Resource Language Pair // Proceedings of 43rd International Convention on Information, Communication and Electronic Technology (MIPRO). 2020. P. 1034–1039.
- Ghazvininejad M., Choi Y., Knight K.** Neural Poetry Translation. // Proceedings of NAACL-HLT. 2018. P. 67–71.
- Halliday M. A. K., Hasan R.** Cohesion in English. London: Longman. 1976. 374 p.
- Humblé P.** Machine Translation and Poetry. The Case of English and Portuguese // Ilha do Desterro. 2019. Vol. 72 (2). P.41–56.
- Rifqi M.** Cohesion in poem. A Case Study in 'Marks' and 'the way and the way things are' // LITE: Jurnal Bahasa, Sastra, Dan Budaya. 2009. Vol. 5(1). P. 58–80.

References

- Bochaver S. Yu.** Non-conventional tools of cohesion in contemporary Russian poetry. *Trudy` Instituta russkogo yazy`ka im. V. V. Vinogradova*, 2016, vol. 1 (7), pp. 338–347. (in Russ.)
- Grebenyuk A. V.** Towards building a taxonomy of machine translation errors: versatility vs targeted approaches. *Vestnik Moskovskogo Universiteta. Seriya 22 Teorija Perevoda – Moscow University Bulletin on Translation Studies*, 2024, vol. 2, pp. 7–25 (in Russ.)
- Grebenyuk A. V., Mitrenina O. V.** Analysis of the Distribution of 'Prosaic' Errors in the Machine Translation of Poetic Texts. *Vestnik SPbGU. Yazy`k i literature*, 2025, vol. 3, pp. 648–663. (in Russ.)
- Drejzis Yu. A.** Classicist parallelism as a cohesion device in contemporary Chinese poetry. *Kul`tura i civilizaciy*, 2016, vol. 4, pp. 192–202. (in Russ.)
- Kupina N. A.** Mass literature today. Moscow, Nauka publ., 2009. 423 p. (in Russ.)
- Leontev A. A.** Features of coherence and cohesion of the text. *Discourse Analysis: Proceedings of the conference, Part I*. Moscow, MSU publ., 1974. 230 p. (in Russ.)
- Morya L. A.** Semantic and stylistic system of the poetic text (based on the poem by B. Pasternak 'January 1919'. *Gumanitarnaia paradigm*, 2020, vol. 4 (15). pp. 35–44. (in Russ.)
- Murzin L. M.** The text and its perception. Sverdlovsk, Ural publ., 1991. 171 p. (in Russ.)
- Norman B. Iu.** Degree of syntactic connectivity of a poetic text as a variable value. *Ekologiya iazyka i kommunikativnaia praktika*, 2019, vol. 4, pp. 83–91. (in Russ.)
- Skulacheva T. V., Kostiuk A. E.** Functions and linguistic peculiarities of verse. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2, Iazykoznanie*, 2020, vol. 3, pp. 155–168. (in Russ.)
- Titscher S., Meyer M., Wodak R., Vetter E.** Methods of Text and Discourse Analysis. SAGE Publications Ltd, 2000. 288 p. (in Russ.)
- Yakovleva M. A.** On the issue of classification of different types of compensation in translation. *Voprosy` filologicheskix nauk*, 2008, vol. 4(33). pp. 46–51. (in Russ.)
- Dunder S. Seljan, Pavlovski M.** Automatic Machine Translation of Poetry and a Low-Resource Language Pair. *Proceedings of 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 2020, pp. 1034–1039.

- Ghazvininejad M., Choi Y., Knight K.** Neural Poetry Translation. *Proceedings of NAACL-HLT*, 2018, pp. 67–71.
- Halliday M. A. K., Hasan R.** Cohesion in English. London: Longman, 1976. 374 p.
- Humblé P.** Machine Translation and Poetry. The Case of English and Portuguese. *Ilha do Desterro*, 2019, vol. 72 (2), pp. 41–56.
- Rifqi M.** Cohesion in poem. A Case Study in 'Marks' and 'the way and the way things are'. *LITE: Jurnal Bahasa, Sastra, Dan Budaya*, 2009, vol. 5(1), pp. 58–80.

Информация об авторе

Гребенюк Андрей Владимирович, аспирант Санкт-Петербургского государственного университета
SPIN-код: 4176-7821
Web of Science Researcher ID: JWO-5091-2024
Scopus Author ID: 57194410531

Information about the Author

Andrey V. Grebenyuk, Postgraduate Student at the Department of Mathematical Linguistics, University of St Petersburg
SPIN-code: 4176-7821
Web of Science Researcher ID: JWO-5091-2024
Scopus Author ID: 57194410531

*Статья поступила в редакцию 13.07.2025;
одобрена после рецензирования 19.03.2026; принята к публикации 27.03.2026*
*The article was submitted 13.07.2025;
approved after reviewing 19.03.2026; accepted for publication 27.03.2026*

К достоверным исследованиям: разработка и профильная экспертная верификация корпуса утверждений о здоровье¹

**Полина Алексеевна Колмогорова
Елена Дмитриевна Артёменко
Галина Владимировна Орешина
Елена Юрьевна Кольцова**

Лаборатория социальной и когнитивной информатики Национального исследовательского университета
«Высшая школа экономики»
Санкт-Петербург, Россия

kolmogorovapa@gmail.com, <https://orcid.org/0000-0003-1708-5437>
edartemenko@hse.ru, <https://orcid.org/0000-0002-7610-0089>
oreshinagalina.kosm@gmail.com, <https://orcid.org/0000-0002-5955-6471>
ekoltsova@hse.ru, <https://orcid.org/0000-0002-2669-3154>

Аннотация

Статья посвящена разработке и многоэтапной верификации специализированного корпуса текстовых утверждений о здоровье на русском языке. Актуальность исследования обусловлена высокой распространенностью медицинской дезинформации в Рунете и дефицитом качественно размеченных ресурсов, верифицированных профильными экспертами. Новизна работы заключается в комплексной междисциплинарной методологии создания корпуса, интегрирующей лингвистический анализ для обеспечения однозначности формулировок и строгую многоуровневую врачебную экспертизу с обязательной опорой на авторитетные источники. Методология включала отбор репрезентативных утверждений по пяти актуальным темам, их балансировку по осям «истинность/ложность» и «позитивная/негативная валентность», лингвистическое пилотирование на ясность и однозначность, а также независимую верификацию шестью врачами и тремя экспертами-медиками с последующим мета-ревью профильным специалистом. Особое внимание уделялось анализу стратегий верификации, используемых экспертами, и оценке их согласованности.

Результаты показали, что среди медицинских специалистов наблюдается значительный разброс мнений при оценке утверждений без обращения к источникам. Многоэтапный процесс верификации с привлечением мета-ревьюера оказался критически важным для достижения консенсуса по спорным утверждениям. Анализ выявил умеренную согласованность оценок врачей, основанных только на знаниях (каппа Флейсса = 0,316), и более высокую согласованность экспертов, работавших с источниками (каппа Флейсса = 0,383), что подтверждает необходимость обязательной эталонной проверки.

Ключевым результатом является создание сбалансированного корпуса из 108 верифицированных утверждений (40 истинных, 40 ложных, 28 спорных). Полученный датасет представляет собой ценный ресурс для few-shot обучения NLP-моделей, анализа стратегий верификации экспертами, а также для междисциплинарных исследований в области компьютерной лингвистики и когнитивных наук. Важной особенностью методологии стала разработка четких критериев для категоризации утверждений, включая категорию «спорно» для случаев частичной истинности, преувеличений и недостатка данных. Полученный датасет открывает новые возможности для разработки интеллектуальных систем, способных воспроизводить процесс проверки медицинских утвержде-

© Колмогорова П. А., Артёменко Е. Д., Орешина Г. В., Кольцова Е. Ю., 2026

¹ Корпус доступен по запросу авторов статьи.

ний, включая формирование поисковых запросов, оценку достоверности источников и построение аргументированных заключений.

Ключевые слова

корпус, утверждения о здоровье, фактчекинг, обработка естественного языка (NLP), экспертная разметка

Благодарности

Авторы выражают благодарность экспертам профильной подготовки Юлии Клиник, Наталье Усоловой, Софье Меньшиковой, а также Дарье Моргачевой, осуществлявшей мета-ревью.

Финансирование

Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ, Санкт-Петербург.

Для цитирования

Колмогорова П. А., Артёмов Е. Д., Орешина Г. В., Кольцова Е. Ю. К достоверным исследованиям: разработка и профильная экспертная верификация корпуса утверждений о здоровье // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 83–96. DOI 10.25205/1818-7935-2026-24-2-83-96

Towards Reliable Research: Development and Medical Verification of a Health Claims Corpus

Polina A. Kolmogorova, Elena D. Artemenko

Galina V. Oreshina, Elena Yu. Koltsova

Social and Cognitive Informatics Lab., National Research University Higher School of Economics (HSE University)
Saint-Petersburg, Russian Federation

kolmogorovapa@gmail.com, <https://orcid.org/0000-0003-1708-5437>

edartemenko@hse.ru, <https://orcid.org/0000-0002-7610-0089>

oreshinagalina.kosm@gmail.com, <https://orcid.org/0000-0002-5955-6471>

ekoltsova@hse.ru, <https://orcid.org/0000-0002-2669-3154>

Abstract

The proliferation of digital information has fundamentally changed health-seeking behavior, with individuals increasingly turning to the internet for medical advice. This trend, evidenced by millions of search queries on Runet, is complicated by highly variable reliability of online health information. While Natural Language Processing (NLP) offers promising solutions for automated credibility assessment, its development is hindered by the scarcity of high-quality, expertly annotated datasets. Existing resources often lack in medical verification, being labeled through crowdsourcing or by non-specialists, which severely limits their practical utility.

This paper presents a novel Russian-language dataset of health claims and details its meticulous development methodology. The dataset construction followed three core principles: representativeness (reflecting actual Runet discourse), balance (across truthfulness and valence), and verifiability (against authoritative sources). A significant innovation is the inclusion of a “Disputed” category for claims involving partial truths, exaggerations, or insufficient evidence.

A multi-stage verification pipeline involved linguistic screening by philologists, independent assessment by six physicians (Fleiss’ $\kappa = 0.316$), and source-backed annotation by medical experts ($\kappa = 0.383$). A journalist refined formulations for accessibility, while a meta-reviewer resolved discrepancies based on argument persuasiveness. The final dataset contains 108 claims (40 True, 40 False, 28 Disputed) across five themes: Vaccination, GMOs, Lifestyle, Alternative Medicine, and Vitamins.

Beyond medical NLP applications, this resource will be used/employed in few-shot learning, psycholinguistics, and AI verification strategies. The study demonstrates that reliable medical data annotation requires expert-driven, multi-layered validation with mandatory source verification, emphasizing that linguistic precision is as crucial as factual accuracy. Thus, the developed dataset makes a valuable contribution to both computational linguistics and digital healthcare research, providing a robust benchmark for future studies in Russian-language medical text processing.

Keywords

corpus, health claims, fact-checking, natural language processing (NLP), expert annotation

Acknowledgements

The authors express their gratitude to experts Yuliya Klink, Natalia Usolova, and Sofia Menshikova, as well as to Daria Morgacheva, who conducted the meta-review.

Funding

This study was carried out within the framework of the Fundamental Research Program of HSE University, St. Petersburg.

For citation

Kolmogorova P. A., Artemenko E. D., Oreshina G. V., Koltsova E. Yu. Towards reliable research: development and medical verification of a Health Claims Corpus. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 83–96. DOI 10.25205/1818-7935-2026-24-2-83-96

Введение

Дигитализация информации и повышение ее доступности детерминировали сокращение роли медицинских институтов в процессе принятия человеком решения о собственном здоровье: в поисках ответов на вопросы люди все чаще обращаются к Интернету. Это создает серьезную проблему, поскольку достоверность информации в открытых источниках неоднородна, а доля недостоверных и непроверенных данных о здравоохранении в Интернете остается очень высокой [World Health Organization, 2022].

Важную роль в решении данной проблемы может сыграть использование больших языковых моделей и LLM-агентов в области здравоохранения. Подобные системы способны отвечать на медицинские вопросы [Gilson et al., 2023], а в отдельных случаях они сопоставимы с врачами по уровню эмпатии [Ayers et al., 2023]. Вместе с тем их выводы не могут рассматриваться как полноценная альтернатива профессиональному медицинскому заключению.

Данная проблема стимулирует активное развитие исследований на стыке искусственного интеллекта и здравоохранения. Методы автоматической обработки естественного языка (Natural Language Processing, NLP) применяются для автоматического анализа медицинской информации [Afsana et al., 2021], для борьбы с медицинской дезинформацией [Deka et al., 2022]; создаются специализированные датасеты [Srba et al., 2022]. Однако ключевым ограничением остается низкое качество данных: они часто получены из непроверенных источников или размечены непрофильными специалистами [Kesiku et al., 2022; Basyal et al., 2020], что снижает надежность, точность и практическую применимость моделей. Особую проблему в области открытой медицинской информации представляют околomedical мифы и утверждения, а также случаи намеренного упрощения или искажения результатов научных исследований [Chou et al., 2021]. Специализированных датасетов для работы с подобными категориями медицинской информации в настоящее время не существует.

Целью данного исследования является создание датасета текстовых утверждений о здоровье на русском языке и разработка методики его многоэтапной врачебной верификации. В работе решаются задачи отбора репрезентативных утверждений, разработки критериев оценки, проведения независимой экспертной валидации с участием медицинских специалистов и обеспечения контроля качества на каждом этапе.

Актуальность датасета для лингвистических исследований

В условиях дефицита качественно размеченных данных в области медицины и здравоохранения обсуждаемый датасет может использоваться для разработки и тестирования методов few-shot обучения – парадигмы машинного обучения, где модель классифицирует объекты, увидев всего несколько примеров каждого класса [Wang et al., 2020]. Наряду с самими утверждениями, корпус содержит список источников информации, использованных врачами при проверке, и их экспертные комментарии. Это позволяет исследовать методы работы с ин-

формацией: как эксперты ищут, анализируют и интерпретируют данные для оценки достоверности.

Датасет важен для исследований в области когнитивной лингвистики и психолингвистики, особенно в контексте изучения восприятия медицинской информации на русском языке. В отличие от большинства существующих ресурсов, сфокусированных на английском языке, данный корпус предоставляет уникальные возможности для изучения специфики работы с медицинским дискурсом в условиях менее ресурсного языка. Сбалансированный по валентности и длине корпус утверждений может использоваться в качестве стимульного материала в психолингвистических экспериментах, например, для изучения того, как лингвистические особенности сообщения (например, негативная валентность) влияют на скорость и точность распознавания его достоверности реципиентами.

Таким образом, высокое качество верифицированности представленных в датасете данных и их сбалансированность позволяют использовать его в качестве эталонного корпуса для решения различных научных и прикладных задач на стыке лингвистики, компьютерных и когнитивных наук.

1. Разработка датасета

В основу датасета положены три взаимосвязанных принципа:

1. Репрезентативность: утверждения должны отражать реальные и актуальные темы медицинского и околomedicalского дискурса в русскоязычном информационном пространстве.
2. Сбалансированность: корпус должен быть сбалансирован по двум ключевым осям: «истинные – спорные – ложные» утверждения и утверждения с «позитивной – негативной» валентностью в рамках каждой темы.
3. Верифицируемость: каждое утверждение должно иметь однозначное толкование и предполагать возможность его проверки с опорой на четкий круг авторитетных источников.

Для реализации этих принципов был применен многоэтапный процесс, включавший сбор первичных данных, их тематическую и лингвистическую фильтрацию, а также многоуровневую экспертизу. Последовательность и взаимосвязь этих этапов представлены на рисунке.

Достоверность утверждений

При формировании корпуса была реализована выборка утверждений, репрезентирующих различные категории эпистемической оценки.

Изначально применялась бинарная разметка («истинно»/«ложно»), однако на этапе экспертной верификации врачами была введена третья категория – «спорно», что позволило учесть градиенты научной определенности. Итоговая категоризация соответствует современному научному консенсусу и опирается на следующие критерии: (1) правда (утверждение подтверждено устойчивым консенсусом и достаточной эмпирической базой); (2) ложь (однозначно опровергнуто научным сообществом); (3) спорно (присваивается при частичной истинности, преувеличении, недостатке или противоречивости данных, невозможности верификации или неясной формулировке).

Окончательная категоризация всех утверждений была приведена в соответствие с действующим научным консенсусом, установленным на основе анализа авторитетных и официальных источников. В качестве официальных источников использовались веб-платформы институций, чья миссия связана с охраной общественного здоровья (например, проект Роспотребнадзора «Здоровая диета», официальный сайт Всемирной организации здравоохранения – WHO.int и др.). Авторитетными источниками считались публикации в рецензируемых научных журна-

Подготовка корпуса



Пайплайн исследования
Pipeline research

лах, соответствующих международным стандартам качества и индексируемых в базах Scopus и Web of Science.

Тематика утверждений

Утверждения, входящие в корпус, описывают 5 основных тем, выявленных на основании анализа частоты ключевых слов в поисковых запросах за 30, 90 и 12 месяцев при помощи аналитического инструмента Google Trends²: вакцинация, употребление генно-модифицированных продуктов в пищу, здоровый образ жизни, применение методов альтернативной медицины в собственной практике лечения, прием витаминов.

² Google Trends – инструмент для анализа популярности поисковых запросов в Google. URL: <https://trends.google.com/trends/> (дата обращения: 07.09.2025)

Валентность утверждений

Для всех утверждений из датасета была определена валентность, которая в настоящем исследовании рассматривалась не как эмоциональная окраска, а как семантическая направленность утверждения, связанная с наличием прямой или косвенной негации. Для операционализации валентности использовалась бинарная классификация по осям «опасно – безопасно» и «полезно – бесполезно»: утверждения, относящиеся к категории «безопасно» и/или «полезно», кодировались как обладающие позитивной валентностью, тогда как относящиеся к категориям «опасно» и/или «бесполезно» – как обладающие негативной валентностью. Данный подход опирается на предположение, что утверждения с отрицательной конструкцией сложнее для когнитивной обработки и могут влиять на точность их верификации респондентами.

Однозначность и ясность

В предварительный корпус утверждений вошли 200 утверждений, каждое из которых дополнительно было верифицировано командой студентов-практикантов, получающих образование в области социальных наук ($N = 5$) под руководством экспертов с опытом работы в филологии, лингвистике и/или СМИ и журналистике. Никто из экспертов не обладал медицинским образованием. Основной задачей экспертов была оценка толкования полученных утверждений по бинарным шкалам ясности и однозначности.

Под ясностью понималось отсутствие в формулировке специализированной медицинской терминологии или описаний, потенциально непонятных аудитории без профильной подготовки.

Под однозначностью понималось наличие только одной смысловой пропозиции в каждом утверждении. Данный критерий вводился для того, чтобы при последующей экспертной оценке оценивалась достоверность одного изолированного высказывания, а не нескольких одновременно. В случаях, когда утверждение содержало более одной пропозиции, оно подвергалось декомпозиции на отдельные формулировки. После верификации на уровне ясности и однозначного толкования в корпус вошли 120 утверждений.

2. Процесс врачебной верификации корпуса

Процедура верификации была организована как многоуровневый процесс. На первом этапе врачи оценивали утверждения без обращения к внешним источникам, что позволило выявить уровень профессионального консенсуса и обозначить спорные формулировки, требующие дальнейшего анализа. Именно тогда возникла необходимость во введении третьей категории достоверности и были сформулированы критерии спорности. Последующие этапы – с участием экспертов-медиков, медицинского журналиста и мета-ревьюера с медицинским образованием и опытом в медицинской журналистике – были направлены на разрешение выявленных разночтений и создание объективно верифицированного датасета (табл. 1).

Таким образом, представленная структура верификации обеспечивала последовательное движение от субъективной профессиональной оценки к объективной, основанной на консенсусе специалистов в области здоровья и подтвержденной авторитетными источниками разметке данных. Первым шагом в этой многоуровневой системе стала верификация врачами.

Таблица 1

Функции и роли привлекаемых для разметки экспертов

Table 1

Functions and roles of experts involved in annotation

Роль	Функция
Врач	Оценка ложности/истинности утверждения с опорой только на собственные знания (без обращения к внешним источникам и справочной информации)
Медик-эксперт	Оценка истинности утверждения с опорой на авторитетные источники, при необходимости – переформулировка для большей однозначности, сопровождаемая записью экрана и устным комментарием (зафиксированным на диктофон)
Журналист	После чтения комментариев экспертов предложение формулировки, которая сделает утверждение более однозначным с точки зрения правдивости его содержания
Мета-ревьюер (врач и медицинский журналист)	Присвоение утверждению окончательных меток истинности, при необходимости переформулирование утверждения, чтобы оно было однозначным. Оценка убедительности коллег, а не самостоятельная проверка информации

Верификация врачами

Корпус из 120 утверждений оценивался шестью врачами различного пола, возраста, медицинской специализации и региона проживания, каждый из которых оценивал утверждения без обращения к источникам и справочной информации. Экспертам необходимо было внимательно ознакомиться с утверждениями и пометить в соответствии с предложенными метками достоверности.

Сходимость полученной экспертной оценки оказалась умеренной согласно коэффициенту конкордации Кендалла ($W = 0,43$) [Kendall, Smith, 1939] и низкой согласно внутриклассовому коэффициенту корреляции (ICC = 0,27) [Fisher, 1925] и коэффициенту α Криппендорфа ($\alpha = 0,314$) [Krippendorff, 2013], что обнаружило дефицит согласованности в оценке медицинских знаний участников-медиков в сфере общего здоровья. Этот результат указывает на то, что вне рамок собственной профессиональной специализации, а особенно в отношении вопросов, относящихся к околomedicalным практикам (например, использование контрастного го душа), эксперты не демонстрируют устойчивого единства в суждениях.

Верификация экспертами-медиками

На следующем этапе разметку корпуса выполнили три независимых эксперта-медика. Каждый оценивал правдивость утверждений на основе актуальных медицинских знаний, подкрепляя решения ссылками на авторитетные источники. Перед началом работы эксперты прошли краткий устный инструктаж, после чего заполняли таблицу автономно – без взаимодействия друг с другом. Во время разметки они могли записывать экран и давать устные комментарии (эксперт № 1 – видео с комментариями, № 2 – только видео, № 3 – только аудиокomментарии). Все записи были расшифрованы, а цитируемые источники добавлены в корпус. Эксперты также имели возможность уточнять формулировки утверждений, чтобы сделать их однозначно правдивыми или ложными.

Анализ стратегий верификации утверждений экспертами на основе анализа источников

Для детального анализа стратегий верификации, используемых экспертами, был проведен контент-анализ источников: первичных (указанных экспертами при разметке – «Источник 1») и дополнительных (выявленных в скринкастах или устных комментариях – «Источник 2»). Анализ охватывал данные всех трех экспертов.

Каждый источник был отнесен к одной из семи категорий: (1) международные научные и медицинские базы данных (например, PubMed, PMC, ScienceDirect); (2) авторитетные медицинские порталы (например, UpToDate, Healthline, WebMD); (3) официальные международные организации (например, WHO, CDC, FDA); (4) официальные органы Российской Федерации (например, Роспотребнадзор, Минздрав РФ, клинические рекомендации); (5) русскоязычные медицинские ресурсы (например, CyberLeninka, Vidal, РМЖ); (6) энциклопедические ресурсы (википедия); (7) новостные ресурсы и СМИ.

Количество указанных источников

Анализ заполнения выявил различную степень полноты и последовательности в указании экспертами источников (табл. 2).

Таблица 2

Статистика заполнения колонок с источниками экспертами

Table 2

Statistics of source column completion by experts

Эксперт	Всего утверждений	Указан источник 1	Указан источник 2	Указаны оба источника	Не указано ни одного источника
№ 1	120	107	89	84	0
№ 2		92	75	69	6
№ 3		95	74	68	9

Как видно из табл. 2, эксперт № 1 продемонстрировал наибольшую последовательность, указав как минимум один источник для всех 120 утверждений и максимальное количество источников в колонке дополнительных источников. Эксперты № 2 и № 3 пропустили указание источников для 6 и 9 утверждений соответственно.

Категориальный анализ источников

Для выявления стратегий верификации был проведен категориальный анализ всех указанных источников с разделением по колонкам (табл. 3).

Стратегия эксперта № 1 демонстрирует четко выраженную ориентацию на англоязычные доказательные ресурсы: на международные научные базы данных и авторитетные медицинские порталы приходится 68,2 % источников в колонке 1 и 67,4 % – в колонке 2. При этом «Источник 2» не используется для смены типа источника, а выполняет функцию уточнения и углубления первичного поиска, что отражается в увеличении доли обращений к международным научным базам во второй колонке. Почти полное отсутствие ссылок на российские официальные ресурсы (в сумме – 4,6 %) указывает на устойчивую исследовательскую стратегию,

ориентированную на глобальные стандарты доказательной медицины и англоязычное научное пространство.

Таблица 3

Распределение источников, указанных экспертами,
по категориям и колонкам

Table 3

Distribution of sources cited by experts across categories and columns

Категория	Эксперт	Источник 1 (кол-во / %)	Источник 2 (кол-во / %)
1. Международные научные / мед. базы	№ 1	41 / 38,3	40 / 44,9
	№ 2	11 / 12,0	12 / 16,0
	№ 3	17 / 17,9	15 / 20,3
2. Авторитетные мед. порталы	№ 1	32 / 29,9	20 / 22,5
	№ 2	3 / 3,3	5 / 6,7
	№ 3	11 / 11,6	11 / 14,9
3. Официальные органы (global)	№ 1	15 / 14,0	12 / 13,5
	№ 2	4 / 4,3	4 / 5,3
	№ 3	26 / 27,4	19 / 25,7
4. Официальные органы (РФ)	№ 1	4 / 3,7	5 / 5,6
	№ 2	31 / 33,7	21 / 28,0
	№ 3	14 / 14,7	11 / 14,9
5. Русскоязычные мед. ресурсы	№ 1	10 / 9,3	9 / 10,1
	№ 2	37 / 40,2	23 / 30,7
	№ 3	20 / 21,1	13 / 17,6
6. Википедия	№ 1	3 / 2,8	2 / 2,2
	№ 2	3 / 3,3	5 / 6,7
	№ 3	5 / 5,3	4 / 5,4
7. Новости / СМИ	№ 1	2 / 1,9	1 / 1,1
	№ 2	3 / 3,3	5 / 6,7
	№ 3	2 / 2,1	1 / 1,4

Стратегия эксперта № 2 ориентирована на верификацию в рамках национального нормативного и информационного поля. Российские официальные источники и профильные медресурсы составляют 73,9 % при первичном поиске и 58,7 % во второй колонке. «Источник 2» чаще используется для поиска альтернативных отечественных подтверждений, реже – для обращения к международным данным. Низкая доля международных научных баз (12,0 и 16,0 %) отражает приоритет национальной доказательной базы и ограниченное обращение к зарубежным источникам.

Стратегия эксперта № 3 является гибридной, с заметным уклоном в сторону официальных институциональных источников. В отличие от эксперта № 1 он чаще обращается к ресурсам глобальных организаций (ВОЗ, CDC – 27,4 и 25,7 %) при меньшей доле первичных научных баз (17,9 и 20,3 %). Российские источники выполняют вспомогательную функцию

(35,8 и 32,5 %). Равномерное распределение между колонками указывает, что «Источник 2» используется для дополнительного подтверждения, а не смены стратегии поиска. Повышенное число пропусков (9) свидетельствует о менее систематичном характере работы.

Таким образом, анализ источников позволяет не только оценить степень доказательности выводов экспертов, но и выявить фундаментальные различия в их методологических подходах: от строго научного (эксперт № 1) и нормативно-локального (эксперт № 2) до гибридного с опорой на институциональный авторитет (эксперт № 3). Важно отметить, что в совокупности разметка получилась сбалансированной по оси «российские – международные источники», что обеспечивает репрезентативность методологической базы.

Дополнительная редакторская правка

На промежуточном этапе в процедуру был включен эксперт с журналистским образованием, выступавший в роли медиатора между врачами и финальной редакцией. Он формулировал альтернативные версии утверждений, когда: (а) врачи выявляли ошибку, но не предлагали замены; (б) их формулировки были избыточно техническими или сложными для широкой аудитории. Это позволило сохранить научную точность и обеспечить коммуникативную доступность без искажения смысла.

Верификация мета-ревьюером

Мета-ревьюер (эксперт с медицинским образованием и опытом в медицинской журналистике) осуществлял финальную консолидацию разметки, выполненной врачами. Его задача заключалась не в самостоятельной проверке утверждений, а в оценке убедительности аргументации экспертов и авторитетности привлеченных источников на основе предоставленных материалов.

Если мнения врачей совпадали и были хорошо обоснованы, мета-ревьюер присваивал метку большинства. При полном расхождении мнений утверждение исключалось из корпуса. В промежуточных случаях мета-ревьюер опирался на качество аргументации и надежность источников, избегая собственного поиска информации и не полагаясь исключительно на оценки врачей (поскольку те не верифицировали утверждения напрямую). При наличии четкого пути к консенсусу он мог переформулировать утверждение и присвоить итоговую метку.

В финальную версию корпуса вошли 108 утверждений: 40 – «ложь», 40 – «правда» и 28 – «спорно».

3. Результаты

Предобработка данных

На этапе предобработки данных было выявлено минимальное количество меток «99» (предложение об удалении утверждения из базы): эксперты предложили удалить 4 утверждения (3,3 % от общего числа), врачи – только 2 утверждения (1,7 %). Данный показатель свидетельствует о высоком качестве первоначальной подготовки корпуса утверждений, их релевантности и понятности для экспертов. Такой низкий уровень отсева указывает на эффективность проведенных этапов пилотирования и лингвистической экспертизы.

Метрики согласованности

Анализ межэценочной согласованности выявил существенные различия между группами экспертов (работавших с источниками) и врачей (оценивавших без внешней верификации).

Абсолютное согласие – доля случаев, в которых все оценщики в группе выставили идентичную метку – составило 43,1 % у экспертов и лишь 23,7 % у врачей. Каппа Флейсса, корректирующая согласие на случайное совпадение, составила 0,383 и 0,316 соответственно.

Значения в диапазоне 0,30–0,40 указывают на умеренное согласие: оценщики сходятся в базовой категоризации, но расходятся при интерпретации пограничных случаев. Более высокая каппа у экспертов свидетельствует, что систематическая работа с авторитетными источниками и обоснование решений не только повышают точность, но и способствуют формированию более согласованных критериев классификации сложных утверждений. Разница в 0,067 эквивалентна достижению консенсуса в дополнительных ~7 % спорных случаев за счет структурированной верификационной процедуры.

Попарный анализ с использованием каппы Коэна [Cohen, 1960] выявил неравномерность распределения согласия между разными парами оценщиков (№ 1 vs № 2 – 0,548; № 1 vs № 3 – 0,499; № 2 vs № 3 – 0,594). Такая вариабельность указывает на формирование «микрогрупп» с общими подходами к оценке, что характерно для сложных экспертных задач.

Метрики консервативности оценок

Анализ использования ответа «не знаю», измеряемого как процент ответов «0» от общего количества данных оценщиком ответов, показал, что врачи использовали эту метку значительно чаще (7,5 % в среднем), чем эксперты (2,9 % в среднем). Двукратная разница между группами интерпретируется как следствие различных методологических подходов: врачи, не имевшие доступа к источникам, чаще выбирали стратегию осторожности, в то время как эксперты были вынуждены искать ответ в источниках.

Анализ сложных случаев

Среднее разногласие, измеряемое как стандартное отклонение оценок для каждого утверждения по шкале от 0 до 3, составило у экспертов 0,444, в то время как у врачей этот показатель был значительно выше – 0,664. Данная метрика, где высокое стандартное отклонение означает большой разброс мнений по конкретному утверждению, свидетельствует о большей субъективности оценок врачей при отсутствии доступа к источникам.

Были идентифицированы топ-5 самых спорных случаев для экспертов: утверждение «Чрезмерное употребление алоэ может вызвать проблемы с пищеварительной системой» (разногласие: 1,732; переформулировано с присвоением класса «ложь»), утверждение «Генно-модифицированные растения быстрее истощают почву» (1,528) (присвоен класс «спорно»), утверждение «Ежегодно в России выявляются незарегистрированные линии ГМО» (1,528) (удалено из датасета), утверждение «Сахарозаменитель можно использовать как лекарство для похудения» (1,155) (присвоен класс «спорно»), утверждение «Сахарозаменители вызывают диабет» (1,155) (присвоен класс «спорно»).

Корреляция уверенности и сложности

Была выявлена умеренная положительная корреляция (коэффициент Спирмена ~0,47) [Spearman, 1904] между сложностью утверждений, измеренной через стандартное отклонение оценок, и использованием ответа «не знаю» как для экспертов (0,476), так и для врачей (0,474). Данная метрика, рассчитываемая как корреляция между ранжированными списками утверждений по сложности и по частоте использования «не знаю», является доказательством внутренней валидности исследования. Она свидетельствует, что оценщики не ставили оценки

случайно, а последовательно использовали метку «не знаю» для более сложных утверждений, и подтверждает объективный характер сложности как параметра.

Заключение

Проведенное исследование продемонстрировало высокую методологическую сложность и трудоемкость создания достоверных датасетов в медицинской сфере. Разработанный корпус медицинских утверждений стал первым верифицированным ресурсом подобного типа для русскоязычного пространства, обеспечивающим сбалансированное представление тем, уровней достоверности и оценочной валентности.

Ключевым методологическим результатом работы стало подтверждение принципиальной недостаточности краудсорсинговой или непрофильной разметки для медицинских данных. Даже при участии специалистов-врачей наблюдаются расхождения в интерпретации утверждений, что подчеркивает необходимость обязательной эталонной проверки по внешним авторитетным источникам. Введение многоступенчатой процедуры верификации с участием мета-ревьюера позволило не только согласовать противоречивые оценки, но и выработать объективный консенсус по спорным утверждениям, обеспечив высокую внутреннюю надежность итогового корпуса.

Полученные результаты выявили, что источником недостоверности нередко служит не только содержательная, но и лингвистическая структура сообщений – двусмысленные формулировки, преувеличения и эвристические обобщения, затрудняющие когнитивную оценку достоверности информации. Этот вывод подчеркивает необходимость комплексного анализа дезинформации на пересечении медицинской, лингвистической и когнитивной перспектив.

Несмотря на тематическую фокусировку и ограниченный объем (108 утверждений), созданный датасет представляет собой надежный и методологически обоснованный ресурс, предназначенный прежде всего для задач обучения по нескольким примерам (few-shot learning), тестирования и валидации NLP-моделей. Он может служить эталонной базой для последующего масштабирования – как в рамках расширения тематического охвата, так и при разработке аналогичных верифицированных корпусов для других языков и культурных контекстов.

Таким образом, проведенная работа не только предложила комплексную и воспроизводимую методику построения надежных датасетов для чувствительных областей, но и обозначила новые направления междисциплинарных исследований: от совершенствования процедур экспертной верификации до изучения когнитивных механизмов восприятия медицинской информации в цифровой среде.

Список литературы / References

- Ayers J. W., Poliak A., Dredze M., Leas E. C., Zhu Z., Kelley J. B., Faix D. J., Goodman A. M., Longhurst C. A., Hogarth M., Smith D. M. Comparing physician and artificial intelligence Chatbot responses to patient questions posted to a public social media forum // *JAMA Internal Medicine*, 2023. Vol. 183 (6), 589. <https://doi.org/10.1001/jamainternmed.2023.1838>
- Afsana F., Ashad K. M., Naeem H., Manoranjan P. Automatically Assessing Quality of Online Health Articles // *IEEE Journal of Biomedical and Health Informatics*, 2021. Vol. 25. No. 2. P. 591–601. <https://doi.org/10.1109/JBHI.2020.3032479>
- Basyal G. P., Bhandari G., Adhikari S., Maharjan S., Shakya S. A Systematic Review of Natural Language Processing for Knowledge Management in Healthcare // *CSIT*. 2020. P. 21–27. <https://doi.org/10.5121/CSIT.2020.100921>

- Chou W.-Y. S., Gaysynsky A., Vanderpool R. C.** The COVID-19 Misinfodemic: Moving beyond fact-checking // *Health Education & Behavior*. 2021. Vol. 48. No. 1. P. 9–13. <https://doi.org/10.1177/1090198120980675>
- Cohen J.** A coefficient of agreement for nominal scales // *Educational and Psychological Measurement*. 1960. Vol. 20. No. 1. P. 37–46. <https://doi.org/10.1177/001316446002000104>
- Deka P., Jurek-Loughrey A., De Campos C. P.** Evidence Extraction to Validate Medical Claims in Fake News Detection // *Computational Data and Social Networks: 11th International Conference, CSoNet 2022, Proceedings*. 2022. P. 3–15. https://doi.org/10.1007/978-3-031-20627-6_1
- Fleiss J. L.** Measuring nominal scale agreement among many raters // *Psychological Bulletin*. 1971. Vol. 76. No. 5. P. 378–382. <https://doi.org/10.1037/h0031619>
- Infodemics and misinformation negatively affect people's health behaviours – new WHO review finds // *World Health Organization: WHO*. 2022. URL: <https://www.who.int/europe/ru/news/item/01-09-2022-infodemics-and-misinformation-negatively-affect-people-s-health-behaviours-new-who-review-finds> (дата обращения: 03.09.2025)
- Gilson A., Safranek C. W., Huang T., Socrates V., Chi L., Taylor R. A., Chartash D.** How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language Models for Medical Education and knowledge assessment // *JMIR Medical Education*. 2023. Vol. 9, e45312. <https://doi.org/10.2196/45312>
- Kendall M. G., Smith B. B.** The Problem of m Rankings // *The Annals of Mathematical Statistics*. 1939. Vol. 10. No. 3. P. 275–287. <https://doi.org/10.1214/aoms/1177732186>
- Kesiku C., Oyeboode O., Famiyesin O., Ogunleye G.** Natural Language Processing Techniques for Text Classification of Biomedical Documents: A Systematic Review // *Information*. 2022. Vol. 13. No. 10. P. 499. <https://doi.org/10.3390/info13100499>
- Krippendorff K.** Content analysis: An introduction to its methodology. 3rd edition. Thousand Oaks, CA: Sage, 2013. <https://doi.org/10.4135/9781071878781>
- Mann H. B., Whitney D. R.** On a test of whether one of two random variables is stochastically larger than the other // *Annals of Mathematical Statistics*. 1947. Vol. 18. No. 1. P. 50–60. <https://doi.org/10.1214/aoms/1177730491>
- Shrout P. E., Fleiss J. L.** Intraclass Correlations: Uses in Assessing Rater Reliability // *Psychological Bulletin*. 1979. Vol. 86. No. 2. P. 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Spearman C.** The Proof and Measurement of Association between Two Things // *The American Journal of Psychology*. 1904. Vol. 15. No. 1. P. 72–101. <https://doi.org/10.2307/1412159>
- Srba I., Moro R., Tomlein M., Pecher B., Simko J., Stefancova E., Kompan M., Hreckova A., Podrouzek J., Bielikova M.** Monant Medical Misinformation Dataset: Mapping Articles to Fact-Checked Claims // *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2022. P. 3269–3275. <https://doi.org/10.1145/3477495.3531726>
- Wang Y., Yao Q., Kwok J. T., Ni L. M.** Generalizing from a Few Examples: A Survey on Few-Shot Learning // *ACM Computing Surveys (CSUR)*. 2020. Vol. 53, Issue 3. P. 1–34. <https://doi.org/10.1145/3386252>

Информация об авторах

Колмогорова Полина Алексеевна, стажер-исследователь Лаборатории социальной и когнитивной информатики НИУ ВШЭ, аспирант Школы по филологическим наукам НИУ ВШЭ, Санкт-Петербург

Артеменко Елена Дмитриевна, кандидат филологических наук, заместитель заведующей Лабораторией социальной и когнитивной информатики НИУ ВШЭ, Санкт-Петербург

Орешина Галина Владимировна, стажер-исследователь Лаборатории социальной и когнитивной информатики НИУ ВШЭ, преподаватель департамента социологии НИУ ВШЭ, Санкт-Петербург

Кольцова Елена Юрьевна, доктор филологических наук, заведующая Лабораторией социальной и когнитивной информатики НИУ ВШЭ, профессор департамента социологии НИУ ВШЭ, Санкт-Петербург

Information about the Authors

Polina A. Kolmogorova, Research Assistant, Laboratory for Social and Cognitive Informatics, HSE University; PhD Student, School of Philological Studies, HSE University, St. Petersburg

Elena D. Artemenko, Candidate of Science in Philology, Deputy Head of the Laboratory for Social and Cognitive Informatics, HSE University, St. Petersburg

Galina V. Oreshina, Research Assistant, Laboratory for Social and Cognitive Informatics, HSE University; Lecturer, Department of Sociology, HSE University, St. Petersburg

Elena Yu. Koltsova, Doctor of Science in Philology (Dr. habil.), Head of the Laboratory for Social and Cognitive Informatics, HSE University; Professor, Department of Sociology, HSE University, St. Petersburg

*Статья поступила в редакцию 27.10.2025;
одобрена после рецензирования 14.12.2025; принята к публикации 19.12.2025*

*The article was submitted 27.10.2025;
approved after reviewing 14.12.2025; accepted for publication 19.12.2025*

Научная статья

УДК'81

DOI 10.25205/1818-7935-2026-24-2-97-111

Методы компьютерной лингвистики для идентификации маркеров языка травмы

**Татьяна Ивановна Попова
Виолетта Владимировна Наумова**

Национальный исследовательский университет «Высшая школа экономики»
Санкт-Петербург, Россия

tipopova13@gmail.com, <https://orcid.org/0000-0003-2066-7868>
vvnauмова_2@edu.hse.ru, <https://orcid.org/0009-0001-8903-3065>

Аннотация

В работе исследуется язык травмы – вербальные и невербальные стратегии, использованные при описании пережитых трагических событий. На материале интервью бывших заложников теракта в театральном центре на Дубровке (2002) анализируются лексические и хезитационные закономерности, которые могут служить индикаторами рассказа о пережитом трагическом событии, который предположительно отличается от повседневного нарратива. Были использованы такие методы, как составление частотных словарей, расчет лексической дисперсии отдельных словоформ и подсчет знаков хезитации. Многоуровневый эмпирический анализ пользовательского подкорпуса позволил получить результаты лингвистических и паралингвистических маркеров травмы. На лексико-грамматическом уровне было выявлено обилие местоимения «я», что связано с особенностями нарратива, и пояснительных конструкций («потому что», «то есть») над другими функциональными словами, что может отражать высокую потребность в рационализации пережитого опыта. Была обнаружена вариативность в номинации террористов, которая заметно коррелирует с временной дистанцией между событием и моментом интервью. Например, лексема «чеченцы» встречается в ранних интервью (2003–2004 гг.) и в интервью российских эмигрантов в США, тогда как вариант «террористы» преобладает в поздних интервью (2019–2023 гг.). Также встречались опосредованные обозначения агрессоров («ребята с автоматами»), что может свидетельствовать о стратегии дистанцирования от травмы. При анализе паралингвистических особенностей было установлено, что максимальная плотность хезитационных маркеров наблюдается при описании кульминационных событий (три дня в заложниках). Это может свидетельствовать о когнитивной и эмоциональной нагрузке информантов. Так вдохи и выдохи не только регулируют речь, но и маркируют стресс в середине нарратива. Вздохи же являются наименее частотной хезитационной особенностью, но отражают наивысшую степень аффективности, так как они сопровождают моменты плача и описания смерти. Комбинация методов доказала свою эффективность с точки зрения изучения языка травмы.

Ключевые слова

trauma studies, язык травмы, корпусная лингвистика, лексическая дисперсия, прагматические маркеры

Финансирование

Статья подготовлена в ходе проведения исследования в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ) в 2025 году.

Для цитирования

Попова Т. И., Наумова В. В. Методы компьютерной лингвистики для идентификации маркеров языка травмы // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 97–111. DOI 10.25205/1818-7935-2026-24-2-97-111

© Попова Т. И., Наумова В. В., 2026

Computational Linguistics Methods for Identifying Language of Trauma Markers

Tatyana I. Popova, Violetta V. Naumova

National Research University Higher School of Economics (HSE University)
Saint-Petersburg, Russian Federation

tipopova13@gmail.com, <https://orcid.org/0000-0003-2066-7868>
vvnaumova_2@edu.hse.ru, <https://orcid.org/0009-0001-8903-3065>

Abstract

The study explores the language of trauma by analyzing verbal and non-verbal strategies employed in recounting traumatic experiences. Using interview data from former hostages of the 2002 Dubrovka theatre siege, the research explores lexical and hesitation patterns that may serve as indicators of narratives which are distinct from everyday discourse. Methods such as frequency dictionary compilation, lexical dispersion calculation of individual word forms, and hesitation marker counts were utilized. A multi-level analysis of the subcorpus yielded empirical findings on linguistic and paralinguistic markers of trauma. At the lexico-grammatical level, the prominence of the pronoun “I” and explanatory constructions e.g., “because”, “that is”, over other function words suggests a heightened need to rationalize the events experienced. Variability in the naming of terrorists correlates with temporal distance and the speakers’ cultural context; for instance, the term “Chechens” appears in early interviews (2003–2004) and those by Russian emigrants in the USA, whereas “terrorists” dominates in later interviews (2019–2023). Indirect references to perpetrators e.g., “guys with guns”, also occur, potentially indicating a distancing strategy from the trauma. Paralinguistic analysis found that the highest density of hesitation markers occurs in descriptions of the event climax (three days in captivity), reflecting cognitive and emotional strain. Both inhalations and exhalations do not regulate speech but mark stress mid-narrative, while sighs, though less frequent, correspond to moments of crying and death descriptions, indicating peak affective intensity. The combined methodological approach proved to be effective for investigating the language of trauma.

Keywords

trauma studies, trauma language, corpus linguistics, lexical dispersion, pragmatic markers

Funding

The article/review was prepared within the framework of the Basic Research Program at HSE University.

For citation

Popova T. I., Naumova V. V. Computational linguistics methods for identifying language of trauma markers. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 97–111. DOI 10.25205/1818-7935-2026-24-2-97-111

Введение

Коллективная травма – это переживание травматического события в сообществе. Феномен интересен не только для социологов и психологов [Александр, 2012; Erikson, 2012], но и лингвистов: травматическое событие, обладая устойчивыми эмпирическими характеристиками, оказывает влияние как на культуру, так и на язык. Широкий академический интерес позволил активно развиваться новому направлению исследований – trauma studies. Социологи представляют культурную травму как общественный конструкт, который влияет не только на людей, переживших шокирующее событие, но и на последующие поколения [Александр, 2012]. В лингвистических работах исследуется понятие языка травмы, которое заключается в том, что язык может дать представление о психологическом состоянии человека. Ю. Митчелл пишет: «Психологическая травма нарушает условия безопасности и создает длительный стресс, нарушая когнитивные и, впоследствии, языковые способности человека» [Mitchell, 1998. Р. 121]. В настоящей работе под языком травмы понимается процесс вербализации травматического события, который наполнен лингвистическими и паралингвистическими коррелятами. Для выявления лингвистических маркеров и хезитационного контура травматического нарратива были использованы методы количественного анализа текста, а также подсчет знаков хезитации.

Существуют исследования, посвященные изучению связи между травмой и памятью. Согласно Ван дер Колку и Ван дер Халту, пациент не способен к адекватной наррации, вспоминая о травмирующей ситуации, так как чаще всего это связано как с нереализованными представлениями о себе в контексте травмирующего события, так и с функцией психики забывать подобные моменты [Van der Kolk, Van der Halt, 1991. Р. 427]. Таким образом, травматический нарратив, как правило, связан не логической последовательностью событий, а вспышками фрагментов из сенсорной памяти: звуками, запахами или образами, – эти изолированные отпечатки и создают рассказ о пережитом событии [Ван дер Колк, 2014. С. 67].

В то же время Б. Буш и Т. МакНамара считают, что травматический нарратив сопровождается психологическими феноменами, которые варьируются в зависимости от реакции пациента. Травматический дискурс рассматривается как континуум от структурно полных до фрагментированных нарративов с паузами, гезитациями и отвлечениями [Busch, McNamara, 2020. Р. 330].

Авторы, анализируя лингвистические механизмы травмы на примере рассказов о Второй мировой войне, отмечают, что травматические нарративы выходят за пределы обычного языка, проявляясь через афазии и паузы, равнозначные словам по смыслу. Исследователи предлагают, что необходимо изучать не только лексику, но и паралингвистические корреляты – «place holders for the unsayable» («хранители несказанного», пер. мой – В. Н.), которые могут проявляться в уклонении от темы или метафорах. Кроме того, важны грамматические изменения: нарратор может не ставить себя субъектом действия, объективизировать себя, что отражает психологическую динамику жертвы насилия [Busch, McNamara, 2020. Р. 331].

Для выявления специфики языка травмы ключевое значение имеет исследование повседневного дискурса, который может служить базой для сравнительного анализа. Важны работы по мультиканальному дискурсу, где детально описаны характеристики спонтанной коммуникации. Так, исследования на базе корпуса «Рассказы и разговоры о грушах» демонстрируют ряд устойчивых параметров: высокая плотность гезитации (7–10 гезитаций на 1000 слов), частотные редупликации (по типу «ну-ну», «так-так»), а также эллиптические конструкции (до 20 % высказываний) [Кибрик, 2018]. Устный дискурс представляет собой иерархическую структуру, где просодические элементы (паузы, интонация) служат ключевыми маркерами, сигнализирующими как о глобальной организации текста, так и о когнитивных процессах говорящего [Кибрик, Подлесская, 2009. С. 55–101]. Более того, просодический уровень также может сигнализировать о функционировании когнитивной системы говорящего: так, длинные паузы (более 1 секунды) можно охарактеризовать как переходы между эпизодами нарратива, а интонационный контур может служить полноценным дискурсивным маркером [Кибрик, Подлесская, 2006].

Корпус «Один речевой день» (ОРД) также представляет эмпирические данные для исследования повседневной речи в различных лингвистических аспектах: от социолингвистики до дискурсивного анализа [Богданова-Бегларян и др., 2019]. ОРД фиксирует состояние современного разговорного языка, что позволяет получить объективную информацию об использовании тех или иных дискурсивных и прагматических единиц в естественной речи. Под прагматическим маркером понимается неосознанно порожденная единица, вербализующая «отношение говорящего к процессу речепорождения», но не несущая в себе информативной нагрузки [Богданова-Бегларян, 2021]. Более того, обращение к данным ОРД позволяет отследить динамику языковых изменений, так как в корпусе присутствуют записи, собранные 7–15 лет назад [Шерстинова, 2023].

Различные методы дискурсивного анализа дают полноценное представление о выражении языка в повседневном и контекстуальном выражении. Перспективным направлением представляется сопоставление параметров повседневного дискурса с данными корпуса рассказов о пережитом.

В данной статье исследуются особенности лексического выбора (номинаций) и хезитаций в речи людей, переживших травмирующее событие, что позволит впоследствии провести сопоставление с данными повседневного общения. Данная работа не только выявляет особенности языка травмы, но и формирует критерии для автоматической обработки соответствующих текстов с помощью методов компьютерной лингвистики.

Материал исследования

Поскольку язык травмы не имеет устоявшейся методологии исследования, его изучение требует междисциплинарного подхода, объединяющего методы анализа эмоциональной речи из смежных областей: психолингвистики, корпусной лингвистики, дискурс-анализа и компьютерной лингвистики.

Чтобы работать с коллективной травмой в дискурсивном аспекте, был сформирован пользовательский подкорпус, в который входили рассказы о травмирующих событиях. В рамках данного исследования материалом анализа послужили транскрипты интервью бывших заложников трагедии «Норд-Ост». Были проанализированы два документальных фильма: «Террор в Москве» (2003) и «Московская осада» (2004). Также были найдены интервью бывших заложников, записанные после 2019 года, что помогло сформировать корпус речи, в которой может присутствовать язык травмы. Всего было проанализировано 5 интервью, общей продолжительностью 175 минут.

Видео были транскрибированы с помощью модуля Whisper¹ в среде программирования Python. Полученный корпус состоит из 24 665 токенов. Среди информантов две женщины и четверо мужчин, на момент интервью все говорящие старше 30 лет (табл. 1).

Таблица 1

Информация о говорящих/информантах

Table 1

Information about the speakers					
Информант	Год интервью	Тип пострадавшего	Пол	Возраст	Метаданные
И1	2021	косвенный	Ж	60	Потеряла сына в теракте
И2	2021	прямой	М	50	Преподаватель в театре, на момент теракта находился в заложниках со своими учениками
И3	2004	прямой	—	—	Документальный фильм, речь не сегментирована на информантов
И4	2003	прямой	—	—	Документальный фильм, речь не сегментирована на информантов
И5	2019	прямой	М	30	На момент теракта был подростком 14 лет
И6	2019	косвенный	М	50	На момент теракта работал дублером, смог выбраться из театра в первые минуты теракта
И7	2023	прямой	М	35	На момент теракта был подростком 15 лет
И8	2020	прямой	Ж	60	На момент теракта являлась гражданкой США

¹ Whisper: Automatic Speech Recognition (ASR) system: [программный код] / OpenAI. Версия 1.0. 2022. URL: <https://pytorch.org/project/whisper/> (дата обращения: 23.01.2025).

Поскольку в документальных фильмах сюжет формируется путем монтажа высказываний разных людей, транскрипты не были разбиты по отдельным говорящим, а отмечены в таблице как И3 и И4.

Методы корпусной лингвистики предполагают составление частотных словарей лемм и словоформ, позволяющих определить закономерности в реализации тех или иных лексем. Наиболее частотные леммы в словаре могут отразить, с помощью каких функциональных слов говорящие преимущественно строят свою речь, в то время как гапакс может дать представление о лексическом наполнении анализируемого корпуса [Ляшевская, Шаров, 2008].

Результаты исследования

Количественный анализ: составление частотных словарей лемм и расчет лексической дисперсии

Лемматизация корпуса и присвоение характеристик частей речи производилась в программе MyStem². Составление частотных словарей и расчет статистик были произведены в программе Excel.

Частотность возникновения слова в корпусе не может отразить равномерность употребления леммы среды всех информантов. Некоторые говорящие могут использовать одно слово чаще, чем другие, что повлияет на общие результаты статистик, поэтому корпус был также проанализирован с помощью пакета «qanteda»³ на языке программирования R, что позволило рассчитать лексическую дисперсию для интерпретации анализируемых лексем. Эта программа позволяла отследить, насколько часто анализируемая словоформа или словосочетание появлялись во всех текстах корпуса.

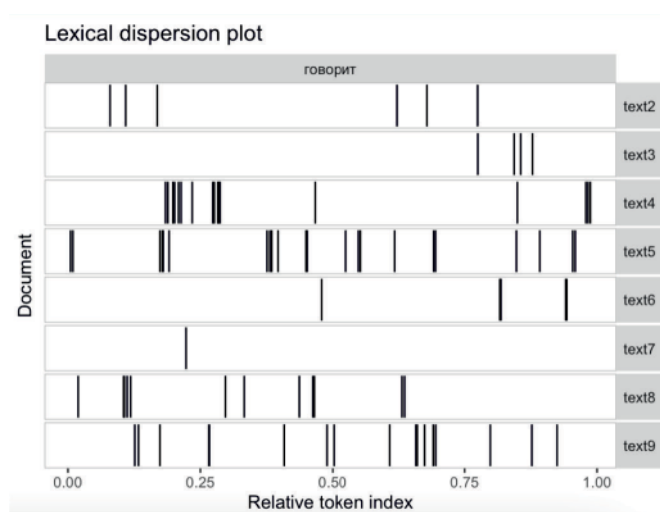


Рис. 1. Лексическая дисперсия словоформы «говорит»

Fig. 1. Lexical dispersion of «говорит»

В рамках анализа частотных словарей лемм и словоформ были обнаружены закономерности, характерные для устного дискурса в целом: использование прагматических маркеров (ПМ). Например, маркер ввода чужой речи (ксенопоказатель) «говорить» (отн. част. 0,932 %, ipm: 9324,95) (рис. 1) помогает вводить слушателя в контекст ситуации. В примере (2) можно

² MyStem: [программное обеспечение] / Yandex. 2023. URL: <https://yandex.ru/dev/mystem/> (дата обращения: 23.01.2025).

³ Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Mu 'ller, and Akitaka Matsuo. 2018 "qanteda: An R package for the quantitative analysis of textual data. Journal of Open Source Software 3(30), 774.

увидеть ПМ «говорю», который выполняет также и хезитационную функцию, часто встраиваясь несколько раз в одну реплику:

(1) И2: я **говорит** тоже / был заложником Норд-Оста / и мне кто-то из ребят потом передал вашу записочку листочек // я как сидел вот так вот / и читал *В / и начался штурм // вот когда началось // вот потом так вот / когда все закончилось / я просто открыл глаза *П с листочком в руках все // встал и пошел // *П *О спасибо **говорит** за листочек // Я **говорю** ну *В это не мне спасибо //;

(2) И1: звонит и **говорит** / слушай / твой Данила дома или где? // я **говорю** / он на работе // а он тебе звонил или нет? // я **говорю** / нет / не звонил // а что случилось? // захватили Норд-Ост террористы // я **говорю** как? // она **говорит** вот так // включи телевизор //.

Другой частотный ПМ «вот» (отн. част. 1,399%, ipm: 13987,43) выступает в роли маркера старта или финала (примеры (3)-(7)). Ср.:

(3) И1: все **вот** *В это **вот** ужас *Х //;

(4) И2: и **вот** его и я // он и более боялся // и **вот** его я посадил рядом с собой // **вот** он сидел рядом со мной //;

(5) И3: *В **вот** // что-то по-чеченски очень сильно ругался / *В **вот** //;

(6) И4: **вот** прям **вот** как () будто по команде // и **вот** весь зал захрапел // *В **вот** прям громко так //;

(7) И5: **вот** например помню эпизод / когда реально там у человека началась () *В паника / если мы все умрем (...) там все будет плохо //.

Оба этих маркера характерны для всех текстов подкорпуса, что показывает их равномерное распределение (рис. 2). Устойчивая частотность и функциональная нагрузка маркеров характерна для повествовательных структур, независимо от эмоциональной окраски текста, так как они выполняют текстоорганизующие функции [Богданова-Бегларян, 2021].

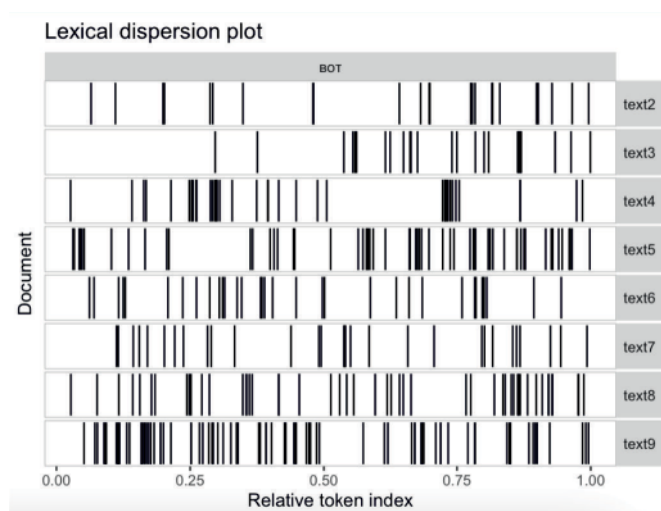


Рис. 2. Лексическая дисперсия словоформы «вот»

Fig. 2. Lexical dispersion of «вот»

Закономерности сохраняются при различной эмоциональной нагрузке текстов, что согласуется с концепцией базовой нарративной структуры, которая остается устойчивой независимо от тематики или эмоционального регистра [Кибрик, 2010].

Часто встречались союз «потому что» (отн. част. 0,58 %) и пояснительный союз «то есть» (отн. част. 0,51 %). Если употребление прагматического маркера довольно очевидно для устной речи, то появление уточняющих союзов может свидетельствовать об определенном паттерне разговора о травме, так как подавляющее большинство говорящих используют их не только

для пояснения происходящих вокруг действий, но и в качестве некоторого оправдания своих эмоций. Ср.:

(8) И8: *и периодически мы просто играли в ка(:)рты // потому что любой нормальный человек такого длительного напряжения / *В не сойдя с ума / не выдержит //*

(9) И3: *то есть тогда очень стало страшно // то есть мы поняли / что мы никому не нужны // совершенно никому //*

В примере (8) говорящий объясняет не совсем типичное для заложника поведение, а появление одного и того же маркера в высказывании несколько раз, как в примере (9), может свидетельствовать о фокусировании на эмоциональном состоянии.

Очевидно, что среди самых часто употребляемых слов будут находиться служебные части речи, которые служат каркасом для построения высказывания. Однако не стоит пренебрегать их анализом, так как, по мнению Пеннебейкера, функциональные слова могут отражать особенности высказывания, а также внутренний мир говорящего [Pennebaker, 2011]. При анализе незначительных частей речи, которые занимают верх частотного словаря лемм, интересно, что первое место занимает не союз «и» (отн. част. 3,211 %, ipm: 32110,3), а местоимение «я» (отн. част. 4,131 %, ipm: 41313,6), что расходится с данными Нового частотного словаря русской лексики [Ляшевская, Шаров, 2009], где «и» занимает лидирующую позицию, а лемма «я» (ipm 12684,4) находится под пятым рангом. Предполагаю, что различие в относительной частоте употребления лемм «я» и «и» может говорить о том, что тексты демонстрируют акцент на личном опыте говорящих, нежели повседневный нарратив.

Анализируя частотные знаменательные части речи, стоит обратить внимание на существительное «человек» (отн. част. 0,681 %, ipm: 6811,27). Интересно, что большинство информантов с помощью него либо вводят информацию о нераспознанном референте, причем множественную форму «люди» намного чаще (абс. част. 121), чем единственную (абс. част. 47), также они используют местоимение «они» (отн. част. 1,476 %), вне зависимости террорист это или, наоборот, заложник:

(10) И5: *вот например помню эпизод / когда реально там у человека началась () *В паника / если мы все умрем (...) там все будет плохо // *В я видел как несколько раз выходил на сцену человек / выставил телевизор / говорил там смотрите что о вас говорят ваши власти //*

(11) И2: *и к ним / еще двоих моих ребят / присоединили // они пошли на выход // по крайней мере хотя () бы двоих // правда через полчаса мои двое вернулись // *Х *В их на выходе сказали что(:) / они / еще / *П не такие маленькие / вернули обратно в группу // *В и когда я спрашивал / ну это же дети / почему вы их держите? // они же(:) / и так взрослых много в заложниках // *В они сказали / у нас в этом возрасте воюет уже там // и они потенциальные воины // то есть вот такие были их аргументы //*

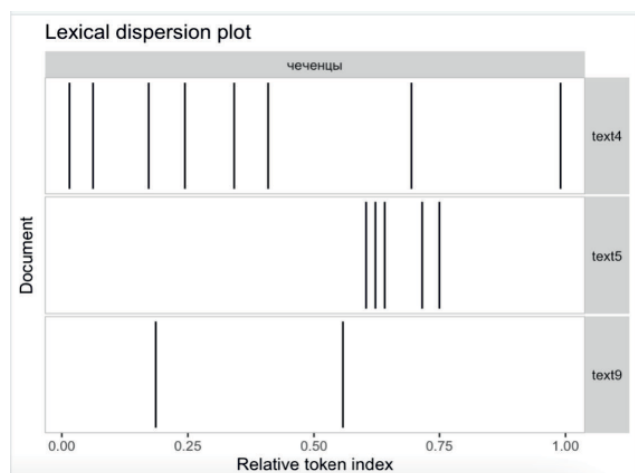


Рис. 3. Лексическая дисперсия слова «чеченцы»

Fig. 3. Lexical dispersion of «чеченцы»

Вопрос референции особенно важен, когда говорящие упоминают террористов в своих нарративах. Так, слово «чеченец» (отн. част. 0,122 %) и «террорист» (отн. част. 0,118 %) непродуктивны в дискурсе некоторых говорящих, что показал расчет лексической дисперсии. Так, «чеченцами» террористов называют герои документального фильма 2003 года (рис. 3), которые давали интервью после произошедших событий (на графике распределения лексической дисперсии интервью отмечены как text4 и text5), а также информантка, иммигрировавшая в США, интервью с которой было записано спустя 20 лет после произошедших событий.

Словоформа «террористы» встречается в большем количестве текстов, хотя имеет меньшую частотность по всему корпусу (рис. 4). Исключением стали интервью женщины, которая потеряла сына, но не была свидетельницей террористических действий, а также интервью женщины, которая на данный момент является гражданкой США. Обе женщины не используют слово «террористы».

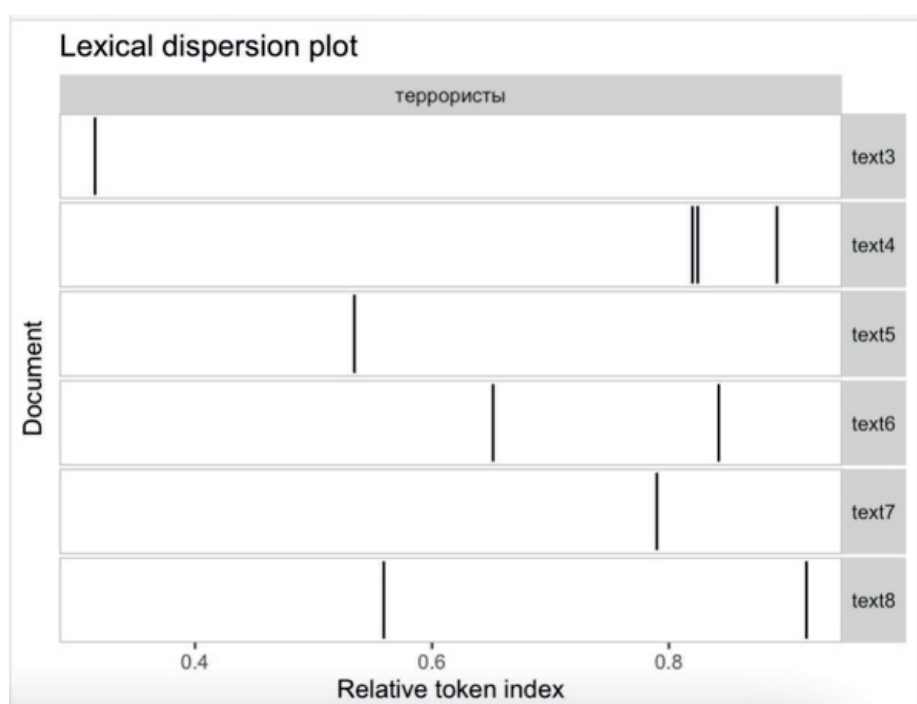


Рис. 4. Лексическая дисперсия слова «террористы»

Fig. 4. Lexical dispersion of «террористы»

Заметим, что, помимо референции, связанной с указанием на этническую принадлежность или деятельность террористов, несколько разных говорящих используют также лексику «ребята» (рис. 5), а триграмма «ребята с автоматами» встречается в корпусе три раза, как в примерах (12) – (14):

(12) И5: *опять же / это факты // *В что здесь сидят девушки с(:) бомбами / что там есть **ребята с автоматами** //*

(13) И8: *и как только мы вернулись / буквально 5–10 минут после начала второго акта / пришли **ребята** // ничего я не поняла / ничего //*

(14) И4: **В и тут же в зал там *П соков много стали раздавать *Х // некоторые *П молодые террористы **ребята** что мы их увидели / они сняли маски // *В ну мне казалось что они устали / потому что мы-то спали / а эти **ребята** они все время что-то ходили / там нервничали / переживали //*

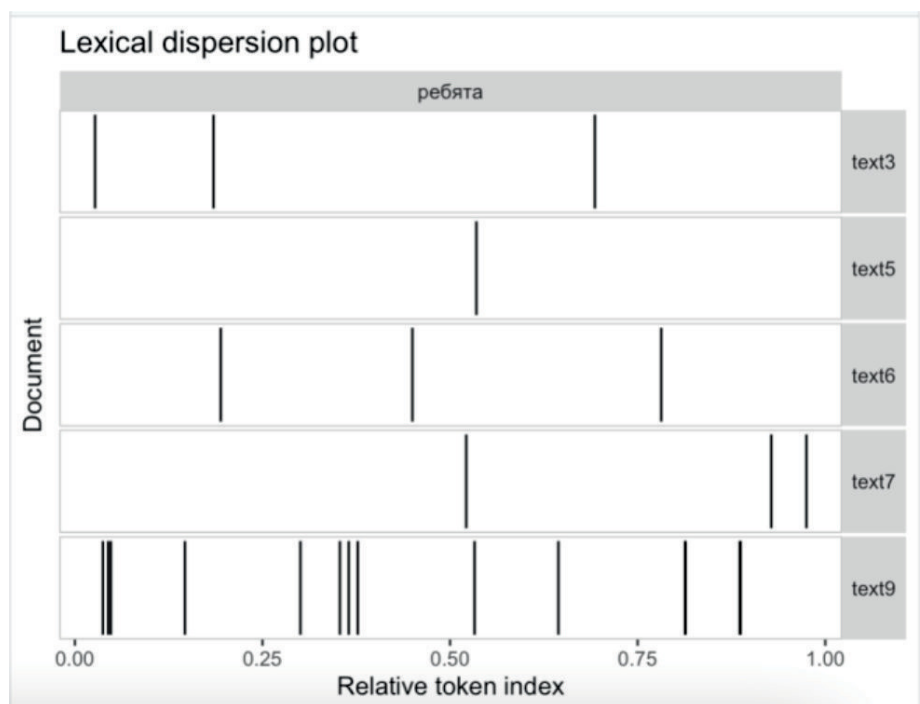


Рис. 5. Лексическая дисперсия слова «ребята»

Fig. 5. Lexical dispersion of «ребята»

Конечно, это не единственная референция, которая реализуется с помощью этой словоформы. Информанты отсылали и к детям, с которыми они проводили дни в заложниках.

Помимо анализа некоторых лексических и прагматических аспектов, в данной работе была проведена ручная разметка хезитационных особенностей и их подсчет.

Подсчет знаков хезитации

Помимо количественного анализа, стоит также обратиться к хезитационным особенностям, которые могут возникать в процессе рассказа о пережитом травматическом событии. Для поиска и подсчета паралингвистических маркеров пользовательский подкорпус был размечен согласно системе аннотирования проекта «Один речевой день» [Шерстинова, 2009]. При шифровке текстов учитывались вдохи (*В), выдохи (*Х), вздохи (*О), заминки (.), затяжки звука (:), паузы (*П), а также хезитации (...). Предполагается, что информация о частотностях хезитационных особенностей не только может различаться с данными о повседневной речи, но и будет характерна для конкретного участка нарратива. Именно поэтому для выявления закономерности между хезитационным контуром речи и описываемым событием, каждый текст был поделен на четыре части:

- 1) начало повествования,
- 2) ввод в контекст,
- 3) описание действий до теракта: рассказ о пребывании в заложниках в течение трех дней (самый длинный по токенам отрезок),
- 4) описание штурма и своих ощущений от действия газа и окончание повествования – рассказ о жизни после теракта и о продолжающихся страхах.

В результате подсчетов хезитационных явлений была получена таблица распределения паралингвистических коррелятов, возникавших во всех высказываниях на конкретном участке рассказа. Данные представлены в процентном соотношении.

Таблица 2

Распределение паралингвистических коррелятов в речи информантов

Table 2

Distribution of paralinguistic correlations in the speech of informants

Элементы хетизации	Начало повествования, %	Три дня в заложниках, %	Штурм, %	Жизнь после теракта, %
Вдохи	10,50	48,81	12,18	28,51
Выдохи	5,68	61,36	17,05	15,91
Вздохи	12,66	54,43	8,86	24,05
Паузы	4,78	52,61	13,48	29,13
Заминки	13,24	53,88	14,16	18,72
Хезитации	15,42	59,29	11,46	13,83
Затяжки звука	7,06	60,39	17,25	15,29

Чаще всего в корпусе встречаются вдохи (абс. част. 1010), что характеризует общую саморегуляцию при построении речевого высказывания. Однако в середине нарратива большинство интервьюируемых дышат не только перед главной клаузой, но и в ее середине, как в примере (15):

(15) И4: на всякий случай просто у меня возникла мысль / надо подсесть поближе к этой бомбе / познакомиться *В с женщиной / которая *В контролирует взрыватель // *П еще вдруг удастся там не знаю *П / в какой-то *В критический момент (...) / отвести руку или выдернуть шнур // мы разговаривали очень о многом *Х / о роли женщины *В в исламе / о(:)б искусстве // () ей было неловко в этом сознаться / но очевидно ей понравился спектакль //.

Вторым по частотности хезитационным феноменом является затяжка звука (абс. част. 255), которая может отражать, как работает мышление рассказчика. Тем не менее нельзя утверждать, что затяжка распространена среди всех исследуемых говорящих: например, в монологе женщины, которая не находилась в заложниках, но которая потеряла сына, данная особенность встречается лишь трижды. Это может отражать корреляцию между типом травматического опыта (прямое или косвенное переживание события) и хезитационными особенностями речи. Однако в речи большинства информантов затяжка звука демонстрирует устойчивую частотность, при этом ее употребление достигает пика в средней части высказывания, как показано в следующем примере (16).

(16) И4: они сказали / да знаем вас(:) / все эти ФСБшные штучки // во(:)т ее значит открывают дверь / они ее туда заводят / вот так ее рукой толкают () / в(:) спину // и(:) три пули ей / из Калашникова / пустил // это было видно / я прям сидел напротив двери//.

Такая же степень частотности распределения по всему корпусу у хезитации, что отражает сомнения или возможность подумать над выбором лексического наполнения высказывания. Хезитации особенно часто возникают в тех фрагментах нарратива, где говорящие воспроизводят диалоги или пытаются восстановить в памяти детали пережитого. Ср.:

(17) И2: и в этот момент подошел ко мне один из (...) террористов / там был такой / небольшого росточка / какой-то странный парень // он () подошел ко мне / сел / начал рассказывать / хвалиться каким-то там (...) пистолетом / (...) необычным оружием // они были довольны () / интересно вооружены // то есть их оттуда (...) / снабдили // то есть они были упакованы *В очень нормально // я имею в виду так *В основной / боевики // те которые пу-

шечное мясо / они в шароварах / () в свитерах / в балаклавах // и все // *В и вот он подошел / что-то мне рассказывает // знаете как интересно? // *В то есть (...) я его слушал / (...) сцена была за мной //.

Далее по частотности следуют абсолютные паузы (остановка речи на срок от 0,3 секунды и больше) со значениями 230. Паузы также являются типичным для речепроизводства дискурсивным элементом, так как после них может вводиться новый топик, или же говорящие могут использовать это время для отдыха перед произношением новой клаузы: в этом случае довольно часто возникает комбинация из паузы и вдоха. Также паузы возникают в середине клаузы (см. (18)–(29)), что может свидетельствовать о том, что говорящим тяжело произносить особенно эмоциональную информацию. Как правило, паузы возникают в середине малого дискурсивного элемента, когда интервьюируемые говорили об убийствах или других особенно страшных событиях.

(18) И2: *В но потом *П нашли его тело в(:) подвале / что его тоже расстреляли //;

(19) И2: но потом / когда была попытка () наших *П зрительниц двух девушек / они попросили в туалет / а потом *П из окна туалета выпрыгнули / и началась перестрелка между(:) нашими *П спецназовцами и тем / кто был между ними // (...) кто-то из них был даже ранен //.

Аналогичная закономерность наблюдается и для заминок – они короче пауз и характерны для быстрой речи, возникая в моменты поиска подходящей языковой формы при спонтанном порождении речи, а также в ситуациях эмоционального напряжения при вербализации травматического опыта, как в примере (20). Ср.:

(20) И2: это от меня было () в десяти метрах // и он делает скрутку / соединяет две бомбы / *В а дальше понимает что произойдет / и так далее // вот тут / *П был момент когда / *П стало страшно // я просто закрылся / () чтобы мои эмоции не могли увидеть окружающие / то есть дети //.

Самыми последними хезитационными особенностями по частотности являются выдохи (абс. част. 88). Выдохи, как правило, составляют комбинацию с вдохом, обрамляя большую или малую клаузу с двух сторон:

(21) И4: а мы с Андрюшей настолько сильно дрожали / и все остальные вокруг нас настолько сильно дрожали / что *В наш ряд *Х просто ходил ходуном //;

(22) И4: *В в зал вошла девушка *Х // это было несколько неожиданно / но я почему-то (...).

Вздохи же представляют из себя самую малочисленную группу хезитационных феноменов, но они являются самой сильной группой из всех особенностей. Они возникают в период сильных эмоций, могут сопровождаться паузами или плачем:

(23) И2: от меня скрывали // *П *О что-то частенько я // раньше был покрепче // *П каждый раз 26 октября / в день () освобождения / когда мы () перезваниваемся / переписываемся / поздравляем друг друга со вторым днем рождения *О // () вот я / это уже традиционно // я езжу туда на кладбище захожу / в храм покупаю свечечки / *В подаю записочки поминальные / прихожу на могилу / *О приношу цветы / зажигаю свечки / вспоминаю общаюсь с ребятами //.

Подсчет знаков аннотации подтверждает, что особенный подъем всех паралингвистических особенностей можно наблюдать в тех фрагментах нарратива, где говорящие описывают три дня, проведенные в заложниках в театре. Особенно интересно, что выдохи не были характерны для рассказчиков в начале повествования, но возрастают в десять раз и становятся самым частым коррелятом из всех анализируемых, что также наблюдается и при разговоре о штурме театра.

Проведенный анализ хезитационных и прагматических маркеров позволил выделить закономерности в построении нарративов о травматическом событии. Распределение паралингвистических коррелятов зависит от фазы повествования. Максимальная плотность хезитационных особенностей наблюдается в середине нарратива, что связано с высокой эмоциональной нагрузкой, которая осложняет вербализацию травматического опыта. Вдохи и выдохи чаще обрамляют смысловые блоки, выполняя регулятивную функцию при построении речи, но рост

показателей в эмоционально напряженных фрагментах (при описании штурма) указывает, что вдохи и выдохи можно отнести к маркерам стресса. Вздохи же являются наименее частотными, но и наиболее аффективно насыщенными, так как они сопровождают кульминационные моменты рассказа, часто сочетаясь с плачем или длительными паузами. Также тип травматического опыта влияет на паралингвистический профиль нарратива: в речи прямых свидетелей событий можно обнаружить большую вариативность хезитаций, особенно при описании деталей штурма, в то время как в речи косвенных участников (родственники погибших) чаще встречаются длинные паузы и вздохи, что коррелирует с фокусом на эмоциональных последствиях, а не деталях события.

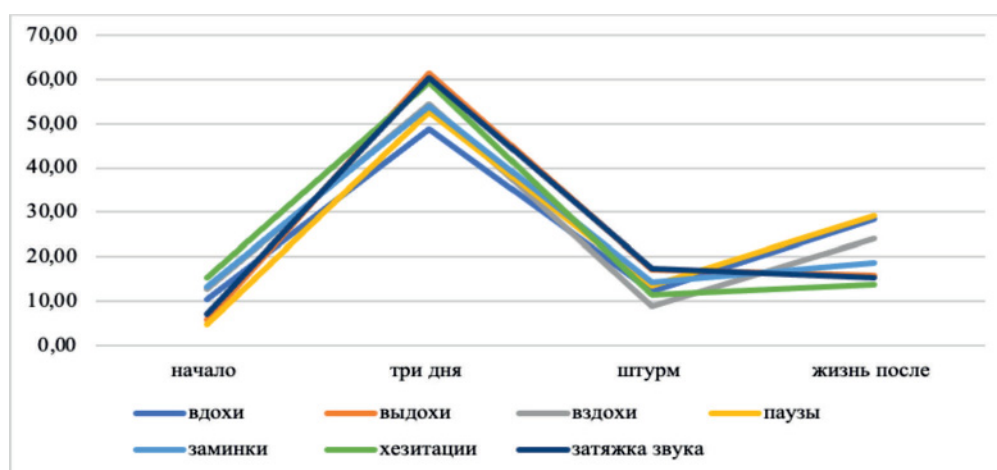


Рис. 6. Распределение паралингвистических коррелятов в речи информантов
Fig. 6. Distribution of paralinguistic correlations in the informants' speech

Также стоит обратить внимание на лексико-грамматический уровень исследуемых текстов. Например, доминирование местоимения «я» и пояснительных конструкций, с одной стороны, подтверждает персонализацию нарратива, типичную для рассказа от первого лица, но с другой стороны, отражает потребность в рационализации опыта коллективной травмы. Это может также проявляться в референции к террористам («чеченцы», «террористы», «ребята с автоматами»), которая варьируется в зависимости от временной дистанции и культурного контекста говорящего, отражая стратегии дистанцирования или опосредованного обозначения. Прагматические маркеры («говорить», «вот») сохраняют текстообразующую функцию.

Таким образом, язык травмы представляет собой уникальный лингвистический феномен, где лингвистические и паралингвистические корреляты становятся знаковыми элементами смыслопорождения. Дальнейшие исследования требуют расширения корпуса и разработки специализированных аннотационных схем, учитывающих взаимодействие вербальных и невербальных компонентов травматического дискурса.

Исследование языка травмы в лингвистическом аспекте имеет свои перспективы развития. Предполагается, что, помимо хезитационного и лексического уровней, большую роль в интерпретации выражения переживания может играть и интонационный уровень. Более того, сравнение полученных лингвистических маркеров с языковым материалом, в котором нарратив не связан с рассказом о травмирующем событии, может помочь сформировать представление об особенной реализации языка травмы, который отличается от повседневного дискурса. Первичные результаты позволили обнаружить особенную косвенную номинацию агрессора, тяготение к построению причинно-следственных связей и повышенную хезитационную активность. Расширение корпуса может выступать одним из перспективных направлений

для дальнейшего исследования языка травмы. Включение нарративов о других коллективных травмах может помочь выявить универсальные маркеры коллективной травмы, независимые от контекста. Также исследование может быть продолжено после более глубокой разметки материала с точки зрения интонационных паттернов.

Список литературы

- Александр Дж., Куракин Д. Ю.** Культурная травма и коллективная идентичность // Социологический журнал. 2012. №. 3. С. 005–040.
- Богданова-Бегларян Н. В.** и др. Корпус русского языка повседневного общения «Один речевой день»: текущее состояние и перспективы // Тр. Ин-та русского языка им. В. В. Виноградова. 2019. Т. 21. С. 101–110.
- Богданова-Бегларян Н. В.** (ред.). Прагматические маркеры русской повседневной речи: словарь-монография. Nestor-Istoriia, 2021.
- Ван дер Колк Б.** Тело помнит все. Какую роль психологическая травма играет в жизни человека и какие техники помогают ее преодолеть. М.: Эксмо, 2014.
- Кибрик А. А., Подлесская В. И.** Проблема сегментации устного дискурса и когнитивная система говорящего // Когнитивные исследования. 2006. С. 138–158.
- Кибрик А. А., Подлесская В. И.** (ред.) Рассказы о сновидениях: Корпусное исследование устного русского дискурса. М.: ЯСК, 2019.
- Кибрик А. А.** Лингвистическая теория нарратива У. Лабова // Моделирование языка и когнитивных процессов. М.: Языки славянских культур, 2010. С. 85–112.
- Лукашевич Н. В., Левчик А. В.** Создание лексикона оценочных слов русского языка РуСентилекс // Тр. конф. OSTIS-2016. 2016. С. 377–382.
- Ляшевская О. Н., Шаров С. А.** Частотный словарь национального корпуса русского языка: концепция и технология создания. URL: <http://www.dialog-21.ru/digests/dialog2008/materials/html/53.htm>. 2008.
- Ляшевская О. Н., Шаров С. А.** Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009.
- Мультиканальный (мультимодальный) корпус «Рассказы и разговоры о грушах» / А. Кибрик, В. Подлесская. URL: <https://multidiscourse.ru/corpus/> (дата обращения: 31.03.2025).
- Федорова О. В., Кибрик А. А.** Мультиканальный корпус русского языка: принципы аннотирования и анализ дискурсивных маркеров // Компьютерная лингвистика и интеллектуальные технологии. 2018. № 17 (24). С. 780–795.
- Шерстинова Т. Ю., Рыко А. И., Степанова С. Б.** Система аннотирования в звуковом корпусе русского языка «Один речевой день» // Формальные методы анализа русской речи. Материалы XXXVIII Междунар. филол. конф. СПб.: СПбГУ. 2009. С. 66–75.
- Шерстинова Т. Ю.** Моделирование повседневного речевого поведения: корпус устной речи молодежи, или ОРД V. 2.0. 2023.
- Busch B., McNamara T.** Language and trauma: An introduction // Applied Linguistics. 2020. Vol. 41. No. 3. P. 323–333.
- Erikson K. T.** Everything in its path. Simon and Schuster, 2012.
- Golubev A., Loukachevitch N.** Transfer learning for improving results on Russian sentiment datasets // arXiv preprint arXiv:2107.02499. 2021.
- Mitchell J.** Trauma, recognition, and the place of language // Diacritics. 1998. Vol. 28. No. 4. P. 121–133.
- Pennebaker J. W.** The secret life of pronouns // New Scientist. 2011. Vol. 211. No. 2828. P. 42–45.
- Pennebaker J. W., Chung C. K.** Expressive writing, emotional upheavals, and health // Foundations of health psychology. 2007. P. 263–284.

Van der Kolk B. A., Van der Hart O. The intrusive past: The flexibility of memory and the engraving of trauma // *American imago*. 1991. Vol. 48. No. 4. P. 425–454.

References

- Alexander J., Kurakin D.** Cultural trauma and collective identity. *Sociological Journal*, 2012, no. 3, pp. 005–040. (in Russ.)
- Bogdanova-Beglarian N. V.** et al. The Russian Everyday Communication Unit «One Day of Speaking»: current state and prospects. *Works of the Institute of Russian Language. BB Vinogradova*, 2019, vol. 21, pp. 101–110. (in Russ.)
- Bogdanova-Beglarian N. V.** (ed.). Pragmatic markers of everyday Russian speech: dictionary-monograph. Nestor-Istorii a, 2021. (in Russ.)
- Busch B., McNamara T.** Language and trauma: An introduction. *Applied Linguistics*, 2020, vol. 41, no. 3, pp. 323–333.
- Erikson K. T.** Everything in its path. Simon and Schuster, 2012.
- Golubev A., Loukachevitch N.** Transfer learning for improving results on Russian sentiment datasets. *arXiv preprint arXiv:2107.02499*. 2021. (in Russ.)
- Kibrik A. A., Podlesskaya V. I.** The problem of segmentation of oral discourse and the speaker's cognitive system. *Cognitive Studies*, 2006, pp. 138–158. (in Russ.)
- Kibrik A. A., Podlesskaya V. I.**, eds. Stories about dreams: Corpus study of oral Russian discourse. Moscow: YASK publ., 2019. (in Russ.)
- Kibrik A. A.** Linguistic theory of narrative by W. Labov. In: *Modeling of language and cognitive processes*. Moscow: Languages of Slavic cultures, 2010, pp. 85–112. (in Russ.)
- Mitchell J.** Trauma, recognition, and the place of language. *Diacritics*, 1998, vol. 28, no. 4, pp. 121–133.
- Lukashevich N. V., Levchik A. V.** Creating a lexicon of evaluative words of the Russian language RuSentilex. *Proceedings of the OSTIS-2016 conference*, pp. 377–382. (in Russ.)
- Lyashevskaya O. N., Sharov S. A.** Frequency dictionary of the National Corpus of the Russian language: Concept and creation technology. URL: <http://www.dialog-21.ru/digests/dialog2008/materials/html/53.htm>. 2008. (in Russ.)
- Lyashevskaya O. N., Sharov S. A.** Frequency dictionary of modern Russian language (based on the materials of the National Corpus of the Russian language). Moscow: Azbukovnik publ., 2009. (in Russ.)
- Multi-channel (multimodal) corpus «Stories and talks about pears». A. Kibrik, V. Podlesskaya. URL: <https://multidiscourse.ru/corpus/> (accessed: 31.03.2025). (in Russ.)
- Sherstinova T. Y., Ryko A. I., Stepanova S. B.** System of annotations in the sound body of the Russian language «One speech day». *Formal methods of analysis of Russian speech. Mat-ula XXXVIII Int. phil. conf. Spa: SPbHU*, 2009, pp. 66–75. (in Russ.)
- Sherstinova T. Y.** Everyday speech behavior: the body of youth oral speech, or ORD V. 2.0. 2023. (in Russ.)
- Van der Kolk B.** The Body Keeps the Score. What psychological trauma plays in human life and what techniques help to overcome it. Moscow: Eksmo publ., 2014. (in Russ.)
- Van der Kolk B. A., Van der Hart O.** The intrusive past: The flexibility of memory and the engraving of trauma. *American Imago*, 1991, vol. 48, no. 4, pp. 425–454.
- Fedorova O. V., Kibrik A. A.** Multi-channel corpus of the Russian language: principles of annotation and analysis of discursive markers. *Computational linguistics and intellectual technologies*, 2018, no. 17(24), pp. 780–795. (in Russ.)
- Pennebaker J. W.** The secret life of pronouns. *New Scientist*, 2011, vol. 211, no. 2828, pp. 42–45.

Pennebaker J. W., Chung C. K. Expressive writing, emotional upheavals, and health. *Foundations of health psychology*, 2007, pp. 263–284.

Информация об авторах

Попова Татьяна Ивановна, кандидат филологических наук, доцент департамента филологии НИУ ВШЭ СПб

Наумова Виолетта Владимировна, студентка НИУ ВШЭ СПб

Information about the Authors

Tatyana I. Popova, Candidate of Philology, Associate Professor of the Department of Philology HSE University Saint-Petersburg

Violetta V. Naumova, Student HSE University Saint-Petersburg

*Статья поступила в редакцию 06.11.2025;
одобрена после рецензирования 14.12.2025; принята к публикации 06.02.2026*

*The article was submitted 06.11.2025;
approved after reviewing 14.12.2025; accepted for publication 06.02.2026*

Научная статья

УДК 81'322.2

DOI 10.25205/1818-7935-2026-24-2-112-123

**«На Гавайские, Липарские острова не поскачешь.
Не советские они еще пока что...»:
топонимы в русском рассказе 1930-х годов**

Яна Константиновна Хлусова

Национальный исследовательский университет «Высшая школа экономики»
Санкт-Петербург, Россия

yakkhlusova@edu.hse.ru, <https://orcid.org/0009-0004-3619-8191>

Аннотация

В статье представлены результаты корпусного исследования именованных сущностей, обозначающих географические локации (LOC), в русском рассказе 1930-х годов. Материалом для исследования послужили 100 текстов из Корпуса русского рассказа XX века, отобранных по принципу равномерного распределения по годам (по 10 рассказов на каждый год 1931–1940 гг.), и принадлежащих разным авторам, что позволило сформировать репрезентативную выборку. Автоматическое извлечение локаций осуществлялось с помощью библиотеки *Natasha*, нормализация – на основе большой языковой модели *Gemma2-9b-it*, верификация результатов проводилась вручную. В результате было выявлено более 400 уникальных географических названий; для 331 из них определены точные координаты, что позволило отразить их на карте. Пространственное распределение охватывает как СССР, так и зарубежные страны, с наибольшей концентрацией точек в европейской части СССР, Средней Азии и на юге Сибири. Проведен контекстуальный анализ упоминаний западных стран, часто выступающих в роли культурного контраста к Советскому Союзу, восточных топонимов, упоминающихся эпизодически и без выраженной идеологической нагрузки, а также союзных республик, изображаемых как органическая часть единого советского пространства. Вымышленных локаций практически не обнаружено, что отражает реалистическую парадигму литературы данного периода. Результаты исследования вносят вклад в изучение особенностей малой русской прозы прошлого столетия и могут быть использованы как в литературоведческих, так и в междисциплинарных исследованиях, включая культурную географию и цифровую гуманитаристику.

Ключевые слова

русский рассказ, корпусные исследования литературы, именованные сущности, топонимы, локации в художественном тексте, 30-е годы XX века

Финансирование

Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ (HSE-BR-2025-032).

Для цитирования

Хлусова Я. К. «На Гавайские, Липарские острова не поскачешь. Не советские они еще пока что...»: топонимы в русском рассказе 1930-х годов // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 112–123. DOI 10.25205/1818-7935-2026-24-2-112-123

© Хлусова Я. К., 2026

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2
Vestnik NSU. Series: Linguistics and Intercultural Communication, 2026, vol. 24, no. 2

“You Can’t Just Hop to the Hawaiian or Aeolian Islands: They’re Not Soviet Yet...”: Toponyms in Russian Short Stories of the 1930s

Yana K. Khlusova

HSE University
Saint Petersburg, Russian Federation
yakkhlusova@edu.hse.ru, <https://orcid.org/0009-0004-3619-8191>

Abstract

The article presents the results of a corpus-based study of named entities denoting geographic locations (LOC) in Russian short stories of the 1930s. The research material consists of 100 texts from the Corpus of Russian Short Stories of the 20th century, selected to ensure an even distribution across years (ten stories for each year from 1931 to 1940) and authored by different writers, thus forming a representative sample. Automatic extraction of locations was performed using the Natasha library, normalization was carried out with the large language model Gemma2-9b-it, the results were verified manually. The analysis identified more than 400 unique geographic names; precise coordinates were determined for 331 of them, enabling their visualization on a map. The spatial distribution covers both the USSR and foreign countries, with the highest concentration of locations in the European part of the USSR, Central Asia, and the south of Siberia. A contextual analysis of references to Western countries, often serving as a cultural contrast to the Soviet Union, to Eastern toponyms, appearing sporadically and without pronounced ideological connotations, as well as to the Union republics, depicted as an organic part of the unified Soviet space, was conducted. Fictional locations were virtually absent, reflecting the realist paradigm of literature of the period. The findings contribute to the study of the distinctive features of the past century Russian short prose and may be applied in both literary scholarship and interdisciplinary research, including cultural geography and digital humanities.

Keywords

Russian short stories, corpus-based literary studies, named entities, toponyms, locations in a literary text, 1930s

Funding

This article is an output of a research project HSE-BR-2025-032 implemented as part of the Basic Research Project at HSE University.

For citation

Khlusova Y. K. “You Can’t Just Hop to the Hawaiian or Aeolian Islands: They’re Not Soviet Yet...”: Toponyms in Russian Short Stories of the 1930s. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 112–123. DOI 10.25205/1818-7935-2026-24-2-112-123

Введение

Глубокие социальные и политические преобразования XX века нашли отражение в языке и литературном творчестве русского народа и всего русскоязычного культурного пространства. Для анализа этих изменений был создан Корпус русского рассказа XX века¹ – аннотированная коллекция, включающая 1000 текстов 1006 писателей, работавших на протяжении всего столетия [Шерстинова, Кирина, Хлусова, 2025]. Выбор жанра обусловлен его популярностью среди писателей и читателей, а также скоростью реакции языка малой прозы на происходящие в мире потрясения [Мартыненко и др., 2018 (а), (б)].

На материале Корпуса русского рассказа выполнен ряд исследований по текстам 1900–1930 гг. [Шерстинова, 2022]. Их предметом являлась лексическая составляющая [Гребенников, Скребцова, 2019; Гребенников, Марусенко, 2020; Скребцова, Гребенников, Шерстинова, 2021; Grebennikov, Skrebtsova, 2024], тематический аспект [Скребцова, 2020; Шерстинова и др., 2024], эмоциональная окраска [Sherstinova et al., 2025]. Данная работа продолжает изучение особенностей языка и художественного мира русского рассказа, исследуя употребление именованных сущностей типа LOC, обозначающих географические названия² [Grishman, Sundheim,

¹ <https://russian-short-stories.ru/>

² От англ. location.

1996]), в текстах более позднего времени – 30-х гг. (1931–1940 гг.). Выбранный период отмечен индустриализацией Советского Союза, коллективизацией, перетоком населения из деревень в города, ускоренными темпами производства, а также массовыми политическими репрессиями. Господствующим направлением в искусстве в то время являлся социалистический реализм, предписывающий изображение событий в их естественном течении к революции [Первый Всесоюзный съезд советских писателей, 1934. С. 716]. Интерес к топонимам данного периода обусловлен целым рядом факторов, среди которых:

- социальные и политические изменения, приведшие к переименованию географических объектов и появлению новых названий (проспект Фрунзе, Ленинский поселок, улица Лепсе);
- реформа территориального деления страны, в ходе которой губернии, уезды и волости были заменены областями, краями, округами и др.;
- смещение фокуса внимания с одних территорий на другие;
- транслирование ценностей через литературу и, в частности, через топонимы (повышенное внимание к союзным республикам Средней Азии, противопоставление СССР западным странам).

К именованным сущностям (ИС, от англ. *named entities*) традиционно относят имена людей (тип PER), названия организаций (тип ORG) и географических объектов (тип LOC) [Grishman, Sundheim, 1996]. Также в этот список могут входить даты, валюта, проценты, названия событий и др. [там же; Лагутина, Васильев, Зафиевский, 2023]. Инструменты для автоматического извлечения ИС расширяют категорию LOC, выявляя не только топонимы в тексте, но и любые именованные локации (например, дворы персонажей – ср. *Ванькин двор*, *двор Лизунихи*). В данном исследовании анализируются исключительно стандартные топонимы – географические объекты, зафиксированные в официальной или общепринятой топонимике, локальные и неформальные названия не рассматриваются.

Анализ распределения географических объектов в текстах с помощью автоматического извлечения ИС проводился ранее на материале детской литературы Сибири конца XIX – начала XX в. [Меховский, Кижнер, 2025] и советских песен 1964–1991 гг. [Колмогорова, 2025]. Обе работы направлены на реконструкцию картины мира русского человека в разные исторические периоды. Настоящее исследование предлагает иное фокусное смещение – оно обращено к литературе 1930-х гг., ориентированной на взрослого читателя и рассматривает географическое пространство глазами советского гражданина эпохи первых пятилеток и предвоенных лет.

1. Материал и методология исследования

Материалом исследования послужили 100 рассказов, дата первой публикации³ которых пришлась на 1931–1940 гг. (по 10 текстов на каждый год). Выборка включает как классиков (М. Горького, М. Зощенко, В. Шишкова), так и менее известных писателей, отобранных случайно. Формирование проведено по методологии Корпуса русского рассказа [Мартыненко и др., 2018 (а) и (б)]: отобраны русскоязычные непереводные тексты, описывающие реалии XX в., не являющиеся детской литературой. Объем произведения не учитывался. Методика вдохновлена системным подходом Ю. Н. Тынянова, предполагающим охват максимального количества писателей эпохи независимо от их известности [Тынянов, 1929]. Объем исследовательского подкорпуса составил 2 984 486 словоупотреблений.

Для выделения ИС использовалась утилита библиотеки *Natasha*⁴. Данные были нормализованы большой языковой моделью (БЯМ) *Gemma2-9b-it*⁵ [Gemma Team, 2024] и проверены

³ При отборе материала мы старались, чтобы и дата написания (если она известна) приходилась на указанный период.

⁴ <https://natasha.github.io/>

⁵ Настройки генерации, а также все таблицы, упомянутые в статье, выложены в облачное хранилище по ссылке: https://docs.google.com/spreadsheets/d/1Omu7r7Pxz75UnMjXWOZdjwvxv_pssfqDt/

экспертно. Нормализация включала объединение вариантов написания ИС, их лемматизацию, а также фильтрацию ложно извлеченных единиц (междометий, глаголов с заглавной буквы) и ИС неверного типа (например, имен или названий организаций, ошибочно отнесенных к локациям). Кроме того, были удалены дубликаты внутри рассказов, чтобы избежать смещения выборки [Меховский, Кижнер, 2025]: подсчитывалось число рассказов, содержащих локацию, а не общее количество ее упоминаний по корпусу.

После нормализации координаты объектов были получены через Nominatim API⁶ для построения карты с помощью библиотеки Plotly⁷ и прошли экспертную сверку. Спорные случаи (распространенные названия вроде проспекта Фрунзе или отсутствующие в API населенные пункты) уточнялись вручную по Яндекс/Google Картам, Wikipedia и другим источникам, в том числе с опорой на контекстные ИС рассказов. Локации с неустановленным местоположением исключены из визуализации.

2. Результаты

2.1. Общая статистика и избранные контексты

Всего было извлечено более 400 уникальных имен, для 331 из них были найдены координаты. Составленная карта изображена на рисунке.



Географическая карта топонимов,
упомянутых в русском рассказе 1930-х гг.
Geographical Map of Toponyms Mentioned
in Russian Short Stories of 1930s

Большая часть топонимов расположена в европейской части СССР (*Москва, Петербург, Крым, Украина* и др.). Вторая по концентрации область – южная часть Сибири (*Байкал, Алтай, Саяны*, прилегающие деревни) и Средняя Азия (*Узбекистан, Таджикистан, Туркменистан, Казахстан*). Отдельные объекты расположены в Заполярье (*Диксон, Земля Франца-Иосифа, Пустозерск, мыс Челюскин*), на Дальнем Востоке (*Владивосток, Магадан*). Топонимы вне СССР находятся в основном на территории Европы, отдельные точки отмечены в Америке

⁶ <https://nominatim.org/>

⁷ <https://pypi.org/project/plotly/>

(Нью-Йорк), Аргентине (Буэнос-Айрес), Африке, Монголии, Китае, Японии, Мировом океане (Тихий океан, Атлантический океан, Гавайи, Мартиника).

В список объектов, не нанесенных на географическую карту, вошли имена, местоположение которых сложно точно определить: топоним мог быть переименован, малоупотребляем или, наоборот, быть слишком распространенным, как *Привокзальная улица*, *Лесной поселок* и др. Для ряда локаций установлено примерное местоположение с помощью других географических названий, упомянутых в рассказах: село Петровка, например, находится где-то в Сибири; деревни *Латоново*, *Кирсановка*, *Куракино* – недалеко от Ростова-на-Дону; село *Попонка* – около Полтавы (Полтавской губернии); *Милятинский монастырь* находится или находился возле села Милятин на Украине; *Печкинская*, *Трубочистинская волости* могут быть вымышленными названиями, но находящимися возле Воронежа (*Воронежской губернии*, *Бобровского уезда*); *Лесной переулок* – в Москве, у фабрики «Габай» («Ява»); *Юлайское устье*, *Девичий лес*, *Курье озеро* – на Урале; *Красное Самсоновище* – недалеко от Самары и т. д.

Отдельную сложность представили названия деревень, поселков и малых городов: некоторые из них уже исчезли с карт, как, например, *Пустозерье* из рассказа Ю. П. Германа «Николка Баженов» (1936). В этом и подобном случаях удалось идентифицировать положение с помощью других локаций текста и цитат: в рассказе упомянут Архангельск, вблизи которого действительно существовал город *Пустозерск*. Также произведение начинается словами: «Николка Баженов родился на дальнем Севере, за полярным кругом, в рыбацкой деревеньке Пустозерье», что еще раз доказывает правильность найденных координат. С другой стороны, названия многих деревень могут дублироваться в разных регионах страны. В таких случаях координаты, найденные в Nominatim API, перепроверялись по другим локациям в тексте и, чаще всего, удалялись, поскольку были неверными. Подобные населенные пункты на карту не нанесены.

Наконец, координаты не указаны для таких топонимов, как *Житомирское направление* и *Тверская дорога* из-за их протяженности.

В 12 рассказах не удалось выделить локации. Эти тексты сравнительно коротки (их медианная длина составляет около 1500 слов). Они фокусируются вокруг конкретных эпизодов из жизни персонажей – самих действиях и/или диалогах, что отводит местоположение героев на второй план. Герои этих текстов представлены вне географических координат, что делает их образы обобщенными и в определенной степени обезличенными. Особенно ярко это видно в рассказе Д. Е. Зозули «Красный мешочек» (1937), где герой даже не имеет имени – повествование ведется от 3-го лица, называющего персонажа *он*. Текст повествует о некоем человеке, прошедшем через Гражданскую войну. «Сейчас он взрослый, серьезный человек, хороший член партии», но нередко он мысленно возвращается в прошлое, наполненное активной борьбой. Душу персонажа греет «красный мешочек» – подарок для красноармейцев, содержащий самое необходимое для жизни в боевых условиях («нитки, иголка, алюминиевая кружка, пуговицы»). Этот мешочек он хранит до сих пор, даже после победы «красных».

Извлеченные локации были экспертно распределены на следующие категории: Страны; Регионы; Губернии / уезды / волости; Города; Села / поселки / деревни / слободы; Станицы; Хутора; Острова / полуострова / архипелаги / мысы; Горы / холмы / хребты; Леса; Степи; Пустыни; Другие природные объекты; Косы; Океаны; Моря; Озера; Реки; Заливы / бухты; Проливы; Другие гидронимы; Монастыри; Направления; Железная дорога. Среди них самыми многочисленными стали «Города» и «Села / поселки / деревни / слободы».

Дополнительно топонимы были разделены на полярности «Запад» и «СССР», поскольку для 1930-х гг. было свойственно противопоставление «своего» «чужому» (= западному). Оказалось, что 87–88 % упомянутых в исследовательском подкорпусе локаций расположено на территории СССР. Несколько топонимов (примерно 2–3 %) могут быть отнесены к восточной части земного шара. Наконец, 10 % локаций расположено в западной части мира.

Несмотря на изменения в административном делении страны в 1920-е гг., в нескольких рассказах упоминаются уезды, волости и губернии. Это ретроспективные тексты, повествующие о событиях Гражданской войны.

Улицы, площади, парки, дворцы были объединены в группы по городам. Больше всего подобных топонимов относится к Москве и Петербургу. Также в рассказах упомянуты известные улицы Тамбова и Киева и различные точки, расположенные на полуострове Крым. Единственная западная локация, улицы и достопримечательности которой упоминаются в текстах – район Шарлоттенбург в Берлине, один из центров русской эмиграции после революции 1917 г. Рассказ, в котором идет речь об этом районе и прилегающих к нему топонимах, посвящен противостоянию комсомольцев и фашистов (Д. Лебедев «Люди и гуси», 1934).

Самой частотной локацией оказалась *Москва*: она упомянута в 39 рассказах, в текстах перечислено 17 относящихся к ней топонимов (улиц, парков и др., см. выше). Гораздо реже упоминается Петербург в разных формах (*Петроград*, (*Санкт*)-*Петербург*, *Ленинград*) (11 рассказов), 9 топонимов относится к нему. В 11 рассказах использованы имена *Россия*, *РСФСР* или *Русь*.

Русь в рассказе В. Я. Шишкова «Дивное море» (1934) обозначает «красных» – весь русский народ во время Гражданской войны. С другой стороны, у В. А. Никифорова-Волгина «Дорожный посох» (1938) *Русь* – земля отцов: «Не так ли и предки наши уносили святыни свои в леса, в укромные места, во дни татарского нашествия на Русь?» Это единственный рассказ в выборке, не совсем вписывающийся в рамки соцреализма. Повествование в нем ведется от лица верующего человека, скорбящего об антирелигиозных настроениях общества. В тексте перечисляются реальные репрессированные духовные лица. Так, координаты *Белграда-Курского* (Белгорода) были найдены с помощью жизнеописания сщмч. Никодима, некогда бывшего курским священником и расстрелянным в районе Белгорода [Субботин, 2018].

Имя *Россия* появляется в разных текстах: некоторые из них посвящены событиям Гражданской войны или более ранним, когда Советский Союз еще не был установлен; рассказы, повествующие о событиях после создания нового государства, употребляют имя *Россия* в нейтральном контексте (как обозначение страны). *РСФСР* употребляется в контексте перечисления входящих в СССР республик.

2.2. Контексты употребления топонимов западных стран

Несмотря на определенную закрытость Советского Союза в 1930-е гг., в рассказах данного периода упоминаются западные страны.

Франция – родина образцового коммуниста Филоша (В. А. Герасимова «Филош», 1932), осознавшего всю несправедливость господствующего строя еще в детском возрасте, не имея никаких источников, кроме собственных размышлений и наблюдений. Образ юноши отражает всемирность революционной идеи – коммунисты борются за правду не только на территории СССР, но и за границей. Французские *Ментон* и *Мартиника* использованы в текстах без идеологической нагрузки. Первый является местом временного пребывания одного из персонажей во времена его молодости, вторая – локацией «знаменитого землетрясения» (вероятно, 1902 г.). Наконец, в одном из рассказов *Париж* показан как точка, весьма далекая от СССР.

Америка единожды употребляется в составе фразы «наследство дядюшки из Америки», означающей везение человека. В остальных случаях эта страна употребляется в противопоставлении Советскому Союзу – она более прогрессивна: «Это уже давно делают в Америке» (В. Н. Борахвостов «Химия», 1933), в текстах описывается определенная зависимость от ее разработок, звучат призывы улучшить собственное производство и добиться технологической самостоятельности. Америка изображается как далекое и во многом мифологизированное пространство, вызывающее у советского человека удивление и ощущение технических чудес:

«Слыхал ты, там построили мост над морем, висит мост на воздухе и – ничего, висит!» (М. Горький «Бык», 1935); «Им до смерти, как до Америки, – далеко!» (там же). Говорится, что эта страна является чересчур бюрократичной: «Как в Америке, начинаем работать. Не успели еще со станка брак снять – уже, пожалуйста в кабинет директора...» (П. Русин «Маховик», 1935). В одном из эпизодов упоминается *Аляска* как родина собаки-вожака у чукчей.

Германия предстает в текстах как важный источник технических знаний и промышленного опыта, к которому Советский Союз обращается в рамках модернизационных изменений. Из нее импортируют сахар, у нее учатся станкостроению. Специально выписанный немец консультирует рабочих советского завода по листопрокату. Однако такой технологический обмен подается как временный и стратегически обусловленный шаг, направленный на достижение полной промышленной самостоятельности. Как уже было замечено выше, в рассказе Д. Лебедева «Люди и гуси» (1934) коммунисты прямо противопоставляются фашистам, что отражает идеологическую дистанцию между странами.

Дания и *Швейцария* так же описаны как более развитые в индустриальном плане страны.

Финляндия, *Швейцария*, *Швеция*, *Норвегия*, *Стокгольм* и *Скандинавия* в целом эпизодически упоминаются в рассказах, действие которых относится к событиям 1910–1920-х гг., *Ирландия* – в текстах, связанных с более ранними историческими периодами. Употребление этих топонимов не несет выраженной идеологической нагрузки: они фигурируют преимущественно в нейтральных контекстах – как страны происхождения предметов, принадлежащих персонажам, или как эпизодические точки их перемещений.

Лондон, *Буэнос-Айрес* и *Нью-Йорк* перечислены в ряде городов, импортирующих шкуры из Узбекистана. Рассказ Эль-Регистана «Хайбер, сын пустыни» (1934) демонстрирует возможности Советского Союза в экономической борьбе с капиталистическими странами: «Ведь только в пустынях и оазисах далекой Средней Азии водится каракулевая овца, шкурка которой пользуется неизменным спросом и в Нью-Йорке, и в Иркутске, и в Буэнос-Айресе, и в Москве, и в Лондоне, и в Александрии Херсонской». Иными словами, продукция союзных республик ценится по всему миру. Любопытно, но *Нью-Йорк*, *Сан-Франциско*, *Альпы* и *Карпаты* упомянуты в качестве названий пивных и посудных лавок на западе Беларуси в 1939 г. (Е. И. Габрилович «В пути», 1939). В этом случае перечисленные имена использованы не как топонимы. Данные заведения герой встречает на пути к кладбищу, причем сначала он видит «Сан-Франциско» и «Карпаты», а ближе к точке назначения – «Нью-Йорк» и «Альпы». По мере приближения к кладбищу улицы становятся беднее, дома – мельче, но повышается престиж названий лавок, что носит явный сатирический оттенок.

Англия (*Великобритания*, *Ливерпуль*), с одной стороны, как и *Ирландия*, упоминается в качестве страны-производителя предметов, принадлежащих персонажам рассказов, с другой – как посредник между СССР и Польшей во время советско-польской войны, который преследует корыстные цели – спасение панской Польши и белого генерала Врангеля (К. А. Федин «Рисунок с Ленина», 1939).

Польша, *Варшава*, *Лодзь*, *Висла* упомянуты в контексте советско-польской войны. Название страны употребляется с прилагательным «панская», что подчеркивает ее классовую природу и идеологические различия с Советским Союзом. Также упоминание города *Лодзь* имеет ностальгический оттенок: «В мирное время Карагодов знал, что, если в Лодзи прошла гроза с ливнем или в Казани наступило похолодание, он завтра известит об этом петербуржцев: у него было много добровольных корреспондентов» (Н. С. Атаров, «Календарь русской природы», 1938) – ранее этот город был частью Российской империи. Персонаж-фенолог предсказывал погоду в Петербурге по данным, полученным из разных мест, в том числе из Польши, теперь это стало невозможно.

Бывшие привилегированные сословия пытались бежать от социалистических преобразований и вернуть собственность, используя для этого *Латвию*: «Восемнадцатый год прошел грядой преступлений, когда после национализации земли, фабрик, заводов и банков – купцы,

землевладельцы, фабриканты, дворяне – задним числом, датами от 1916-го, от 1915-го годов, подделывая нотариальные записи, продавали национализированные владения различным иностранцам, в том числе даже эстонцам и финнам, и тем, которые имели возможность патриироваться в Польше и Латвии» (Б. Пильняк «Рождение человека», 1934). Данные исторические события не являются основной темой рассказа, но используются для раскрытия образа героини, работающей прокурором и имеющей дело с уголовными преступлениями.

Липарские и *Гавайские* острова показаны как источник важного для промышленности селена и использованы во фразе, показывающей веру во всемирное распространение коммунизма (см. полужирное выделение): «Только с селеном дело плохо. Добывается он в небольшом количестве на Гавайских, Липарских островах. Туда не поскачешь. Не советские они **еще пока что**» (А. Р. Беляев «Заочный инженер», 1931).

Европа в целом фигурирует в негативных контекстах: советские люди культурнее европейских; у европейцев принята ненужная вычурность; эти люди готовы покупать у СССР всякий мусор, а именно – лягушек: «Ну и большевики!.. Уже добрались и до лягушек! За такое дерьмо будут получать золото!..» (Н. Н. Никандров «Зеленые лягушки», 1935). Одновременно *Европа* является символом прогресса и местом, где можно перенять опыт.

Атлантический океан – место путешествий персонажа в книге, читаемой персонажами одного из рассказов. При упоминании *Тихого океана* имеется в виду та его часть, что относится к Дальнему Востоку СССР.

Таким образом, западные страны представлены преимущественно как пространства «инакового» – более технологически развитого, но культурно и идеологически чуждого, – а также чаще всего фигурируют в ретроспективных нарративах, связанных с дореволюционным или раннесоветским временем.

2.3. Контексты употребления восточных топонимов

Топонимов, расположенных в восточной части земного шара и не входящих в число союзных республик, в исследовательском корпусе сравнительно мало.

Монголия, *Синьцзян* и *Япония* являются местом, куда бежали «белые» во время Гражданской войны. *Афганистан* упомянут как соседнее с Таджикистаном государство. *Мекка* – город, посещенный одним из персонажей («он видел Мекку», Л. Канторович «Конец Ибрагим-бека», 1936). В *Турцию* были выселены черкесы, также из этой страны персонажи привозили вещи. *Константинополь* – место временного пребывания персонажей, также упоминается как *Царьград* в историческом нарративе. *Африка (Тунис)* – место, где герой рассказа В. А. Герасимовой «Филош» (1932) проходил службу. Юноша был отправлен туда с расчетом на то, что европейцу будет тяжело находиться в непривычном климате, однако Филоша это не поколебало. *Иерусалим* упомянут не как реальная точка на географической карте, а как Небесный Город (В. А. Никифоров-Волгин «Дорожный посох», 1938).

Таким образом, восточные локации упоминаются эпизодически без идеологической нагрузки – как места, в которые направляются или в которых побывали персонажи, в ретроспективных и религиозных контекстах.

2.4. Контексты употребления топонимов союзных республик

Ряд топонимов располагается на территории нынешнего Казахстана, Таджикистана, Узбекистана, Туркменистана, Грузии. Действие одного рассказа разворачивается в Армении.

Рассказы, в которых упомянуты эти локации, не отличаются по смыслу от тех, которые повествуют о Москве и Петербурге. Территории союзных республик мыслятся как часть единой коммунистической страны, в каждом уголке которой ведется борьба за светлое будущее.

Единственная особенность, которую удалось выявить, – тексты, действие которых происходит в среднеазиатской части СССР, содержат больше описаний природы и природных топонимов (рек, гор, озер, пустынь и др.), что объясняется ландшафтом территорий.

Выделяется рассказ Эль-Регистана «Хайбер, сын пустыни» (1934). Его главный персонаж – собака Хайбер, в образе которой воплощены черты, приписываемые идеальному советскому гражданину. Пес идеально выполнял обязанности пастуха и был настолько хорош, что «многие колхозы мечтали заполучить Хайбера к себе». Тем самым показано, что даже животное нацелено на благо своего колхоза и страны в целом.

Рассказ Л. Первомайского «Джалил» (1934) чем-то перекликается с упомянутым выше «Филошем». Здесь также представлен образ идеального человека, но уже на территории союзной республики. Ему принадлежат слова «Не искушай меня: у меня есть партия и Таджикистан, а у тебя ничего не осталось в жизни...». Герой не будет спать и жалеть себя, но защитит поля от саранчи. Даже когда все говорят ему, что бороться с насекомыми бесполезно, герой не сдается: «...никто не сможет сказать, что Джалил Урунов отступил, испугавшись трудностей...».

Напротив, в рассказе Л. Канторовича «Конец Ибрагим-бека» изображен персонаж-басмач, служивший англичанам («Знал не хуже своих хозяев-англичан...») и занимающийся контрреволюционной деятельностью. Герой описывается следующими словами: «Ибрагим-бек, враг своей страны, человек без родины, человек, в душе которого остались только злоба и месть, не дешево продаст красным свою жизнь. Кровью и огнем рассчитается он за отнятое у него революцией богатство». Он предпринимает попытку численным превосходством (500 джигитов против 13 красноармейцев) разбить советских пограничников и терпит поражение. Это текст о мужестве и силе советских солдат, с одной стороны, и мелочности контрреволюционеров – с другой.

Таким образом, география Советского Союза широка и объединяет сильных, мужественных, честных людей, воодушевленных революционной идеей.

Заключение

Настоящее исследование позволило выявить характерные особенности географической представленности в русском рассказе 1930-х гг. Используя методы автоматической и экспертной аннотации, удалось выделить и проанализировать более 400 топонимов, значительная часть которых была точно локализована. Итоги картографирования и контекстуального анализа позволяют говорить о пространственной структуре мира русского рассказа 1930-х гг., в которой важное место занимает европейская часть Советского Союза, Южная Сибирь, а также союзные республики, представленные как равноправные участники общего проекта построения социалистического будущего. Западные страны, напротив, играют роль символического «другого» – технологически развитого, но морально чуждого пространства. Действие большинства рассказов происходит на территории СССР, страны вне его границ упомянуты как место временного пребывания или транзита персонажей, источник вещей и опыта, а также в ретроспективных контекстах. Вымышленные локации в текстах практически отсутствуют. Также было отмечено, что многие рассказы 1930-х гг. посвящены Гражданской войне в России.

Полученные результаты вносят вклад в изучение художественных и культурных особенностей малой прозы прошлого столетия, рассматриваемых сквозь призму репрезентации географического пространства в русском рассказе 1930-х гг., и могут быть использованы как в литературоведческих, так и в междисциплинарных исследованиях, включая культурную географию и цифровую гуманитаристику. В перспективы исследования входит сопоставительный анализ с текстами других периодов XX в., включая построение географических карт по текстам других десятилетий, с целью выявления динамики литературной географии.

Список литературы

- Гребенников А. О., Марусенко Н. М.** Корпус русского рассказа начала XX века. Пример лингвостатистического анализа // Компьютерная лингвистика и вычислительные онтологии. Вып. 4 (Тр. XXIII Международ. объединенной науч. конф. «Интернет и современное общество», IMS-2020, Санкт-Петербург, 17–20 июня 2020 г.: сб. науч. ст.). СПб.: Университет ИТМО. 2020. С. 21–28. DOI 10.17586/0000-0000-2020-4-21-28.
- Гребенников А. О., Скребцова Т. Г.** Языковая картина мира в русском рассказе начала XX века // Философия и гуманитарные науки в информационном обществе. 2019. № 3. С. 82–92.
- Колмогорова А. В.** Топонимы в советском песенном дискурсе (1964–1991): корпусное исследование // Вопросы ономастики. 2025. Т. 22, № 2. С. 148–182. DOI 10.15826/vopr_onom.2025.22.2.020.
- Лагутина Н. С., Васильев А. М., Зафиевский Д. Д.** Задачи в области распознавания именованных сущностей: технологии и инструменты // Моделирование и анализ информационных систем. 2023. Т. 30. № 1. С. 64–85. DOI 10.18255/1818-1015-2023-1-64-85.
- Мартыненко Г. Я., Шерстинова Т. Ю., Мельник А. Г., Попова Т. И.** Методологические проблемы создания Компьютерной антологии русского рассказа как языкового ресурса для исследования языка и стиля русской художественной прозы в эпоху революционных перемен (первой трети XX века) // Компьютерная лингвистика и вычислительные онтологии. Вып. 2 (Тр. XXI Междунар. объединенной конф. «Интернет и современное общество», IMS-2018, Санкт-Петербург, 30 мая – 2 июня 2018 г.: сб. науч. ст.). СПб.: Университет ИТМО, 2018а. С. 97–102.
- Мартыненко Г. Я., Шерстинова Т. Ю., Попова Т. И., Мельник А. Г., Замирайлова Е. В.** О принципах создания корпуса русского рассказа первой трети XX века // Тр. XV Междунар. конф. по компьютерной и когнитивной лингвистике «TEL 2018». Казань, 2018б. С. 180–197.
- Меховский В. А., Кижнер И. А.** Мир глазами образованного человека г. Минусинска конца XIX – начала XX века: распределение частотности географических названий в книгах Минусинской общественной библиотеки // Историческая информатика. 2025. № 1. С. 174–189. DOI: 10.7256/2585-7797.2025.1.72586
- Первый Всесоюзный съезд советских писателей, 1934: стенограф. отчет. М.: Худ. лит., 1934. 718 с.
- Скребцова Т. Г.** Динамика тем русских рассказов начала XX века // Философия и гуманитарные науки в информационном обществе. 2020. № 3. С. 45–60.
- Скребцова Т. Г., Гребенников А. О., Шерстинова Т. Ю.** Динамика лексического состава русской художественной прозы (на материале частотных словарей корпуса русских рассказов 1900–1930) // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной междунар. конф. «Диалог». М., 2021. С. 646–660. DOI 10.28995/2075-7182-2021-20-646-659.
- Субботин П. Ю.** Священномученик Никодим, епископ Белгородский: Жизнь, трагическая гибель, обретение мощей. Белгород: Изд. дом Белгородской митрополии, 2018. 135 с.
- Тынянов Ю. Н.** Архаисты и новаторы. Л.: Прибой. 1929. 595 с.
- Шерстинова Т. Ю.** Мир русского рассказа сквозь призму современных цифровых технологий // Социо- и психолингвистические исследования, 2022, вып. 10. С. 22–28.
- Шерстинова Т. Ю., Кирина М. А., Москвина А. Д.** Тематическое моделирование художественной прозы: оценка и интерпретируемость результатов (на примере русского рассказа 1900–1930) // Вестник Томского государственного университета. Филология. 2024. № 89. С. 127–151. DOI 10.17223/19986645/89/6.

- Шерстинова Т. Ю., Кирина М. А., Хлусова Я. К.** Корпус русского рассказа XX века: текущее состояние и перспективы развития // Тр. междунаrod. конф. «Корпусная лингвистика – 2025». СПб.: СПбГУ (в печати).
- Gemma Team.** Gemma 2: Improving Open Language Models at a Practical Size //arXiv preprint arXiv:2408.00118. 2024.
- Grebennikov, A. O., Skrebtsova, T. G.** Russian short stories: changes in word group frequencies over the last century. In: *Proceedings of the International Conference “Corpus Linguistics – 2023”*. St. Petersburg, Saint Petersburg University Press, 2024, pp. 260–267.
- Grishman R., Sundheim B.** Message Understanding Conference-6: A Brief History. In: *Proceedings of the 16th International Conference on Computational Linguistics (COLING 96)*, Copenhagen, August 1996, 1996, pp. 466–471.
- Sherstinova T., Moskvina A., Kirina M., Karysheva A., Kolpashchikova E., Maksimenko P., Seinova A., Rodionov R.** Sentiment Analysis of Literary Texts: A Study of Theme and Readers’ Preferences in Russian Short Stories from 1900–1930s. In: *Literature, Language and Computing: Russian Contribution from the LiLaC – 2023*. Springer, 2025, pp. 23–35. DOI 10.1007/978-981-96-0990-1_3.

References

- Gemma Team.** Gemma 2: Improving Open Language Models at a Practical Size. arXiv preprint arXiv:2408.00118. 2024.
- Grebennikov A. O., Marusenko N. M.** The Early XX-century Russian Short Stories Corpora. An Example of Lingvo-statistical analysis. In: *Computer Linguistics and Computing Ontologies*, Vol. 4 (Proceedings of the XXIII International Joint Scientific Conference «Internet and Modern Society», IMS-2020, St. Petersburg, June 17–20, 2020). St. Petersburg: ITMO University, 2020, pp. 21–28. DOI 10.17586/0000-0000-2020-4-21-28. (in Russ.)
- Grebennikov A. O., Skrebtsova T. G.** Russian Short Stories: Changes in Word Group Frequencies over the Last Century. In: *Proceedings of the International Conference “Corpus Linguistics – 2023”*. St. Petersburg, Saint Petersburg University Press, 2024, pp. 260–267.
- Grebennikov A. O., Skrebtsova T. G.** World through the Prism of the Early XX-Century Russian Short Stories. In: *Philosophy and Humanities in Information Society*, 2019, no. 3, pp. 82–92. (in Russ.)
- Grishman R., Sundheim B.** Message Understanding Conference-6: A Brief History. In: *Proceedings of the 16th International Conference on Computational Linguistics (COLING 96)*, Copenhagen, August 1996, 1996, pp. 466–471.
- Kolmogorova A. V.** Toponyms in Soviet Song Discourse (1964–1991): A Corpus-Based Study. In: *Problems of Onomastics*, 2025, vol. 22, no. 2, pp. 148–182. DOI 10.15826/vopr_onom.2025.22.2.020. (in Russ.)
- Lagutina N. S., Vasilev A. M., Zafievskiy D. D.** Name Entity Recognition Tasks: Technologies and Tools. In: *Modeling and Analysis of Information Systems*, 2023, vol. 30, no. 1, pp. 64–85. DOI 10.18255/1818-1015-2023-1-64-85. (in Russ.)
- Martynenko G. Ya., Sherstinova T. Yu., Melnik A. G., Popova T. I.** Methodological Issues Related with the Compilation of Digital Anthology of Russian Short Stories (the First Third of the 20th Century). In: *Computer Linguistics and Computing Ontologies*, Vol. 2 (Proceedings of the XXI International Joint Scientific Conference «Internet and Modern Society», IMS-2018, St. Petersburg, May 30–June 20, 2018). St. Petersburg: ITMO University, 2018, pp. 97–102. (in Russ.)
- Martynenko G. Ya., Sherstinova T. Yu., Popova T. I., Melnik A. G., Zamirajlova E. V.** On the Principles of Creation of the Russian Short Stories Corpus of the First Third of the 20th Century.

In: *Proceedings of the XV International Conference on Computational and Cognitive Linguistics "TEL 2018"*: in 2 vols. Kazan, 2018, vol. 1, no. 18, pp. 180–197. (in Russ.)

Mekhovskii V. A., Kizhner I. A. The World through the Eyes of an Educated Person in Minusinsk of the Late XIX–early XX Centuries: Distribution of the Frequency of Geographical Names in the Books of the Minusinsk Public Library. In: *Historical Informatics*, 2025, no. 1, pp. 174–189. DOI 10.7256/2585-7797.2025.1.72586. (in Russ.)

Sherstinova T. Yu. The World of Russian Short Stories from the Perspective of Modern Digital Technologies. In: *Socio Psycho Linguistic Research*, 2022, iss. 10, pp. 22–28. (in Russ.)

Sherstinova T. Yu., Kirina M. A., Moskvina A. D. Topic Modeling of Prose Fiction: Model Assessment and Interpretability (the Case of Russian Short Stories of the 1900s–1930s). In: *Tomsk State University Journal of Philology*, 2024, no. 89, pp. 127–151. DOI 10.17223/19986645/89/6. (in Russ.)

Sherstinova T. Yu., Kirina M. A., Khlusova Ya. K. The Corpus of Russian Short Stories of the 20th Century: Current State and Prospects for Development. In: *Proceedings of the International Conference "Corpus Linguistics – 2025"*. St. Petersburg, Saint Petersburg University Press (in print). (in Russ.)

Sherstinova T., Moskvina A., Kirina M., Karysheva A., Kolpashchikova E., Maksimenko P., Seinova A., Rodionov R. Sentiment Analysis of Literary Texts: A Study of Theme and Readers' Preferences in Russian Short Stories from 1900–1930s. In: *Literature, Language and Computing: Russian Contribution from the LiLaC – 2023*. Springer, 2025, pp. 23–35. DOI 10.1007/978-981-96-0990-1_3.

Skrebtsova T. G. Early 20 Century Russian Short Stories Themes and Their Dynamics of Change. In: *Philosophy and Humanities in Information Society*, 2020, no. 3, pp. 45–60. (in Russ.)

Skrebtsova T. G., Grebennikov A. O., Sherstinova T. Yu. The Dynamics of Vocabulary in Russian Prose (Based on Frequency Dictionaries of the Corpus of Russian Short Stories 1900–1930). In: *Computational Linguistics and Intellectual Technologies, Papers from the Annual International Conference "Dialogue"*. Moscow, 2021, pp. 646–660. DOI 10.28995/2075-7182-2021-20-646-659. (in Russ.)

Subbotin P. Yu. Hieromartyr Nikodim, Bishop of Belgorod: Life, Tragic Death, Discovery of Relics. Belgorod, Publishing House of the Belgorod Metropolis, 2018, 135 p. (in Russ.)

The First All-Union Congress of Soviet Writers, 1934: Stenographic Report. Moscow, Khudozhestvennaya literatura, 1934, 718 p. (in Russ.)

Tynyanov Yu. N. Archaists and Innovators. Leningrad, Priboy, 1929, 595 p. (in Russ.)

Информация об авторе

Хлусова Яна Константиновна, стажер-исследователь лаборатории языковой конвергенции Национального исследовательского университета «Высшая школа экономики», Санкт-Петербург

Information about the Author

Yana K. Khlusova, Intern Researcher, Laboratory of Language Convergence, National Research University «HSE University», St. Petersburg

Статья поступила в редакцию 18.08.2025;

одобрена после рецензирования 02.03.2026; принята к публикации 06.03.2026

The article was submitted 18.08.2025;

approved after reviewing 02.03.2026; accepted for publication 06.03.2026

Научная статья

УДК 81'27

DOI 10.25205/1818-7935-2026-24-2-124-135

Полевые исследования языка немцев-меннонитов Сибири: экспедиционный опыт из д. Неудачино и д. Аполлоновка

Алена Сергеевна Григорьева

Новосибирский государственный университет
Новосибирск, Россия

a.grigoreva7@g.nsu.ru, <https://orcid.org/0000-0001-8329-040X>

Аннотация

В статье представлены результаты комплексных полевых социолингвистических исследований языка *Plautdietsch*, проведенных в 2021 и 2024 гг. в деревнях Неудачино (Новосибирская область) и Аполлоновка (Омская область), являющихся местами компактного проживания сибирских немцев-меннонитов. Эмпирическую базу исследования составляют аудиозаписи спонтанной речи, полуструктурированные интервью, включенное наблюдение, материалы лингвистической документации, социолингвистические опросы и полевые дневники, полученные в ходе экспедиционной работы. На основе социолингвистического, контактно-лингвистического и этнолингвистического анализа реконструируется современная языковая ситуация и выявляются факторы, определяющие витальность, функциональное распределение, институциональную поддержку и межпоколенческую трансмиссию *Plautdietsch* в локальных сообществах. Показана историческая обусловленность формирования *Plautdietsch* как «голландизированного» нижненемецкого диалекта в Данцигском регионе и его последующая трансформация в условиях миграций в Российскую империю и СССР. Установлены различия в степени сохранности и функциональном статусе языка в двух исследуемых локациях: в Аполлоновке *Plautdietsch* сохраняет функции повседневной коммуникации, внутриобщинного взаимодействия и межпоколенческой передачи, тогда как в Неудачино фиксируются процессы функционального сужения, доминирование русского языка, возрастная стратификация языковой компетенции и ограничение коммуникативных доменов употребления идиома. Выявлены некоторые устойчивые лексические признаки голландского и немецкого субстратов (на материале базовой лексики), а также системные русские адстратные влияния, проявляющиеся на лексическом, синтаксическом и дискурсивном уровнях. Научная новизна исследования состоит в сопоставительном анализе двух локальных социолингвистических моделей функционирования *Plautdietsch* и в эмпирическом выявлении роли общины, религиозных практик и внутриобщинных коммуникативных норм как ключевых механизмов поддержания языковой витальности. Делается вывод о том, что *Plautdietsch*, несмотря на отсутствие письменной традиции и давление контактного языка, сохраняет статус центрального маркера этнокультурной идентичности меннонитов, а перспективы его сохранения напрямую связаны с устойчивостью общинных институтов, межпоколенческой передачей, стабильностью внутриобщинных коммуникативных практик, а также возрастающим интересом к *Plautdietsch* у молодежи.

Ключевые слова

Plautdietsch, немцы-меннониты Сибири, полевые исследования, социолингвистика, языковая витальность, языковые контакты

Благодарности

Автор выражает искреннюю благодарность кандидату филологических наук Е. А. Либерт за неоценимое содействие в организации полевых экспедиций; всей семье Шелленберг и З. П. Тевс (д. Неудачино) за гостеприимство и активное участие в исследовании; Е. А. Штеффен (д. Неудачино) за помощь в информировании населения и предоставленную возможность работать в Доме-музее им. А. Я. Штеффена; А. Я. Паулсу (д. Аполлоновка) за помощь в организации полевой работы, предоставление литературы и содействие в сборе лингвистического материала, а также его семье за теплый прием; И. А. Шелленберг и М. И. Эпп за помощь в расшифровке аудио-

© Григорьева А. С., 2026

записей и консультации. Особая признательность – всем жителям деревень Неудачино и Аполлоновка, принявшим участие в данном исследовании.

Для цитирования

Григорьева А. С. Полевые исследования языка немцев-меннонитов Сибири: экспедиционный опыт из д. Неудачино и д. Аполлоновка // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 124–135. DOI 10.25205/1818-7935-2026-24-2-124-135

Field Research on the Language of Siberian Mennonites: Expedition Experience from the Villages of Neudachino and Apollonovka

Alena S. Grigoryeva

Novosibirsk State University
Novosibirsk, Russian Federation

a.grigoreva7@g.nsu.ru, <https://orcid.org/0000-0001-8329-040X>

Abstract

This article presents the results of a comprehensive field-based sociolinguistic study of the *Plautdietsch* language, conducted in 2021 and 2024 in the villages of Neudachino (Novosibirsk Region) and Apollonovka (Omsk Region), areas of compact settlement of Siberian German Mennonites. The empirical base includes audio recordings of spontaneous speech, semi-structured interviews, participant observation, linguistic documentation, sociolinguistic surveys, and field diaries. Using sociolinguistic, contact-linguistic, and ethnolinguistic approaches, the study reconstructs the current linguistic situation and identifies the key factors shaping the vitality, functional distribution, and intergenerational transmission of *Plautdietsch* in these communities. The historical development of *Plautdietsch* as a “Dutchified” Low German dialect in the Danzig (Gdańsk) region, as well as its subsequent transformation through migration into the Russian Empire and the USSR, is outlined. A comparative analysis reveals pronounced differences between the two localities. In Apollonovka, *Plautdietsch* functions as a language of everyday communication, intra-community interaction, and stable intergenerational transmission. In Neudachino, processes of functional narrowing are observed, characterized by the dominance of Russian, age-stratified language competence, and the restriction of *Plautdietsch* to limited communicative domains. The analysis identifies persistent lexical features reflecting Dutch and German substrate influences in the core vocabulary, alongside systematic Russian adstratal influence at the lexical, syntactic, and discursive levels. The scientific novelty of the study lies in the comparative analysis of two local sociolinguistic models and in the empirical identification of community structures, religious practices, and internal communicative norms as key mechanisms sustaining language vitality. The study concludes that *Plautdietsch*, despite the absence of a written tradition and increasing pressure from contact languages, remains a central marker of Mennonite ethnocultural identity. Its future preservation is directly linked to the resilience of community institutions, stable intergenerational transmission, and a growing interest among younger generations.

Keywords

Plautdietsch, Siberian German Mennonites, field research, sociolinguistics, language vitality, language contact

Acknowledgements

The author wishes to express profound gratitude to E. A. Libert, PhD (Philology), for her invaluable support in the organization of the field expeditions. Sincere thanks are also extended to the Schellenberg family and to Z. P. Tevs (Neudachino) for their hospitality and active participation in the research. The author is grateful to E. A. Steffen (Neudachino) for assistance in community outreach and for granting access to the A. Ya. Steffen House-Museum as a research site. The author further acknowledges A. Ya. Pauls (Apollonovka) for his support in the organization of fieldwork, for providing relevant literature, and for his assistance in the collection of linguistic data, as well as his family for their generous hospitality. Special thanks are due to I. A. Schellenberg and M. I. Epp for their assistance with the transcription of audio materials and for their expert consultations. Finally, the author expresses deep appreciation to all residents of the villages of Neudachino and Apollonovka who participated in this study.

For citation

Grigoryeva A. S. Field research on the language of Siberian Mennonites: Expedition experience from the villages of Neudachino and Apollonovka. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 124–135. DOI 10.25205/1818-7935-2026-24-2-124-135

Введение

Изучение языков и диалектов этнических групп на территории России представляет собой важное направление современной лингвистики. Особую значимость в этом контексте приобретают полевые исследования языков Сибири, позволяющие не только фиксировать уникальный языковой материал, но и анализировать экстра- и социоллингвистические факторы, влияющие на их витальность. В последние годы такие комплексные исследования успешно проводятся на материале различных сибирских идиом, например, бурятского языка и его диалектов [Дырхеева и др., 2020], языков миноритарных народов, таких как кумандинский [Добрынина и др., 2024], барабинских и чатских татар [Рыжикова и др., 2024]. При этом язык сибирских немцев-меннонитов – *Plautdietsch*, несмотря на его уникальность и сложную историю формирования, до недавнего времени [Шиндрова и др., 2025] оставался на периферии подобных полевых исследований. Настоящая работа, основанная на экспедициях в д. Неудачино Новосибирской области (2021 г.) и д. Аполлоновка Омской области (2024 г.), направлена на комплексное описание современной языковой ситуации в двух локальных сообществах носителей *Plautdietsch*.

Наименование «меннониты» имеет религиозное происхождение. Так называли мирных крещенцев – последователей бывшего католического священника Менно Симонса, который стремился остановить жестокое, не соответствующее духу Евангелия бунтарское движение крещенцев, распространившееся в Западной Фрисландии (ныне Голландии), с помощью проповеди. В основе их верования лежали несогласие с милитаристскими режимами и стремление к мирному существованию, что остается характерной чертой и современных меннонитских обществ. Таким образом, многие меннониты-беженцы оказались в г. Данциг (ныне г. Гданьск, Польша).

Развитие диалектов в районе Данцига и его окрестностей, включая Вердер, Нерунг и Нидерунг, происходило под влиянием целого ряда исторических, социальных и лингвистических факторов. С XVI в. в эти районы начали прибывать голландские меннониты, получившие особые привилегии за свои навыки в мелиорации и освоении заболоченных земель. Первоначально меннониты были голландоязычными, однако к середине XVIII в. в быту перешли на нижненемецкие диалекты, сохраняя до середины XIX в. голландский язык в религиозной и письменной сферах. Важно подчеркнуть, что эти нижненемецкие диалекты уже содержали элементы голландского влияния, что облегчило языковой переход. Влияние голландского языка прослеживается в лексике и фонетике, формируя «голландизированный» вариант нижненемецкого диалекта. Языковая ситуация была осложнена и многоуровневым городским языковым ландшафтом Данцига: нижненемецкий использовался преимущественно ремесленниками и рабочими, тогда как верхненемецкий ассоциировался с образованными слоями населения. Как крупный торговый центр, город также стал важным узлом языкового обмена, что способствовало процессам стандартизации языка [Jähnig, Letkemann, 1985].

Лишь немногие могли легально пребывать на территории Польши – им было позволено оставаться на заболоченных территориях между Данцигом и Восточной Пруссией и заниматься освоением этих земель, развивая сельское хозяйство. Трудолюбие и умения меннонитов позволили превратить некогда неблагоприятные для земледелия территории в цветущий край. Однако в конце XVIII в. после первого раздела Польши места поселения меннонитов оказались под влиянием милитаристской Пруссии. В это же время меннониты были приглашены Екатериной II в Российскую империю для освоения территорий. Так меннонитами основана колония Хортица на правом берегу Днепра, а в начале XIX в. во время правления Александра I самая большая колония – Молочная. В Сибири меннониты оказались уже в конце XIX в. во времена массового переселения крестьян в Сибирь (1890–1912 гг.) [Эпп, 2011].

Соответственно, становление языковой и культурной специфики меннонитов неразрывно связано с их миграционной историей: от голландоязычных переселенцев в окрестностях

Данцига до носителей нижненемецкого диалекта, а затем – *Plautdietsch*. Таким образом, *Plautdietsch* демонстрирует редкий пример устойчивой языковой адаптации в условиях постоянных политических и социальных изменений. Сложный языковой ландшафт Данцига, влияние различных немецких и голландских диалектов, а также последующее переселение в Российскую империю создали основу для формирования уникальной языковой и этноконфессиональной общности, которая сохраняется и сегодня, несмотря на географическое и культурное рассеяние.

Морфология населенных пунктов и конфессиональные особенности

Особый уклад жизни меннонитов проявляется уже в пространственной организации деревень: Аполлоновка и Неудачино визуально отличаются от типичных российских поселений. Характерная планировка домов с единообразными изгородями и аккуратной застройкой формирует узнаваемый меннонитский архитектурный облик, позволяющий отличать дома местных жителей от построек неместного населения.

В Неудачино проживает около 400 человек, из которых около 25 % являются меннонитами. На территории около 1,5 км² находятся администрация, Дом культуры, школа (около 90 учащихся) и два молельных дома, исторически отделившихся друг друга и символично располагающихся на противоположных улицах. В деревне есть Дом-музей им. А. Я. Штеффена, знаменитого часового мастера. Авторские башенные часы А. Я. Штеффена можно увидеть в Санкт-Петербурге, Татарске и в Доме культуры д. Неудачно. В доме-музее собрана коллекция часов и различных старинных вещей А. Я. Штеффена, хранятся реликвии истории меннонитов и книги на *Plautdietsch*.

В д. Аполлоновка проживает примерно 970 человек, 70 % из которых – меннониты. Точных данных о площади деревни нет, однако она значительно крупнее Неудачино, состоящей из четырех улиц, и насчитывает 11 улиц. В Аполлоновке также имеется школа, рассчитанная на небольшое количество учеников. Работает фельдшерско-акушерский пункт, оказывающий базовую медицинскую помощь. Конечно же, центральное учреждение деревни – большой дом молитвы, вмещающий всех членов общины.

Религией современных меннонитов является баптизм – одно из направлений протестантского христианства, определяющее базовые ценностные установки, включая добровольное принятие жизненных испытаний и приоритет семьи. Х. Ангер отмечает, что уже в 1930-е гг. религиозные ограничения и притеснение веры становились основной причиной эмиграционных настроений среди меннонитов, подчеркивая их исключительную религиозную приверженность и пацифистские установки [Anger, 1930. S. 33]. Тяжелое положение российских меннонитов после Октябрьской революции, массовые депортации 1930-х гг., репрессии в годы Второй мировой войны и последующее расселение в Сибири и Казахстане подробно описаны Т. де Графом и Р. Нивебуром [de Graaf, Nieuweboer, 1995].

Культурная и духовная жизнь меннонитских деревень сосредоточена вокруг домов молитвы, где проводятся богослужения, коллективные чтения и обсуждения Библии, песнопения и общинные собрания. Исторически службы велись на *Plautdietsch* и немецком языке, однако в условиях расширения прихода они постепенно перешли на русский язык. В Неудачино сохраняется эпизодическое использование немецкого языка, а в Аполлоновке на *Plautdietsch* читаются наставления детям. В ходе экспедиции 2024 г. в д. Аполлоновка был зафиксирован и расшифрован один из таких поучительных рассказов (И. А. Шелленберг, 2002 г. р.; М. И. Эпп, 2004 г. р.; см. Прил. 1), представляющий собой нарратив с выраженной моральной доминантой, основанной на ценностях семьи, ответственности, послушания и заботы о ближних. Помимо этого, функционируют молодежные собрания, ориентированные на совместное обсуждение религиозных текстов, музыкальные практики и групповую коммуникацию в неформальной среде.

Как в Неудачино, так и в Аполлоновке гости деревень и молельных домов принимаются с радушием, без строгого требования соблюдения внутриобщинных норм: приобщение к вере и принятию правил общины носит добровольный характер. Отличительной чертой Аполлоновки является наличие на территории деревни небольшой православной церкви. Жители отмечают, что сосуществование двух конфессий не влияет на межличностные отношения и основывается на взаимном уважении.

Немаловажно отметить наблюдения о высокой степени развития методов ведения домашнего хозяйства: «...меннониты традиционно выделялись среди немцев высоким уровнем развития экономики и культуры, использованием наиболее современной техники и приемов ведения земледелия и животноводства» [Савин, 2009, с. 3].

Одной из особенностей Аполлоновки является развитый предпринимательский сельскохозяйственный сектор: в деревне функционируют кооперативы по переработке сельхозпродукции, зернохранилище, мельница и комбикормовый завод. Особую роль в экономике деревни играет ООО «Виллок Фарм» – современное агропредприятие, активно сотрудничающее с хозяйствами региона. Предприятие участвует в развитии инфраструктуры деревни, обеспечивая рабочие места и поддержку местных инициатив. Компания берет начало в Канаде, где существует одноименное хозяйство, принадлежащее Валтеру Вилламсу; в результате его интереса к сибирским территориям в 2002 г. было запущено финансирование проекта, а в 2005 г. началась активная деятельность предприятия. В настоящее время завод успешно функционирует, не уступая по качеству ведущим аграрным предприятиям, и больше не является собственностью канадских владельцев. Сотрудниками предприятий преимущественно являются меннониты, что связано с их трудовой этикой, сельскохозяйственными навыками и социальной надежностью. В д. Неудачино разрабатывался проект сыроварни, однако не был реализован.

Бережное отношение меннонитов к благоустройству деревень отмечал еще Х. Ангер [Anger, 1930]. По его наблюдениям, меннонитские поселения в Сибири отличались высокой степенью типологической схожести и легко распознавались в ландшафте благодаря планировочной структуре: населенные пункты меннонитов отличались малочисленностью и линейной планировкой с одной основной дорогой [Там же]. Ангер также указывает на более высокий уровень хозяйственного устройства по сравнению с лютеранскими, католическими немецкими и русскими деревнями: дома были, как правило, крупнее и лучше построены, жилые и хозяйственные помещения часто объединялись под одной крышей, развиты были садоводство и огородничество. Несмотря на интерес к Германии, меннониты, по наблюдениям автора, сохраняли прежде всего религиозную и культурную идентичность, воспринимаясь окружающим населением как немцы. Важно, на наш взгляд, отметить открытость, любовь к жизни и окружающему миру местных жителей. Новый человек в закрытом сообществе меннонитов воспринимается не как чужак, а как гость. Главное правило – взаимоуважение и уважение личных границ другого человека.

Самоидентификация и язык

Самоидентификация меннонитов Аполлоновки и Неудачино – немцы. Однако местные жители говорят, что несмотря на то, что считают себя этническими немцами и относятся к немцам на территории России, в Германии они – русские. Также российские меннониты являются желанными работниками в Германии, поскольку обладают трудо- и жизнелюбием. На наш взгляд, меннонитов Сибири можно описать как носителей немецкой этнической идентичности, сформированной в российском социокультурном контексте.

Особенность и уникальность своего языка, *Plautdietsch*, осознают и сами меннониты. Ввиду отсутствия единой письменности, язык передается в устной форме от поколения к поколению. Удивительным, с нашей точки зрения, выступает факт единства понимания меннонитов,

несмотря на некоторые различия на лексическом и фонетическом уровнях в семьях в пределах одного сообщества (в нашем случае – д. Неудачино и д. Аполлоновка).

Меннониты обладают высокой степенью другоцентричности: появление новых прихожан в молельных домах привело к угасанию *Plautdietsch* – приезжие не прилагали усилий к изучению языка местных жителей, доступность церковных служб стала возможной лишь на русском языке. В послевоенное время в связи с приездом в д. Неудачино учителей, не владеющих *Plautdietsch*, общение учащихся на языке стало также невозможным. Негласный «запрет» на использование *Plautdietsch* привел к упадку престижа языка среди молодежи. До недавнего времени молодежь избегала общения на языке – использование *Plautdietsch* считалось непопулярным. Только в последние годы молодые меннониты, осознавая важность сохранения своей идентичности и передачи ее новым поколениям, начали проявлять интерес к овладению, сохранению и распространению *Plautdietsch*. Небезынтересна языковая особенность молодежи меннонитов: говоря как о прошлом, так и о будущем, самих лексем «прошлое» и «будущее» молодые девушки и юноши не знают, хотя данные лексические единицы в языке существуют. С накоплением жизненного опыта понимаются уникальность языка и единственная возможность передачи языка новым поколениям через устную речь. Также приходит осознание, что *Plautdietsch* – важнейшая часть меннонитского самосознания. От этого язык ценится еще больше. Неудачинцы даже замечают, что между людьми, говорящими на *Plautdietsch*, существуют особая связь.

Языковая ситуация в Аполлоновке отличается более высоким уровнем функционирования языка. В первую очередь это обусловливается тем, что большая часть жителей – этнические немцы-меннониты. Кроме того, язык повседневности Аполлоновки – *Plautdietsch*, а дети дошкольного возраста практически не владеют русским. Также старшие члены общины регулярно проводят занятия с детьми на языке.

Сам *Plautdietsch* действительно уникален: он обладает своей структурой, словарным составом и правилами. Однако сопоставление лексических единиц *Plautdietsch*, выявленных в ходе сбора языкового материала для исследования и общения с носителями, с голландскими и немецкими аналогами показывает, что язык меннонитов сохранил заметные следы своего исторического происхождения. Устойчивое сходство базовой лексики с голландским подтверждает, что *Plautdietsch* развивался на пересечении двух языковых систем – голландской и нижненемецкой:

- *фраувэ* ‘вдова’: гол. *weduwe*, нем. *Witwe*, нем. *Frau* ‘женщина’;
- *юа* ‘год’: гол. *jaar*, нем. *Jahr*;
- *ольт* ‘старый’: гол. *oud*, нем. *alt*;
- *лу:ть* ‘поздно’: гол. *laat*;
- *енух* ‘достаточно’: гол. *genoeg*, нем. *genug* – в данном случае сходство особенно очевидно при обращении к фонетической транскрипции голландской лексемы *genoeg* – [ʏə'nuχ].

При этом для носителей *Plautdietsch* языковое родство неочевидно – не все меннониты владеют литературным немецким языком, практически никто не знает голландского языка. Фигурируют в *Plautdietsch* и самостоятельные лексические единицы, так, наименование напитка из зерен пшеницы, ячменя и овса, ранее использовавшегося в качестве кофе, сегодня заменяет саму лексему «кофе» – *препс*, *шварцпрепс* ‘черный кофе’.

Несмотря на практическую идентичность *Plautdietsch* в Аполлоновке и Неудачино, в процессе записи материала для психолингвистического исследования с жителями Аполлоновки выявлены некоторые расхождения в лексическом составе. Так, жители Неудачино используют лексему *шлу:т* в значении ‘поздно’, в то время как для жителей Аполлоновки это ‘конец’. И даже есть поговорка *ди й’эшихът эс йт’мит нэ гольднэ шлу:т* ‘история окончилась золотым концом’.

В речи меннонитов прослеживается и контактность русского языка, например, меннониты говорят *нэ баночка* ‘(одна) баночка’, заимствовав слово из русского языка и добавив к нему

артикуль, для кастрюли существует лексема *кастрол*; практически без изменений, за исключением фонетических особенностей, заимствованы такие слова, как *аллергия* (*аллергия онэ би:нэ* ‘аллергия на пчел’), *багажник*, *документ*, *магазин*, *трактор* и др. Безусловно, нельзя не отметить и некоторые грамматические особенности, нетипичные для русского языка: как в немецком языке, так и в *Plautdietsch* можно часто услышать союз *oder* ‘или’ в конце альтернативного вопроса: *du fährst heute mit dem Traktor aufs Feld, oder?* ‘ты сегодня поедешь на поле на тракторе, да?’. Эта особенность оказывает влияние и на синтаксис русского предложения: Мы сегодня должны закончить раньше управляться¹, или?

Сегодня тенденцию к преобладанию русского языка с сожалением отмечают в обеих деревнях, хотя степень его влияния в Аполлоновке ниже, чем в Неудачино, что подтверждается большим процентом жителей-меннонитов, владением *Plautdietsch* детьми, использованием языка на собраниях в молельных домах.

Заключение

История немцев-меннонитов насыщена переселениями, гонениями и миграциями, что отражает и их язык, *Plautdietsch*: отсутствие единой сформировавшейся письменной традиции и неопределенный языковой статус – самые очевидные атрибуты языка, коррелирующие с историей его носителей.

Опыт, полученный в экспедициях, показал, что этническая, религиозная и языковая идентичность меннонитов в Сибири сохраняется, несмотря на исторические трансформации, миграции и влияние окружающей культуры. *Plautdietsch*, вопреки отсутствию письменной нормы, остается важнейшим элементом культурного наследия и средством внутригрупповой коммуникации. Сравнительный анализ языковой ситуации в двух деревнях позволяет утверждать, что степень влияния русского языка значительно выше в Неудачино: в молельных домах *Plautdietsch* используется лишь эпизодически, а в быту – в основном людьми старшего поколения. В Аполлоновке же *Plautdietsch* – это язык повседневной жизни: им владеют дети, на нем проводятся занятия в общине.

Историческая реконструкция происхождения *Plautdietsch* через нижненемецкие и голландские корни подтверждается как лексическими параллелями, так и фонетическими особенностями. Это свидетельствует о глубоком историческом пласте, заложенном еще в эпоху заселения Данцигского региона.

Религиозная жизнь остается ключевым механизмом поддержания общинной идентичности. Богослужения, молодежные собрания, наставления и совместные песнопения – не только сплавляют членов общины, но и способствуют сохранению языка. Особенно значимо то, что интерес к *Plautdietsch* среди молодежи возрождается.

Таким образом, *Plautdietsch* в условиях сибирских деревень остается живым языком, отражающим сложную историю меннонитов: от голландских переселенцев Данцигской низины до современных общин в сельской Сибири. Язык становится зеркалом исторической памяти и маркером этнокультурной идентичности. Его сохранение возможно лишь при поддержании жизнеспособности общины, межпоколенческой преемственности и уважении к своим традициям.

¹ «управляться» – заниматься хозяйством.

Приложение

**Рассказ-наставление детям из д. Аполлоновка
(Текст на *Plautdietsch* записан носителями языка²)**

– дуф эть эм вольт гун? Эть Вэл ипцир гун.

Тьэмт Лука ну мама би йрант эн фрахьт би де. Дой ходэ эм дарп эн хусь йекофт эн вирзэ нхь лэнст ивайтрокэ. Эн мама тьид оп Ам зо фрнтлхь эн дохт футс он Дой йорэ ос зой клинь Ви. де йфол дот ок Зоя оп виз гун, блэмэ плэтьэ ок эм вольт ипцир эн хе тэмт ок дэ нейнйоше Лорна би эн зайт, эть Вэл ок эм вольт гун, эн дан тэмт нох дот фифйоша брожате Лейстер би эн зайх эть Вэл ок мт дин мэт гун.

Дан захьт мама:

– на дан мот йи би папа гун фрогхе: Папа войт бита, дуф йи гун ода нхь. Эн ранд д Лука ну папа, папа дойт Буте шофэ, эн дат тэмта онйрант, йо, папа хафт онст эрлобт эм вольт гун. На йо, ну гунз эн тьтьзэ, Ду эс н гротэ виз, эн до восэ блэмэ, эн захьт Лорна:

– вот ДТ эс нэ шойнэ виз, хе кон Ман ранэ, эн дан плэтьд зэ блэмэ шмок, эн дан гунз эм вольт эн доньз до зхь доа нох олэс бтьте, оп оймул зайхьт дэ Лейстер:

– тьть мул, тьть мул до опс бом эс эн айхернхен (дост н белка, й) хупст эн Ду эс ойнэ белка, эн дан гунз вида эн дан тьтьз зо, рунт ом бтьтьз зхь олс эн оп оймул Ви д Лука нхь моя ту зойн.

Эн Лука Ви тин йо ёльт, дот Ви д элста, эн дан оп оймул фн эрнт Ву хирз д штэм:

– кумт мул хе, кумт мул хе, эть хаб вот йфонг захьт Лорна:

– ву бэст Ду, Ви зойн ди нхь.

– кумт хе, кумт хе – ропта дан

На йо, гунз ну Дой ЗиД, фэн Ву д штэм ком, эн тьтьз эс эн грота лох, эн Дой эс мт бэша Ол гонс бэвосэ, эн Ду эн мэд штэт дан д Лука, эн ду эс дан нох вот. эн гунз Ду бид бэша фби эн тьтьз штэт ду н оле оле мишень. Эн захьт д Лука:

– тьйт хе мул, хе кон Ман ок дм дак уп мок. Эн мокта дм дак уп, эн зэнд Ду твой зэт, оль твой зэт бэнэ, эн мокта д Дэ уп, эн захьта:

– зат йуньт нэн

– на, захьт Лорна, дуф ви онс эн дз мишень нэн зат?

– эть войт нхь. Эс дот вот йферлхет эн зоне оль мишень зхь нэн зат?

– эть войт нхь – захьт Лорна – Ви мот дохвол ишт би папа фрогхе. Эн тьйт з зо н хэхьт, д зон Дой Ви зо хух, эн захьт з:

– о, ви мот о л ту мэдах нэ хусь гунэ.

– на йо, эн Лука Дой вул нхь вахь гун фн хе, дот ви ам дох зо энтросонт н мишень фнье эм вольт.

Эн нэмт Лорна дм Клестер он хонт эн гунз н хусь, эн муст Лука ок хенярон гунэ. Эн дан, ван з хаб мэдахт ййит, эн дан захьт д Клестер:

– папа, Ви хаб эм вольт н мишень йфонге н оль мишень, эн дан захьта:

– на вот фн мишень, эть хаб ношт нхь йзойн.

– йо, зальт д Лука – фанкта он ту фталэ – дот эс зойн мишень, ду кон ман дм дак упмок, ду кон ман бэнэ зэтэ. Эн захьт папа:

– на эть мот ми дот ишт зэлст бэтьте.

На йо, д ондр знвнт вулз вада гун, эн фрахьт д Лука:

– папа, кон эть эм вольт гун ипцир?

Эн папа тьйт ам зо штрэнь он эн захьта:

– уба нуд мишень дуфст ду нхь би гун

² Информантами, сделавшими расшифровку и перевод текста, выступили женщины 2002 и 2004 г. р., проживающие в д. Неудачино и д. Аполлоновка. Обе женщины имеют / получают высшее образование.

собирать цветы и гулять в лесу. Тут подошла девятилетняя Лорна и говорит: «Я тоже хочу пойти в лес». И тут еще подходит пятилетний братишка Лестер: «Я тоже хочу пойти с ними».

– Тогда спросите у папы, он лучше знает, можно вам идти или нет, – сказала мама.

Папа работал во дворе. Лука подбежал, попросил разрешения, и папа ответил:

– Да, идите.

Дети обрадовались и пошли. Луг был широкий и красивый: можно бегать, цветы пахнут, птицы поют. Увидели белку – крикнули, побежали за ней, но она ловко скрылась.

Вдруг послышался голос:

– Идите сюда, идите сюда, я кое-что нашел!

– Где ты? Мы тебя не видим! – отозвалась Лорна.

Голос звал снова, и дети пошли на звук. За кустами оказалась большая яма, а рядом старая машина. Лука сказал:

– Смотрите, сверху крышка, ее можно открыть.

Они приподняли крышку и заглянули внутрь: два старых сиденья. Тогда Лука открывает дверь и говорит: «Садитесь». Лорна говорит: «А нам можно сюда садиться?» – «Не знаю», – говорит Лука. «Разве это опасно – садиться в такую старую машину?» – «Не знаю, может, спросить лучше у папы», – говорит Лорна и посмотрела вверх, а солнце было уже высоко. «Нам надо уже идти домой на обед», – говорит она.

Луке уходить не хотелось – это так интересно: прямо в лесу найти машину! Но Лорна взяла Клестера за руку, и Луке пришлось идти за ними.

– Папа, мы нашли в лесу машину, старую, – говорит Клестер за обедом.

– Какую еще машину? Я такой не видел, – удивился он.

– Настоящую, – подтвердил Лука. – Крышку можно открыть, сиденья видно.

– Сами туда не лезьте и ничего не открывайте, – строго сказал папа.

На следующий вечер Лука снова спросил:

– Папа, можно нам прогуляться в лес?

– Можно, – ответил папа, – но к машине не подходить.

– Понял, – кивнул Лука.

Клестер услышал, что Лука идет, и сказал:

– Я тоже хочу.

Лорна в этот раз осталась помогать маме, и поэтому она не пошла. Они шли по лесу, Клестер собирал для мамы цветы, увидев жучка, смотрел, что он делает.

Шли, играли по дороге, Лука все хотел пойти туда, в сторону машины, а Лестер говорит: «Папа же запретил подходить к машине». – «Ну, мы можем же просто пойти посмотреть». Дошли до той ямы, снова увидели старую машину.

– Давай откроем багажник, – предложил Лука. – Вдруг там клад?

Руками открыть не получилось. – Нужна палка, – сказал он.

Нашли палку, просунули в щель – багажник резко подскочил, и тут же раздалось громкое «ж-ж-ж-ж»: изнутри вырвался рой пчел. Дети в ужасе бросились бежать. Лука неся что есть сил, а Клестер вдруг сел на землю, закрыл лицо руками и заплакал, а Лука понял, что его ужалила пчела в щеку.

Вы же знаете, что пчелы не кусают, а жалят. У них есть жало, которым они жалят, и это очень больно, потому что в жале есть пчелиный яд, и от него все начинает распухать. Щека Лестера стала быстро пухнуть. Лестер заплакал; бежать он уже не мог, и Лука посадил его себе на спину, и так они добежали до дома.

– У Клестера аллергия на пчелиный яд, – встревожилась мама. – Потому и распухло так сильно.

Семья срочно поехала к врачу. Папа с мамой зашли с Лестером к врачу, а Лука сел возле двери на стуле, думал и корил себя: «А что, если он умрет и умрет из-за меня?» Лорна стояла у окна и ждала.

Наконец родители вернулись с Клестером. Ему велели лежать спокойно – теперь все будет хорошо. Он не умер, опасность миновала.

Дома папа строго сказал Луке:

– Ты понимаешь, ты сделал плохо, и из-за этого Клестер пострадал.

Лука все понимал, и ему было жалко Клестера.

– Даже если бы его пчелы не тронули, все равно это было неправильно. Ты старший, и младшие берут пример со старших. Если ты поступаешь плохо, и они будут делать то же самое. Это неправильно.

Лука попросил прощения у Клестера и у папы и подумал: «И что я в этой жизни сделал еще что-нибудь плохое?»

Поэтому, дети, давайте всегда думать о том, что когда мы плохо себя ведем, это плохо не только для нас, но и для других людей тоже, мы являемся для них плохим примером. Давайте всегда быть хорошим примером.

Список литературы

- Добрынина А. А., Шиндрова К. В., Григорьева А. С., Шестера Е. А.** Интонация диалога на кумандинском языке (результаты полевых исследований) // Томский журнал лингвистических и антропологических исследований. 2024. № 4 (46). С. 47–57. DOI: 10.23951/2307-6119-2024-4-47-57.
- Дырхеева Г. А.** Бурятский язык в условиях двуязычия: особенности трансформации языкового сознания (по результатам ассоциативного эксперимента). Улан-Удэ: Изд-во Бурятского научного центра СО РАН, 2020. 236 с.
- Рыжикова Т. Р., Шиндрова К. В., Плотников И. М., Якимец Н. В.** Интонационные и прагматические характеристики диалогической речи барабинцев и чатов // Языки и фольклор коренных народов Сибири. 2024. № 4 (52). С. 98–113. DOI: 10.25205/2312-6337-2024-4-98-113.
- Савин А. И.** Этноконфессия в советском государстве. Меннониты Сибири в 1920–1930-е годы: эмиграция и репрессии. Новосибирск: Посох, 2009. 752 с.
- Шиндрова К. В., Григорьева А. С., Рыжикова Т. Р., Добрынина А. А.** Интонационные паттерны диалогической речи немцев-меннонитов (опыт первого экспериментального просодического исследования) // Журнал Сибирского федерального университета. Серия: Гуманитарные науки. 2025. Т. 18, № 10. С. 1932–1945. EDN: HERFXZ.
- Эпп П.** Не исчезли по милости Господа. Steinhagen – Исилюль: Samenkorn, 2011. 896 с.
- Anger H.** Die Deutschen in Sibirien. Reise durch die deutschen Dörfer Westsibiriens. Berlin: Ost-Europa-Verlag, 1930. 102 S.
- De Graaf T., Nieuweboer R.** The language of the West Siberian Mennonites // Proceedings of the XIIIth International Congress of Phonetic Sciences: ICPHS 95. Stockholm, Sweden, 13–19 August 1995. P. 180–183.
- Jähnig B., Letkemann P.** Die Mundarten Danzigs und seines Umlandes // Danzig in acht Jahrhunderten / Hrsg. von B. Jähnig, P. Letkemann. Nicolaus-Copernicus-Verlag, 1985. 384 S.

References

- Dobrynina A. A., Shindrova K. V., Grigoreva A. S., Shestera E. A.** Intonatsiya dialoga na kumandinskoy yazyke (rezul'taty polevykh issledovaniy) [Intonation of dialogue in the Kumandin language (results of field studies)]. *Tomskii zhurnal lingvisticheskikh i antropologicheskikh issledovaniy* [Tomsk Journal of Linguistics and Anthropology], 2024, no. 4(46), pp. 47–57. DOI: 10.23951/2307-6119-2024-4-47-57. (in Russ.)

- Dyrkheeva G. A.** Buryatskii yazyk v usloviyakh dvuyazychiya: osobennosti transformatsii yazykovogo soznaniya (po rezul'tatam assotsiativnogo eksperimenta) [The Buryat language in a bilingual environment: features of the transformation of linguistic consciousness (based on the results of an associative experiment)]. Ulan-Ude, Buryat Scientific Center of SB RAS Publ., 2020. 236 p. (in Russ.)
- Ryzhikova T. R., Shindrova K. V., Plotnikov I. M., Yakimets N. V.** Intonatsionnye i pragmaticheskie kharakteristiki dialogicheskoi rechi barabintsev i chatov [Intonational and pragmatic characteristics of the dialogic speech of Baraba and Chat Tatars]. *Yazyki i fol'klor korennykh narodov Sibiri* [Languages and folklore of the indigenous peoples of Siberia], 2024, no. 4(52), pp. 98–113. DOI: 10.25205/2312-6337-2024-4-98-113. (in Russ.)
- Savin A. I.** Etnokonfessiya v sovetskom gosudarstve. Mennonity Sibiri v 1920–1930-e gody: emigratsiya i repressii [Ethnoconfession in the Soviet state. Siberian Mennonites in 1920–1930s: emigration and repressions]. Novosibirsk, Posokh Publ., 2009. 752 p. (in Russ.)
- Shindrova K. V., Grigoreva A. S., Ryzhikova T. R., Dobrynina A. A.** Intonatsionnye patterny dialogicheskoi rechi nemtsev-mennonitov (opyt pervogo eksperimental'nogo prosodicheskogo issledovaniya) [Intonation patterns of the dialogic speech of German Mennonites (experience of the first experimental prosodic study)]. *Zhurnal Sibirskogo federal'nogo universiteta. Seriya: Gumanitarnye nauki* [Journal of Siberian Federal University. Series: Humanities], 2025, vol. 18, no. 10, pp. 1932–1945. EDN: HERFXZ. (in Russ.)
- Epp P.** Ne ischezli po milosti Gospoda [Not perished by the grace of the Lord]. Steinhagen–Isilkul, Samenkorn Publ., 2011. 896 p. (in Russ.)
- Anger H.** Die Deutschen in Sibirien. Reise durch die deutschen Dörfer Westsibiriens. Berlin, Ost-Europa-Verlag, 1930. 102 p.
- De Graaf T., Nieuweboer R.** The language of the West Siberian Mennonites. In: Proceedings of the XIIIth International Congress of Phonetic Sciences: ICPhS 95. Stockholm, Sweden, 13–19 August 1995, pp. 180–183.
- Jähnig B., Letkemann P.** Die Mundarten Danzigs und seines Umlandes. In: Jähnig B., Letkemann P. (eds.). Danzig in acht Jahrhunderten. Nicolaus-Copernicus-Verlag, 1985. 384 p.

Информация об авторе

Григорьева Алена Сергеевна, магистр филологии, аспирант Гуманитарного института Новосибирского государственного университета

Information about the Author

Alena S. Grigoryeva, Master of Philology, PhD Student, Institute for Humanities of Novosibirsk State University

Статья поступила в редакцию 04.02.2026;
одобрена после рецензирования 27.03.2026; принята к публикации 03.04.2026

The article was submitted 04.02.2026;
approved after reviewing 27.03.2026; accepted for publication 03.04.2026

Научная статья

УДК 316.772; 316.772.2

DOI 10.25205/1818-7935-2026-24-2-136-145

Утерянное искусство общения: проблемы межличностной коммуникации в цифровом обществе

Ника Олеговна Кривопалова

Новосибирский государственный университет
Новосибирск, Россия

nika.krivopalova@gmail.com, <https://orcid.org/0009-0002-4664-8614>

Аннотация

В статье анализируются трансформации межличностной коммуникации в условиях цифровизации и распространения технологий искусственного интеллекта. Цель исследования – систематизировать современные знания о типах коммуникации, выявить факторы ее трансформации и определить роль ИИ как когнитивного посредника. Методологической основой послужил критический обзор классических и современных монографий по коммуникации, семиотике и медиаэкологии. Показано, что цифровая среда усиливает фатическое общение и визуальную репрезентацию, что при снижении доли очных контактов ведет к ослаблению межличностной чувствительности и росту социальной изоляции. Выявлен новый эффект – когнитивно-агентная медиация, при которой ИИ участвует в создании, редактировании и интерпретации сообщений, а не просто выступает каналом передачи. Определены позитивные последствия использования ИИ (доступность, инклюзия, эмоциональная поддержка) и риски (снижение эмпатии, антропоморфизация, этические проблемы). Новизна исследования заключается в предложении многоосевой типологии коммуникаций, учитывающей влияние ИИ, и в выявлении механизма когнитивной разгрузки пользователя. Результаты позволяют утверждать, что устойчивость искусства общения возможна только при балансе между использованием цифровых инструментов и развитием личностных коммуникативных компетенций.

Ключевые слова

коммуникация, фатическое общение, искусственный интеллект, эмодзи, визуальная культура, номофобия, цифровое общество

Благодарности

Выражаю искреннюю благодарность научному руководителю, редакционной коллегии журнала и рецензентам за внимательное отношение к моей работе, ценные комментарии и конструктивные предложения, способствовавшие улучшению качества статьи. Исследование выполнено без внешнего финансирования.

Для цитирования

Кривопалова Н. О. Утерянное искусство общения: проблемы межличностной коммуникации в цифровом обществе // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 136–145. DOI 10.25205/1818-7935-2026-24-2-136-145

© Кривопалова Н. О., 2026

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2
Vestnik NSU. Series: Linguistics and Intercultural Communication, 2026, vol. 24, no. 2

The Lost Art of Communication: Interpersonal Communication Challenges in the Digital Society

Nika O. Krivopalova

Novosibirsk State University
Novosibirsk, Russian Federation

nika.krivopalova@gmail.com, <https://orcid.org/0009-0002-4664-8614>

Abstract

This article analyzes the transformations of interpersonal communication in the context of digitalization and the spread of artificial intelligence (AI) technologies. The study aims to systematize current knowledge about communication types, identify factors driving its transformation, and define the role of AI as a cognitive mediator. The methodological foundation is a critical review of classical and contemporary monographs on communication, semiotics, and media ecology. It is shown that the digital environment enhances phatic communication and visual representation, which, combined with a decreasing share of face-to-face interactions, leads to diminished interpersonal sensitivity and increased social isolation. A new phenomenon is identified—cognitive-agential mediation—where AI actively participates in the creation, editing, and interpretation of messages, rather than merely serving as a transmission channel. The paper outlines the positive impacts of AI use (accessibility, inclusivity, emotional support) as well as risks (reduced empathy, anthropomorphism, ethical concerns). The novelty of the research lies in proposing a multi-axis typology of communication that accounts for AI's influence and in identifying the mechanism of user cognitive offloading. The findings suggest that the sustainability of the art of communication is possible only through a balance between the use of digital tools and the development of personal communicative competencies.

Keywords

communication, phatic communication, artificial intelligence, emojis, visual culture, nomophobia, digital society

Acknowledgments

I sincerely thank my academic advisor, the journal's editorial board, and the reviewers for their careful attention to my work, valuable comments, and constructive suggestions that helped improve the quality of the article. This research was conducted without external funding.

For citation

Krivopalova N. O. The lost art of communication: interpersonal communication challenges in the digital society. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 136–145. DOI 10.25205/1818-7935-2026-24-2-136-145

Введение

Коммуникация – центральный элемент человеческого существования, позволяющий устанавливать социальные связи, передавать знания и эмоции. Современное общество переживает кардинальные изменения в сфере коммуникации. Традиционные формы общения, сформировавшиеся тысячелетиями, подвергаются серьезной трансформации под воздействием цифровых технологий. Искусство живого общения, включающее в себя не только вербальные, но и невербальные компоненты коммуникации, постепенно утрачивает свои позиции в пользу опосредованных форм взаимодействия.

Актуальность данной статьи обусловлена необходимостью осмысления происходящих изменений в коммуникативной сфере и их влияния на психологическое состояние человека и общества в целом. Стремительное развитие технологий искусственного интеллекта создает новые вызовы для традиционного понимания межличностного общения и требует переосмысления роли человеческого фактора в коммуникативных процессах.

Цель статьи – систематизировать современные знания о типах и видах коммуникации, определить факторы ее трансформации и влияние ИИ на изменение форм общения.

Основная часть

В литературе дается различное толкование понятия «коммуникация». Возьмем за основу определение автора термина «коммуникация» Чарльза Кули: «Коммуникация – это механизм, посредством которого обеспечивается существование и развитие человеческих отношений, включающий в себя все мыслительные символы, средства их передачи в пространстве и сохранения во времени» [Гуляев, 2008. С. 275].

Исторически коммуникацию описывали в двух традициях: процессуально-инженерной (передача сообщений от источника к получателю через канал при наличии шума: шумовая модель К. Шеннона – У. Уивера, 1949; модель коммуникации «ИСКП: источник – сообщение – канал – получатель» (SMCR: Sender-Message-Channel-Receiver) Д. Берло, 1960; модель У. Шрамма, 1954) и семиотико-социокультурной (производство и циркуляция смыслов в социальных контекстах: Р. Якобсон, 1960; У. Эко, 1976; Р. Барт, 1957; Дж. Кэри, 1989). Границы между ними проницаемы: «код» и «шум» используются в обеих перспективах (модель Шеннона – Уивера часто называют «кодовой»), а инженерные схемы расширяются социокультурными параметрами (см. Д. Берло, развитие модели Шеннона – Уивера). В этой статье мы не противопоставляем традиции и не разводим терминологию; зафиксируем набор параметров анализа: канал и шум; код и функции сообщения; свойства медиасреды; социальные нормы и властные отношения; роли участников и аффективные состояния. Данное множество образует многоосевую схему, соразмерную сложности коммуникативных процессов, а конкретные классификации зависят от исследовательской задачи и выбранных осей.

В данной статье рассмотрим подробнее семиотико-социокультурный подход, который представляет собой одно из наиболее перспективных направлений для анализа современных коммуникативных процессов. «Если Э. Тоффлер писал о трех цивилизационных волнах, революциях образа жизни (аграрной, индустриальной, постиндустриальной, т. е. сервисной) [Тоффлер, 2010], то в настоящее время мы вошли в четвертую волну тотальной медиализации в цифровом формате, которая на глазах трансформирует экономику, бизнес, политику, науку, образование, искусство, личную жизнь» [Тулчинский, 2023. С. 14]. В условиях, когда социальные медиа, искусственный интеллект и цифровые платформы становятся основными средствами коммуникации, понимание того, как создается и трансформируется смысл в этих средах, становится критически важным.

Систематический обзор монографий, посвященных коммуникации, показывает: единой универсальной теории коммуникации нет, поскольку разные традиции фиксируют разные ее аспекты – от канала до ритуала, от функций до норм. Это не недостаток, а ресурс: объединяя инженерную, семиотическую, медиаэкологическую, социально-философскую и психологическую традиции, мы получаем многоосевую схему, где «вид» задает уровень организации взаимодействия, а «тип» – конкретную конфигурацию по ряду измерений. Сформировав многоосевую типологию видов и типов коммуникаций – по структуре взаимодействия участников, среде и медиированности, временной организации, модальностям, институциональности и культурному контексту, – мы неизбежно упираемся в динамику: эти оси не заданы раз и навсегда, они меняются под влиянием внешних условий. Современные тенденции, обусловленные доминированием визуальной культуры, формируют у человека индивидуальное восприятие окружающего мира. Появляется множество видео, инфографики, презентаций, способных объяснить сложную теорию на одном слайде, часто состоящем преимущественно из картинок. Внимание рассеивается: считается, что «современный человек за месяц получает и обрабатывает столько же информации, сколько человек XVII века – за всю жизнь» [Постолатий, 2013], поэтому из-за переизбытка информации новости о войне или голоде оказываются равнозначными новостям о покупке нового предмета одежды.

Существует мнение, что сократится число свежих интересных длинных текстов. «Визуальный контент низкого качества (мемы, селфи) будет превосходить по объему и распространению

нию качественный. Однако сегодняшнюю попытку оцифровать все возможное и невозможное предлагают воспринимать как книгопечатание, пришедшее на смену устным рассказам. Таким образом, культура печатного слова замещается культурой потребления и создания визуального контента, что ведет к активной замене словесных текстов креолизированными. Этот новый этап приводит к трансформации речи, восприятия, привычек и коммуникации современного человека» [Кривопалова, 2025. С. 239]. Технологии (мобильные устройства, платформы, ИИ), правила и нормы (модерация, приватность), новые средства выражения (мультимодальность, эмодзи), а также ролевые и культурные ожидания способны сдвигать границы между типами общения. Так, приватный диалог может легко стать публичным из-за скриншотов и репостов.

Далее мы разберем, какие именно факторы трансформируют коммуникацию, покажем, какие параметры меняются первыми, что запускает цепочки изменений и в каких случаях происходит качественный переход от одного типа общения к другому.

Факторы, влияющие на трансформацию способов коммуникации

Технологические изменения и новые каналы общения

Современная эпоха характеризуется беспрецедентным темпом технологических изменений, которые кардинально трансформируют коммуникативные практики. Каждая новая технологическая волна не просто добавляет новые инструменты коммуникации, но переосмысливает саму природу человеческого взаимодействия. Развитие цифровых технологий вводит новые каналы коммуникации: мессенджеры, соцсети, видеосвязь, виртуальные пространства. Эти каналы меняют характер общения – оно становится мгновенным, но менее личным. Резко возросла роль в цифровой среде фатического общения – повседневные «микроакты» поддержания контакта (приветствия, «как дела?», эмодзи, реакции). Фатическое общение в цифровом формате все чаще выполняет функцию эмоционального присутствия, позволяя людям чувствовать себя связанными даже при отсутствии содержательного диалога. При этом из-за стандартизации форматов (например, однотипные реакции или стикеры) такие акты могут терять искренность, превращаясь в ритуализированные жесты, направленные скорее на поддержание социальной видимости, чем на подлинный контакт.

Особое влияние на трансформацию оказало распространение мобильных устройств – мобильная революция стала первым серьезным вызовом традиционным формам общения. А появление смартфонов создало феномен «всегда включенной» коммуникации, когда люди находятся в постоянной готовности к взаимодействию. «Все больше людей в мире страдают от номофобии – страха остаться без мобильного телефона или без доступа к нему. Из-за разряженной батареи, отсутствия интернет-подключения, по забывчивости или какой-то другой причине. Болезненную привязанность к гаджетам некоторые психиатры даже называют международной проблемой. Исследования показывают: тяга к постоянному просмотру контента на смартфоне вызвана не столько любовью к устройству или его содержанию, сколько нехваткой психологического контакта в реальной жизни и недостаточностью эмоционального отклика в живом общении. И хотя в официальной номенклатуре США по психическим расстройствам и заболеваниям (Diagnostic Manual of Mental Disorders) такого термина пока нет, там уже содержится его описательная характеристика. Некоторые исследователи уже ставят подобное расстройство в один ряд с другими фобиями: боязнью определенных животных (герпетофобия – боязнь змей, арахнофобия – боязнь пауков и др.), боязнью высоты, боязнью замкнутого пространства и т. п. У людей, страдающих номофобией, наблюдаются те же симптомы, что и при многих других фобиях: нервная дрожь, усиленное потоотделение, затрудненное дыхание, резкая взволнованность» [Хвостик, 2024].

Помимо номофобии, желание быть «всегда на связи» привело к фрагментации внимания, формированию новых этикетных норм цифрового общения, а также социального давления

и, впоследствии, выгорания. «Индикаторы статуса, «двойные галочки», автоподсказки эмодзи/реакций формируют «минимальные достаточные» ответы. Фатические жесты становятся быстрее и мультимодальнее (теперь это не только «привет», но и 👍/❤️/🔥 под постом)» [Данези, 2017. С. 176]. Корпоративные чаты поощряют краткие реакции как сигнал «получено/согласен», заменяя развернутые подтверждения. Платформы ранжируют контент с учетом «микрововлеченности», стимулируя фатику как стратегию видимости. В совокупности платформенная экономика изменила структуру коммуникативного пространства. Социальные сети, мессенджеры и другие цифровые платформы создают замкнутые экосистемы общения, каждая из которых имеет свои правила, алгоритмы и ограничения. Это приводит к фрагментации коммуникативного опыта и формированию «пузырей фильтров» (персонализированных замкнутых коммуникационных пространств).

Социокультурные изменения

Изменения в коммуникации не ограничиваются технологической сферой, но тесно связаны с глубокими социокультурными трансформациями современного общества. Индивидуализация общества приводит к изменению характера социальных связей. Традиционные формы коллективной идентичности ослабевают, уступая место более гибким и временным формам социального взаимодействия. Наряду с индивидуализацией, современное общество характеризует скорость, мобильность и разнообразие культур. Это воздействует на нормы взаимодействия, формируя тенденцию к упрощению языка и сокращению личных ничем не опосредованных контактов.

1. Психологические факторы

Трансформация коммуникативных практик сопровождается серьезными изменениями в психологических механизмах общения. Снижение эмоционального интеллекта – одна из причин ухудшения качества коммуникаций. Неспособность слушать, отсутствие эмпатии и склонность к поверхностной коммуникации (фатическому общению) приводят к недопониманию и конфликтам. Также на ухудшение качества общения влияет синдром непрерывной частичной концентрации внимания, который стал характерной особенностью современного коммуникативного поведения. Люди одновременно участвуют в нескольких коммуникативных актах, что снижает качество каждого из них. «Эмодзи/реакции выступают «быстрыми регуляторами» тона: сглаживают резкость, маркируют иронию, выражают эмпатию без длинных текстов. Это снижает порог включения, но уплощает спектр эмоций до стандартизованного набора» [Данези, 2017. С. 176].

2. Потеря навыков живого общения

Эпоха цифровизации снижает повседневную потребность в развитии навыков невербального общения, когда значительная доля взаимодействий перемещается в текст, эмодзи и краткие реакции, репертуар таких сигналов сужается и используется реже [Теркл, 2015. С. 37–38]. Итогом становится ослабление межличностной чувствительности – способности точно распознавать эмоциональные состояния, намерения и потребности собеседника и своевременно на них реагировать – и, как следствие, рост субъективного чувства одиночества и объективной социальной изоляции. Понимание механизмов этих изменений помогает объяснить, почему эффект устойчив: 1) сужение канала сигналов. Текст и «микрореакции» (лайки, эмодзи) заменяют богатые многоканальные «подсказки» лица, голоса и тела. Даже видео- и аудиосвязь страдают от задержек, несоответствия линии взгляда камере, что искажает восприятие контакта глазами и регуляцию пауз; 2) микроформаты. Привычка к коротким сообщениям, «умным ответам» и реакциям поощряет обмен «минимально достаточными» сигналами вместо развернутых реплик с уточнениями и эмпатией (вновь отсылка к фатическому общению, менее содержательному и глубокому); 3) ИИ как посредник в общении. Платформы подсказывают

тон и стиль, сглаживая индивидуальные особенности и перераспределяя усилия: пользователь меньше практикует самостоятельную формулировку предложений, опираясь на подсказки ИИ; 4) Нормативное давление «всегда в сети». Индикаторы «онлайн/прочитано», ожидания быстрых ответов и многозадачность подрывают глубину внимания, необходимую для тонкого считывания контекста и выстраивания эмпатии [Марк, 2018. С. 4–6].

Эти механизмы не только ухудшают точность распознавания эмоций и намерений, но и меняют поведенческий репертуар: увеличивается частота «микрооткликов» вместо содержательной обратной связи; падает терпимость к паузам и неопределенности, возрастает вероятность поспешных интерпретаций. Распространяются практики избегания (ghosting), снижающие навыки конструктивного завершения беседы и разрешения конфликтов (собеседник может занести в «черный список», заблокировать того, с кем не хочет общаться). Возникают новые конфликты ожиданий (например, «почему ответил эмодзи, а не словами?»), что усиливает социальное напряжение. Замещение очных встреч в онлайн-формате лишает человека критически важной практики невербальной коммуникации. Так, удаленная работа и обучение без целенаправленных коммуникативных практик для установления контакта и создания комфортной атмосферы («small talk», встречи с включением камеры, фокус на невербальные сигналы) усиливают отчуждение. Для людей с социальной тревожностью цифровая среда, при отсутствии поддержки, закрепляет стратегии избегания.

Последствия проявляются на нескольких уровнях:

- индивидуальный: падение самооэффективности в трудных разговорах/переговорах, «эмоциональная глухота» к нюансам, утомление от общения;
- групповой: рост недопонимания, обид, снижение доверия и сплоченности;
- социальный: увеличение числа «слабых» контактов при одновременном дефиците «сильных» связей, что ощущается как изоляция даже при высокой онлайн-активности (количество друзей в ВК не равно количеству друзей в реальной жизни).

Итак, снижение практики живого общения в цифровой среде снижает навыки работы с невербальными сигналами и межличностную чувствительность, что способствует социальной изоляции. Но при грамотном выборе каналов, развитии эмоциональной чувствительности эти риски можно существенно уменьшить, сохраняя удобства цифровых инструментов и поддерживая качество человеческих взаимоотношений.

Влияние искусственного интеллекта на коммуникацию

Исторически технологии меняли главным образом каналы и среду передачи – тезис «медиум – это сообщение» фиксирует структурное влияние технологической среды на содержание и эффекты коммуникации [Маклюэн, 2019. С. 15–16]. Искусственный интеллект представляет собой наиболее радикальную трансформирующую силу в современной коммуникации. В отличие от предыдущих технологических инноваций, которые лишь изменяли каналы и способы передачи информации, ИИ становится когнитивным участником – подсказывает, редактирует, переводит, генерирует и интерпретирует, влияя на намерения, форму и интерпретации в межличностном общении. Генеративные модели языка достигли такого уровня, при котором машинно-сгенерированный текст часто неотличим от человеческого. Это создает фундаментальные вопросы об аутентичности коммуникации и природе человеческого творчества. Большие языковые модели способны не только воспроизводить существующие паттерны общения, но и создавать новые формы языкового выражения.

Наряду с негативным влиянием, использование ИИ вносит существенные улучшения в человеческое общение. Алгоритмы снижают барьеры доступности и языка, ускоряют координацию и обмен смыслами, расширяют выразительные средства и поддерживают аффективную

коммуникацию. Ниже рассмотрим ключевые позитивные эффекты – от инклюзии и эффективности до творческого соавторства и эмоциональной поддержки.

Доступность и инклюзия. ИИ заметно снижает барьеры в общении для людей с ограничениями слуха, речи, зрения, моторики и с когнитивными особенностями: автоматическое распознавание речи и синтез голоса обеспечивают субтитры и озвучивание в реальном времени «RogerVoice – приложение, обеспечивающее субтитры для телефонных звонков, делая голосовую связь доступной для людей с нарушением слуха; Seeing AI – приложение от Microsoft для людей с нарушением зрения, которое использует компьютерное зрение для предоставления описаний окружающей среды и объектов» [Буянов, 2024. С. 672–673]; интеллектуальное шумоподавление и корректоры дикции улучшают разборчивость; нейросетевой перевод убирает языковые барьеры в повседневных диалогах; ассистивные интерфейсы и системы дополненной и альтернативной коммуникации (ААС) предлагают предиктивный ввод, простые шаблоны и пиктограммы; генеративные инструменты помогают перефразировать и упрощать текст, делая сообщения более доступными и понятными. Также во многих платформах для видеозвонков (к примеру, Zoom, Skype, Teams и пр.) встроен автоматический протокол встреч и краткий пересказ, повышающий скорость обработки информации и эффективность встреч.

Поддержка эмоциональной выразительности. Использование эмодзи в коммуникации позволяет снизить социальное напряжение и установить дружеский контакт. Эмодзи обеспечивают механизмы передачи тона голоса и эмоциональных нюансов, которые отсутствуют в текстовом сообщении. Отношение современного человека к эмодзи серьезнее, чем к текстовому украшению. Теперь эмодзи являются знаками того, как мы думаем и намерены действовать. «Эмодзи многое рассказывают о нас сегодня, не так ли?» [Данези, 2017. С. 26–27]

Творческое соавторство и обучение. Искусственный интеллект все чаще выступает не только инструментом, но и участником (соавтором) творческой деятельности: он предлагает идеи, продолжает текст, генерирует наброски, уточняет формулировки и дает обратную связь. Практический вклад ИИ в соавторство проявляется в трех звеньях креативного цикла: генерации (идеи, варианты, черновики), трансформации (редактура, резюмирование) и оценке (критика, рецензирование). На этапе генерации модели помогают преодолеть «синдром пустой страницы» и расширяют пространство поиска; на этапе трансформации – ускоряют уточнение формы и структуры; на этапе оценки – формируют быструю обратную связь и чек-листы критериев. Эти же механизмы ложатся в основу обучающих сценариев.

В обучении ИИ выступает как адаптивный наставник и «надстройка» над практиками самообъяснения и рефлексии: он задает наводящие вопросы, предлагает примеры, выявляет пробелы, подсказывает следующий шаг в зоне ближайшего развития, а также составляет учебный трек, подходящий под цели учащегося. Ключевые позитивные эффекты в творческом соавторстве – расширение идейного пространства, снижение транзакционных издержек и ускорение итераций. Пользователь быстрее переходит от замысла к прототипу. Примеры совместного творчества человека и машины: совместное создание литературных произведений с GPT-моделями, музыкальные коллаборации: ИИ предлагает мелодии, а музыкант их дорабатывает; дизайн и изобразительное искусство, где ИИ генерирует концепты, а дизайнер выбирает направления. К примеру, «Пытаясь проснуться» – это первый в мире сборник рассказов, родившийся в результате сотрудничества писателя и его «двойника»-нейросети. Из 24 текстов в нем только половина принадлежит автору – Павлу Пепперштейну – еще дюжину сочинила генеративная нейросеть *tuGPT-3*, дополнительно обученная на рассказах Павла [Шаврина, 2022].

В обучении преимуществами ИИ является персонализация траекторий, своевременная подсказка и развитие метакогнитивных навыков за счет промптов на рефлекссию и самооценку. Эти преимущества, однако, тесно связаны с рисками. Основные риски – стилистическая унификация, размывание границ авторства: права на произведение, зависимость от подсказок. Когнитивная разгрузка и атрофия мыслительных процессов происходят, когда люди делегируют ИИ функции, связанные с анализом информации, принятием решений и творческим

мышлением. Постоянная опора на ИИ-помощников может привести к снижению критического мышления и способности к самостоятельному решению проблем.

Психологические аспекты общения с ИИ. Взаимодействие с искусственным интеллектом все глубже проникает в повседневную жизнь, затрагивая не только бытовые и профессиональные сферы, но и психологические основы человеческого общения. Эти изменения затрагивают фундаментальные аспекты самовосприятия, социальной эмоциональности и способности к межличностному взаимодействию, вызывая как тревожные, так и обнадеживающие последствия.

С одной стороны, психологические риски использования ИИ в коммуникации связаны с феноменами антропоморфизации и иллюзии социальности. Эффект ELIZA – термин, возникший еще в 1960-х годах при тестировании первой программы-чата, – демонстрирует, насколько легко люди склонны приписывать человеческое сознание и эмпатию даже примитивным алгоритмам [Вейценбаум, 1966. С. 38]. Сегодня эта тенденция проявляется особенно остро: пользователи все чаще воспринимают ИИ-аватаров, голосовых помощников и чат-ботов как социальных партнеров, способных на понимание, поддержку и эмоциональный отклик. Такое восприятие может приводить к формированию псевдоэмоциональных привязанностей, особенно у уязвимых групп – пожилых людей, одиноких, переживающих депрессию или социальную изоляцию. В долгосрочной перспективе это несет риск подмены реальных человеческих отношений виртуальными суррогатами, снижая мотивацию к подлинному социальному взаимодействию и способствуя эмоциональной дезориентации.

Однако, несмотря на эти вызовы, взаимодействие с ИИ открывает и значительные позитивные возможности, особенно в сфере психического здоровья. Искусственный интеллект демонстрирует высокий потенциал как инструмент терапевтической и поддерживающей коммуникации. Современные чат-боты для ментального здоровья, такие как Woebot, Wysa или Tess, предлагают доступную, анонимную и круглосуточную психологическую поддержку. Они используют проверенные методы когнитивно-поведенческой терапии [Фитцпатрик, 2017. С. 19], техники мотивационного интервьюирования и элементы внимательности, помогая пользователям осознавать автоматические негативные мысли, справляться с тревогой и управлять стрессом.

Более того, исследования, опубликованные в ведущих психологических журналах, включая *Journal of Medical Internet Research* и *Nature Mental Health*, подтверждают, что взаимодействие с ИИ может снижать уровень тревожности в коммуникации, особенно у людей с социальной фобией или высокой тревожностью [Люкстон, 2011. С. 473]. Для таких пользователей ИИ-собеседник становится безопасным пространством – «тренировочным полигоном», где можно отрабатывать навыки выражения эмоций, формулирования просьб и ведения диалога без страха осуждения. Такой подход способствует повышению коммуникативной уверенности и может стать промежуточным этапом на пути к более активному участию в реальных социальных взаимодействиях.

Таким образом, психологическое воздействие общения с ИИ носит двойственный характер: с одной стороны, оно несет риски искажения социальной реальности и эмоциональной подмены, с другой – открывает новые горизонты для доступной психологической помощи и развития коммуникативного потенциала личности. Устойчивость искусства общения возможна при балансировании использования ИИ и развития личностных коммуникативных компетенций с акцентом на эмпатию, эмоциональную доступность и навыки живого взаимодействия. Не стоит забывать, об этической стороне в ИИ-коммуникации. Использование искусственного интеллекта поднимает фундаментальные этические вопросы. Проблема согласия и прозрачности возникает, когда люди не осведомлены о том, что взаимодействуют с ИИ-системой. Скрытое использование ИИ в коммуникативных платформах может нарушать принципы информированного согласия и автономии личности. Если прежние инновации главным образом расширяли каналы и среду, то ИИ вмешивается в саму ткань межличностной коммуникации: он пере-

распределяет роли, авторскую принадлежность и эмоциональные сигналы, меняет правила интерпретации и подтверждения подлинности, а также «учит» нас новым формам общения – от выбранных эмодзи до совместного письма с машиной. Это требует переосмысления норм, этики и межличностного общения в пространстве, где посредником выступают алгоритмы.

Заключение

Проведенное исследование позволяет сделать вывод о том, что современное общество переживает период беспрецедентной трансформации коммуникативных практик. Традиционное искусство общения, основанное на непосредственном межличностном взаимодействии, не просто дополняется цифровыми формами коммуникации, но подвергается фундаментальному переосмыслению под воздействием искусственного интеллекта и других передовых технологий. Перспективы развития коммуникативной сферы связаны с необходимостью разработки новых образовательных программ, этических стандартов и технологических решений, которые поддерживают человеческую природу общения при использовании преимуществ цифровых технологий.

Список литературы

- Буянов М. П., Чернышева А. С., Рудяк Ю. В. Использование ИИ в мобильных приложениях для людей с ограниченными возможностями // Тольятти: Вестник науки. 2024. Т. 2, № 5 (74). С. 669–676.
- Гуляев Д. Ю. Коммуникация как тип социальной связи: к вопросу о теоретической идентификации // Воронеж: Вестник ВГУ. Серия: Лингвистика и межкультурная коммуникация. 2008. № 3. С. 274–280.
- Кривопалова Н. О. Замещение текста визуализацией // Идеи и идеалы. 2025. Т. 17, № 2. Ч. 1. С. 231–244.
- Маклюэн М. Понимание медиа: внешние расширения человека. М.: Канон-Пресс-Ц, 2019. 496 с.
- Постолатий В. Объем информации в мире будет удваиваться каждые два года. URL: <https://rg.ru/2013/05/14/inf-a-site.html> (дата обращения: 06.12.2025)
- Тоффлер Э. Третья волна. М.: АСТ, 2010. 784 с.
- Тульчинский Г. Л. Междисциплинарный потенциал социальной семиотики и антропология цифровизации // Социокультурные трансформации в цифровую эпоху. 2023. С. 10–18.
- Хвостик Е. ИИ и будущее коммуникаций: как технологии меняют взаимодействие. URL: <https://www.kommersant.ru/doc/6464017> (дата обращения: 03.09.2025)
- Шаврина Т. А., Шевелев Д., Шляжко О., Марков С. Это наконец произошло: нейросеть и человек написали книгу. Вместе! Рассказываем, как им помогали разработчики. URL: <https://habr.com/ru/companies/sberdevices/articles/667582/> (дата обращения: 10.10.2025)
- Danesi M. The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet. Bloomsbury Academic, 2017, 209 p.
- Fitzpatrick K. K., Darcy A., Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial // JMIR Mental Health. 2017. Vol. 4. No. 2. P. 1–11.
- Luxton D. D., McCann R. A., Bush N. E., Mishkind M. C., Reger G. M. mHealth for Mental Health: Integrating Smartphone Technology in Behavioral Healthcare // Professional Psychology: Research and Practice. 2011. Vol. 42. No. 6. P. 505–512.
- Mark G., Gudith D., Klocke U. The cost of interrupted work: More speed and stress // Conference on Human Factors in Computing Systems, CHI. 2008.
- Turkle S. Reclaiming conversation: The power of talk in a digital age. Penguin Press, 2015. 377 p.

Weizenbaum J. ELIZA – A computer program for the study of natural language communication between man and machine // *Communications of the ACM*. 1966. Vol. 9. No. 1. P. 36–45.

References

- Buyanov M. P., Chernysheva A. S., Rudyak Yu. V.** Using AI in mobile applications for people with disabilities. *Vestnik Nauki*, no. 5 (74), vol. 2, pp. 669–676. (in Russ.)
- Danesi M.** The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet. Bloomsbury Academic, 2017, 209 p.
- Fitzpatrick K. K., Darcy A., Vierhile M.** Delivering Cognitive Behavior Therapy to Young Adults with Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 2017, vol. 4, no. 2, pp. 1–11.
- Guliaev D. Iu.** Communication as a Type of Social Connection: On the Question of Theoretical Identification. *Vestnik of Voronezh State University. Series: Linguistics and Intercultural Communication*, 2008, no. 3, pp. 274–280. (in Russ.)
- Krivopalova N. O.** Replacing Text with Visualization. *Idei i Idealy*, 2025, vol. 17, no. 2(1), pp. 231–244. (in Russ.)
- Khvostik E.** AI and the Future of Communications: How Technologies Are Changing Interaction. URL: <https://www.kommersant.ru/doc/6464017> (date of access: 03.09.2025) (in Russ.)
- Luxton D. D., McCann R. A., Bush N. E., Mishkind M. C., Reger G. M.** mHealth for Mental Health: Integrating Smartphone Technology in Behavioral Healthcare. *Professional Psychology: Research and Practice*, 2011, vol. 42, no. 6, pp. 505–512.
- Mark G., Gudith D., Klocke U.** The cost of interrupted work: More speed and stress. *Conference on Human Factors in Computing Systems, CHI*, 2008.
- McLuhan M.** Understanding Media: The Extensions of Man. Moscow: Kanon-Press-Ts., 2019, 496 p. (in Russ.)
- Postolatii V.** The volume of information in the world will double every two years. URL: <https://rg.ru/2013/05/14/infa-site.html> (date of access: 06.12.2025) (in Russ.)
- Shavrina T. A., Shevelev D., Shliazhko O., Markov S.** It finally happened: a neural network and a human wrote a book together! How developers helped them. URL: <https://habr.com/ru/companies/sberdevices/articles/667582/> (date of access: 10.10.2025) (in Russ.)
- Toffler A.** The Third Wave. Moscow: AST publ., 2010, 784 p. (in Russ.)
- Tulchinskii G. L.** Interdisciplinary Potential of Social Semiotics and the Anthropology of Digitalization. *Sociocultural Transformations in the Digital Age*, 2023, pp. 10–18. (in Russ.)
- Turkle S.** Reclaiming conversation: The power of talk in a digital age. Penguin Press, 2015, 377 p.
- Weizenbaum J.** ELIZA – A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 1966, vol. 9, no. 1, pp. 36–45.

Информация об авторе

Кривопалова Ника Олеговна, аспирант Гуманитарного института Новосибирского государственного университета

Information about the Author

Nika O. Krivopalova, Postgraduate Student, Humanities Institute, Novosibirsk State University

Статья поступила в редакцию 14.12.2025;

одобрена после рецензирования 30.01.2026; принята к публикации 06.02.2026

The article was submitted 14.12.2025;

approved after reviewing 30.01.2026; accepted for publication 06.02.2026

Научная статья

УДК 81'367

DOI 10.25205/1818-7935-2026-24-2-146-159

Директивные речевые акты в рекламном дискурсе (на примере английского и русского языков)

Анастасия Владимировна Николаева¹

Любовь Валентиновна Шорина²

Светлана Николаевна Чижикина³

¹ Государственный университет просвещения
Москва, Россия

² Финансовый университет при Правительстве РФ
Москва, Россия

³ Российский государственный аграрный университет –
Московская сельскохозяйственная академия им. К. А. Тимирязева
Москва, Россия

anikolaeva421@gmail.com, <https://orcid.org/0000-0002-9588-427X>

lyvshorina@fa.ru, <https://orcid.org/0009-0008-5233-9125>

arlana.rus@gmail.com, <https://orcid.org/0000-0001-9963-4692>

Аннотация

В условиях высокой конкуренции на рынке товаров и услуг компании вынуждены искать все более эффективные методы привлечения внимания потребителей. В данном исследовании произведен анализ директивных речевых актов на примере англоязычных и русскоязычных рекламных текстов и подробно рассмотрены аспекты их использования. Директивные речевые акты (ДРА) представляют собой сложные и многогранные явления, которые можно анализировать с различных точек зрения. Они могут быть классифицированы по степени обязательности выполнения действия и по вежливости. ДРА выполняют важную коммуникативную функцию, влияя на поведение других людей и выражая социальные отношения между говорящим и адресатом. Ключевую роль в интерпретации ДРА играет контекст общения, включая социальный статус и культурные нормы. Средства выражения ДРА разнообразны и могут включать лексические и грамматические способы, такие как повелительные, вопросительные, утвердительные предложения, модальные глаголы и условные предложения. Таким образом, директивные речевые акты являются неотъемлемым элементом системы коммуникации, играя ключевую роль в социальной организации и межличностном взаимодействии. Проведенное исследование подчеркивает необходимость дальнейшей теоретической разработки и прикладного изучения директивных речевых актов в рекламе для более эффективного использования данного инструмента в маркетинговой практике.

Ключевые слова

директивные речевые акты, рекламный дискурс, побуждение, прямые и косвенные формы, суггестив, прескриптив, реквестив

Для цитирования

Николаева А. В., Шорина Л. В., Чижикина С. Н. Директивные речевые акты в рекламном дискурсе (на примере английского и русского языков) // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 146–159. DOI 10.25205/1818-7935-2026-24-2-146-159

© Николаева А. В., Шорина Л. В., Чижикина С. Н., 2026

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2

Vestnik NSU. Series: Linguistics and Intercultural Communication, 2026, vol. 24, no. 2

Directive Speech Acts in Advertising Discourse (Based on English and Russian)

Anastasia V. Nikolaeva¹, Liubov V. Shorina²
Svetlana N. Chizhikova³

¹ State University of Education
Moscow, Russian Federation

² The Financial University under the Government of the Russian Federation
Moscow, Russian Federation

³ Russian State Agrarian University –
Moscow Timiryazev Agricultural Academy
Moscow, Russian Federation

anikolaeva421@gmail.com, <https://orcid.org/0000-0002-9588-427X>
lyvshorina@fa.ru, <https://orcid.org/0009-0008-5233-9125>
arlana.rus@gmail.com, <https://orcid.org/0000-0001-9963-4692>

Abstract

In the conditions of high competition in the market for goods and services, companies are forced to seek increasingly effective methods of attracting consumers' attention. This study analyzes directive speech acts of English and Russian advertising texts and examines the aspects of their use in detail. Directive speech acts (DSAs) are complex and multifaceted phenomena that can be analyzed from various perspectives. They can be classified according to the degree of obligation to perform an action and by politeness. DSAs perform an important communicative function, influencing the behavior of other people and expressing social relationships between the speaker and the addressee. The interpretation of DSAs is significantly affected by the context of communication, including social status and cultural norms. The Means of expressing DSAs are diverse and can include lexical and grammatical ones such as imperative, interrogative, and affirmative sentences, modal verbs and conditional clauses. Thus, directive speech acts are an integral element of the communication system, playing a key role in social organization and interpersonal interaction. The conducted study emphasizes the need for further theoretical development and applied research on directive speech acts in advertising for more effective use of this tool in marketing practice.

Keywords

directive speech acts, advertising discourse, motivation, direct and indirect forms, suggestive, prescriptive, requestive

For citation

Nikolaeva A. V., Shorina L. V., Chizhikova S. N. Directive Speech Acts in Advertising Discourse (Based on English and Russian). *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 146–159. DOI 10.25205/1818-7935-2026-24-2-146-159

Введение

Реклама играет важную роль в современном обществе, существенно влияя на повседневную жизнь и поведение людей. В условиях растущей конкуренции на рынке товаров и услуг компании вынуждены искать все более эффективные методы привлечения внимания потребителей и стимулирования их к совершению покупок. Рекламные тексты должны быть построены таким образом, чтобы максимально эффективно воздействовать на целевую аудиторию. Именно директивные речевые акты часто являются тем инструментом, который побуждает потребителей к конкретным действиям – покупке продукта, обращение за услугой, посещение веб-сайта и т. д. Несмотря на очевидную значимость директивных речевых актов в рекламе, теоретическая база по этой теме остается недостаточно проработанной. В последние десятилетия исследования в области лингвистики и маркетинга уделяли значительное внимание общей теории речевых актов, однако директивные акты в контексте рекламных текстов оставались вне фокуса большинства из них. Это создает необходимость разработки и уточнения теоретических моделей, способных адекватно описать и объяснить механизмы воздействия директивных актов на аудиторию.

Практическая значимость. В условиях насыщенного и часто перенасыщенного рынка реклама должна быть точной и эффективной. Знание о том, как и какие директивные речевые акты наиболее эффективно воздействуют на разные сегменты аудитории, может существенно повысить эффективность рекламных кампаний. Это особенно важно в эпоху цифрового маркетинга, когда компании обладают возможностью адресно и оперативно взаимодействовать с потребителями через различные медиаканалы.

Психологические исследования показывают, что правильно сформулированные директивные речевые акты могут существенно повышать уровень доверия к рекламе и готовность к выполнению предложенных в ней действий. Поэтому анализ психологических механизмов восприятия директивных актов также является важной частью данного исследования.

В условиях глобализации и роста социальной ответственности бизнеса все больше внимания уделяется этичности рекламных сообщений. Исследование директивных речевых актов поможет определить границы допустимого давления на потребителей, сохраняя баланс между эффективностью рекламы и ее этической пригодностью.

Таким образом, актуальность исследования директивных речевых актов в рекламных текстах обусловлена их центральной ролью в современной коммерческой коммуникации, недостаточной теоретической проработанностью вопроса, высокой практической значимостью для индустрии рекламы, психологическим влиянием на потребителей и необходимостью соблюдения этических стандартов в маркетинговой практике.

Новизна исследования заключается в комплексном подходе к анализу директивных речевых актов в рекламных текстах, что открывает новые горизонты в понимании механизмов воздействия на потребителей. В условиях насыщенного информационного поля и высокой конкуренции на рынке умение правильно сформулировать директивные речевые акты может существенно повысить эффективность рекламы.

В международной научной литературе имеется ряд исследований, посвященных речевым актам в рекламной коммуникации. Среди наиболее значимых можно выделить работы в области лингвистики, психологии и маркетинга. Однако большинство из них концентрируются на изучении иллокутивной силы речевых актов в рекламе в целом, без детального анализа директивных компонентов, что оставляет значительное пространство для дальнейших исследований.

Цель настоящей работы – провести комплексный лингвистический анализ директивных речевых актов в рекламных текстах с целью выявления их специфики и эффективности. Для достижения данной цели были поставлены следующие задачи:

- 1) провести исследование теоретических основ и концептуальных моделей директивных речевых актов;
- 2) проанализировать наличие и разновидности директивных речевых актов в различных видах рекламных текстов;
- 3) охарактеризовать функциональные особенности и роль директивных речевых актов в рекламной коммуникации.

В работе применены как практические методы (метод сплошной выборки – когда анализу подвергается определенный объем текстов, из которых отбираются фрагменты, которые и составили объект нашего исследования), так и аналитические методы: 1) анализ лексических единиц (структурно-семантический, синтаксический), 2) стилистический анализ, 3) статистический анализ (для систематизации и классификации исследуемого материала).

Теоретической основой исследования послужили работы по теории речевых актов (Дж. Остин, Дж. Серл), теории коммуникации (Ю. Хабермас), психоллингвистике (Н. Хомский), а также исследования в области рекламной и маркетинговой лингвистики (Дж. Келлер, Ф. Котлер). Директивные речевые акты часто становятся объектом исследований. Их рассматривают на материале общественного знака (А. В. Анищенко, Т. С. Ларина), рекламных текстов (А. А. Восейкова, Е. Т. Костюшко, А. В. Николаева), вывесок и объявлений (Л. В. Шори-

на), малоформатных текстов (М. Н. Гузь, Н. В. Пигина, Т. В. Морозова), религиозных текстов (О. В. Омельченко) и др.

Результаты исследования

Директивные речевые акты (далее – ДРА) являются сложными и многогранными, их подробный анализ можно провести с учетом нескольких аспектов: их классификация, функция в языке, влияние контекста, средства выражения, а также социолингвистические и прагматические аспекты.

Директивные речевые акты можно классифицировать по различным критериям.

1. По степени обязательности:

- Прескриптив – речевой акт, который предполагает обязательное выполнение действия. Основным видом прескриптива является приказ (например, «Дети! – начала урок Людмила Михайловна. – Скоро вас всех торжественно примут в октябрю! Вы должны подготовить праздник» [Роз, 2022. С. 11])¹.
- Суггестив – речевой акт, предназначенный для предложения действия или рекомендации, но не предполагающий обязательности (например, «Ты лучше сразу по дну шагай, так вернее будет» [Малышев, 2022. С. 121], «Ну, кто сегодня со мной?», «О мой благородный рыцарь, не соизволите ли вы оказать честь одинокой фее и прокружить ее в венском вальсе?»). Самым распространенным видом суггестива является совет.
- Реквестив – менее настойчивый речевой акт просьбы, который можно выполнить или проигнорировать (например, «Можно ещё прийти?», «Ты только бабке своей не говори...», «Мне бы жене бы подарочек бы пожалуйста»). Основными видами реквестивов являются просьба и мольба.

2. По вежливости:

- Мягкие формы выражаются через использование вежливых конструкций (например, «Может, вам дежурную вызвать?», «Может, чайку стаканчик откушаете перед отъездом?»). Как правило, вежливой формой директива является форма вопроса во многих языках – японском, немецком, английском и других [Ермишина, 2023. С. 11].
- Прямолинейные и жесткие формы: открытые и прямые ДРА (например, «Не перебивай, и так говорить тяжело, слушай, что расскажу...»).

Контекст общения играет важную роль в интерпретации ДРА:

- Социальный статус и отношения между говорящим и адресатом. Например, начальник может отдать приказ, в то время как подчиненный в директивах, обращенных вышестоящим лицам, как правило, использует просьбу.
- Культурные нормы играют большую роль: в одних культурах могут быть более приемлемы прямые директивы, в других предпочтительнее мягкие формы. Например, в японской социальной иерархии учитель имеет высокое положение, и ученики не воспринимают его прямые директивы как грубые [Ермишина, 2023. С. 11].

Также использование прямых форм, а именно императива, не является нарушением правил вежливости в русском языке. Э. С. Ветрова утверждает, что «данная особенность поведения позволяет говорить о коммуникативной импозитивности русского речевого этикета – допустимости оказания прямого воздействия на собеседника – как одной из русских коммуникативных доминант» [Ветрова, 2019. С. 86].

Средства выражения ДРА разнообразны. Это могут быть:

- Лексические средства. Использование конкретных слов и выражений, таких как «пожалуйста», «будьте добры», «не в службу, а в дружбу», «будьте любезны», делают речевой акт более вежливым.

¹ Примеры директивов были взяты из произведений современной художественной литературы.

- Грамматические средства. В зависимости от языка могут использоваться различные грамматические формы для выражения директива. Вот несколько распространенных способов:

1. Повелительные предложения, или императив, в том числе и императив третьего лица: это команды или просьбы, которые непосредственно заставляют слушающего выполнить некоторое действие (например, «Закрой дверь»).

2. Вопросительные предложения. Вопросы могут использоваться в качестве вежливой просьбы выполнить действие, часто формулируются как запросы (например, «Не могли бы вы передать соль?»).

3. Утвердительные предложения. Утверждения иногда могут использоваться для косвенных запросов или предложений (например, «Было бы здорово, если бы вы могли помочь с этим»).

4. Модальные глаголы. Использование модальных глаголов, таких как «можно», «могли бы», «следует» и «должен», смягчает директиву, делая ее более вежливой, и формулирует ее как правило (например, «Вам стоит обратиться к врачу»).

5. Условные предложения, или сослагательное наклонение. Они могут подразумевать запрос или предложение, основанное на условии (например, «Если у вас будет время, не могли бы вы помочь мне с этим?»).

6. Будущее время. Использование будущего времени может подразумевать просьбу без прямой команды (например, «Ты pomoжeшь с проектом?»).

7. Косвенные средства. Использование намеков или предложений, требующих интерпретации, может смягчить директиву (например, «Мне действительно нужна помощь с этим»):

8. Интонация и паралингвистика: как произносятся указания, важный аспект, который может смягчить или усилить директивный характер высказывания.

9. Невербальные средства: жесты, мимика, которые могут сопровождать директив и добавлять ему дополнительный смысл [Чижикова, 2016. С. 65].

10. Социоллингвистические и прагматические аспекты раскрываются в ДРА.

11. Импликатуры: подразумеваемый смысл, который может отличаться от буквального. Например, вопрос «Не хочешь ли пойти в кино?» может подразумевать приглашение.

12. Прагматическая сила: то, как они воспринимаются в контексте коммуникации. Например, уместность директива может зависеть от предыдущего контекста разговора, настроения участников и многих других факторов.

13. В рекламных объявлениях на английском языке проявляется не только коммерческий интерес, но и забота о человеке [Тер-Минасова, 2000. С. 234], которая проявляется в виде набора речевых актов.

Так как прагматика рекламного текста заключается в его способности побуждать к действию или изменению мнения покупателя о товаре, то все высказывания в рекламных текстах воспринимаются нами как директивные речевые акты. Выделяют следующие типы речевых актов: репрезентативы, директивы, комиссивы, интерогативы, экспрессивы, высказывания с функцией предупреждения, пермиссивы-концессивы и декларативы. Нас интересуют директивы как акты, побуждающие слушающего совершать обозначенное текстом действие. Анализ данных речевых актов рекламных сообщений необходим, поскольку поможет понять, как происходит коммуникация между говорящим и слушающим.

Для исследования нами были отобраны 1 478 рекламных текстов в американском варианте английского языка и русском языке на сайтах и в прессе. Однозначно определить, к какой группе относится тот или иной речевой акт, иногда бывает затруднительно, и тогда мы, вслед за Е. Т. Костюшко, применяли метод трансформации высказывания, в предложение с перформативным глаголом, например, *Закажите полный прайс-лист...* Мы предлагаем, чтобы вы заказали полный прайс-лист (предложение) [Костюшко, 2020. С. 97].

Первым – самым распространенным – типом ДРА являются суггестивные речевые акты, в которых на первый план выступает явная «польза для адресата» [Костюшко, 2020. С. 95]. Данные речевые акты основаны на большом авторитете, житейском опыте адресанта и включают в себя *советы, предупреждения, предостережения, рекомендации*. Например, как в следующем рекламном тексте представлена рекомендация:

(1) *Difficulty Bathing? Not with our walk-in, lie-down bath. Adjust seat height with easy to use handset. Getting In and Out is simple with Easy-Air Bath Lift. Lower or lift yourself reliably, safely and in comfort. Cushions adjust easily to help you relax. Enjoy your bath again with Easy-Air Bath Lift, Rediscover the pleasures of a deep, relaxing bath today with your Easy-Air Bath Lift with £100 off!! Can be used by all the family.*

В данном тексте рекламируется специальная ванна для пожилых людей, испытывающих возрастные проблемы с передвижением. Хотя это и коммерческое предложение, и адресант явно заинтересован в выгоде, но, кроме рекомендации своего товара, констатирует и выгоду (£100 off) от приобретения адресатом данного товара, в том числе и нематериальную, а именно *you can enjoy a bath again, can be used by all the family, rediscover the pleasures, help you relax*.

(2) *Забудь о походах в офис турагентства! Теперь можно забронировать выгодные туры с вылетом из Москвы без необходимости покидать свой дом. Пегастуристик турагентство предлагает тур онлайн, где вы сможете найти лучшие предложения на отдых по всему миру. Не упустите шанс отправиться в увлекательное приключение и экономить свое время и силы. Начинать свое путешествие прямо сейчас, и вы обязательно найдете то, о чем всегда мечтали.*

Также и в тексте рекламы на русском языке: лексемы *выгодные туры, лучшие предложения, увлекательное приключение, экономить свое время и силы, легче и доступнее* являются языковыми маркерами рекомендации, носящей суггестивный характер.

В русском языке директив смягчается при употреблении модального *надо / нужно / можно* в составе предиката с инфинитивом со значением возможности совершения действий, как в тексте выше: *Теперь можно забронировать выгодные туры с вылетом из Москвы или ... где вы сможете найти лучшее предложения на отдых по всему миру*. Также подобное смягчение можно увидеть в следующих примерах:

(3) Для этого необходимо позвонить по тел. +7 ... *назвать номер Вашей карты* ...

(4) *Профессиональную косметику KORA PROFESSIONAL с доставкой на дом можно приобрести* через наш интернет-магазин ...

Причем в русском языке модальное слово может стоять как в препозиции к инфинитиву, так и после него:

(5) *Повысить собственную безопасность можно*, укомплектовав автомобиль ...

(6) *Узнать условия доставки по Москве и Московской обл. можно*, кликнув по ссылке.

С. Н. Глазкова называет такую конструкцию «ленивым директивом», показывая, что рекомендации с «надо» звучат менее категорично и растянуты во времени [Глазкова, 2011. С. 38]. При использовании модальных глаголов с инфинитивами у адресата создается ощущение, что о нем заботятся, и свой товар ему не навязывают, а выбор он может сделать сам: *можно забронировать* вместо *забронируйте*, *необходимо позвонить* вместо *позвоните* [Костюшко, 2020. С. 98].

Рекомендация может быть подкреплена мнением авторитетного лица, что является характеристикой суггестивного РА:

(7) *Stop shaving. Start living. Why do more people trust American Laser Centers for smoother, more beautiful skin than anyone else in the country?*

Услуги центра эпиляции представлены вниманию покупателя — они подкреплены авторитетом и вынесены в пресуппозицию интеррогатива (people trust). В русскоязычных текстах часто прибегают к утвердительным предложениям, но авторитетным лицом чаще

выступают неопределенные *миллионы* людей, возрастные категории людей, целые страны, а не эксперты:

(8) *Каждому нужен свой Фруктяш. игрушки за наклейки со скидкой до 33%. Их полюбили все: и дети, и взрослые.*

(9) *Выбор миллионов для фиксации зубного протеза. Корега.*

(10) *С тех пор миллионы людей в десятках стран стали поклонниками продукции «Сибирского здоровья».*

(11) *Почему нам доверяют пациенты всего мира более 9 лет?*

(12) *...разработана медико-биологической лабораторией GERnetic Synthese и продается по всему миру.*

(13) *На данный момент почти 270 тысяч человек во всем мире выбрали обновленный Passat CC.*

Суггестивные РА также реализуются в виде *совета и предупреждения*, в случае невыполнения которого может грозить опасность. И здесь они могут быть как прямыми, так и косвенными. Прямое предупреждение:

(14) *Предупреждаем о вреде чрезмерного потребления пива будет содержать соответствующую лексику.*

Косвенные предупреждения могут встречаться в конце рекламного текста и имеют соответствующую структуру телеграфного стиля. Они касаются правильности использования товара (*Not intended for puppies / For indoor use only*); возраста того, кто будет пользоваться (*For ages 8 and older*); доставки (*Ground delivery only*); сборки товара (*Minor assembly required*); невозвратности товаров (*Personalized items are not returnable (non- returnable)*); доплаты за доставку (*Additional shipping charge applies*); оплаты (*Discounts do not apply. / Cancelled / Returned merchandise fee of \$200 applies. / Additional shipping charge applies*); гарантии (*90-day warranty / One-year warranty*). Данный стиль текстов характеризуется нераспространенными предложениями с количеством слов до десяти, в которых опускаются служебные слова. Они уместны и необходимы в рекламе, привлекают внимание и легче запоминаются, придавая тексту динамику [Дьякова, 2010. С. 94].

Также косвенные *предупреждения* реализуются путем перечисления проблем, которые могут возникнуть, если игнорировать предупреждение. Например, покупка данного ополаскивателя весьма желательна, ибо невыполнение действия приведет к проблемам: *the germs, plaque, gingivitis*:

(15) *<...> In fact, rinsing twice a day with Listerine gives you 24-hour protection against the germs that cause plaque and gingivitis. Do it for your mouth. Do it for life* убеждает нас таким образом, что делать это нужно не только для здоровья полости рта, но и качества жизни в целом. В рекламной статье о модных тенденциях окрашивания, читательницам советуют окрашивать волосы только у профессионалов, а игнорирование этого совета может привести к проблеме – ухудшению качества волос:

(16) *Тепло. Теплее. Горячо. Как заполучить модный тон без потери качества волос.*

Вторую по распространенности группу составляют реквестивы. Действие видится как необязательное для адресата, но выгодное для адресата сообщения. Реквестивы включают *просьбу, мольбу* и, иногда, *приглашение* [Шорина, 2007. С. 62]. Прямые реквестивы содержат такие единицы, как *please*, и используются в рекламных текстах, например, для приглашения читателя посетить сайт www.astrazeneca.com в печатной рекламе в журнале:

(17) *Please visit.*

Или при использовании на сайте – для связи с компанией. При заказе товаров онлайн просьбой может быть выбор цвет товара:

(18) *Please specify white or ivory.*

Соответственно приглашение посетить сайт без *please* также рассматриваем как реквестивы:

(19) *Visit MBUSA.com/gLe Mercedes Benz. The best or nothing.*

Одним из лексических средств создания косвенных реквестивов в английском языке видится глагол *discover*, приглашающий к действию, которое выгодно адресату:

(20) *Book now and discover the best destinations to getaway.* – Приглашение воспользоваться услугами сайта по бронированию туров.

(21) *Discover the fun of shopping!* – Приглашение посетить магазины и рестораны в парке аттракционов.

(22) *Discover what's waiting for you on your Rocky Mountaineer train journey. ... Discover our limited-time package that features four days at Moab Music Festival* – Приглашение отправиться в путешествие на поезде.

(23) *Discover DANA POINT in Southern Orange County.*

Ниже приведено больше примеров суггестивов с глаголами, приглашающими к действию:

(24) *Experience the incredible lightness of volume.*

(25) *Experience the better way to buy a luxury vehicle!*

(26) *Welcome to The Brand Room, an exclusive collection showcasing an exquisite array of premium brands ...*

(27) *Welcome to the magic of Macy's.*

(28) *In the realm of fashion, explore the epitome of elegance with Valentino's sophisticated designs ...*

(29) *Indulge in the beauty segment, where the likes of Estee Lauder, MAC Cosmetics and Yves Saint Laurent set the stage.*

В отличие от английского языка, в русском прямые ДРА распространены шире, чем косвенные:

(30) *Не спешите покидать нас!*

(31) *Знакомься, это твоя цена. Проверь ее в приложении. Летуй!*

(32) *Оставьте заявку и получите персональное предложение на первый заказ! Langery.*

Были выявлены основные структурно-синтаксические модели в английском и русском языках в рекламных текстах:

1. Модель (neg) $V_{imp} + (Obj) + (Adv)$. В этой модели есть императив в качестве главного члена предложения. Следует отметить, что отрицательный вариант включен в эту же модель:

(33) *Give your skin an instant shot of brightening hydration with our ultra-fine priming mist! Spritz it on anytime to awaken dehydrated skin and improve skin texture.*

(34) *Слушай ENERGY!*

(35) *ОТЕЛЬ, УБИЙСТВО И КИНОЗВЕЗДА. ИГРА НАЧИНАЕТСЯ! Читайте «Последнее дело Холмса» – захватывающий роман Артуро Переса-Реверте.*

2. Модель (No (Not)) + (Subj)/Gerund+ (Adv) имеет русский аналог (neg)+N+ (Adv). В этой модели мы наблюдаем употребление существительного, называющее тот или иной предмет, в качестве главного члена односоставного предложения и, в некоторых случаях, отрицательное местоимение «но» или «not» в английском языке:

(36) *9 Percale Sheet Sets for the Greatest Snooze Ever.*

(37) *Английский в формате живых диалогов с ИИ репетитором.*

(38) *Юмор FM. Шутки и не-е-е-сни.*

3. В английском языке мы наблюдаем модель No + $V_{ing} + (Adv)$. Она представляет собой односоставное безглагольное предложение, в котором герундий (Gerund) выступает в качестве главного члена предложения в сочетании с отрицательным местоимением «но» в препозиции. Такая конструкция является прямой формой выражения публичного директива в английском языке и встречается в следующем примере:

(39) *Introducing No Bending Over. No Touching Shoes. No Kidding. Sketchers Hands free Slips-in. Just step in and go. Introducing new Sketchers Hands Free Slips-Ins. No bending over. No pulling them on. No hassles. Exclusive Heel Pillow holds your foot securely in place.*

В русском языке не существует герундия, поэтому эта модель не представлена в нем.

4. Многие тексты рекламы оформлены в виде ассертива в изъявительном наклонении по модели Subject + Verb (indicative mood) + (Object) + (Adverbial Modifier), например:

(40) *The Michael Kors Ponti Comb Jacket is a great addition to your casual wardrobe, crafted with a zip fastening front along with a press stud fastening collar and hem for a comfortable fit, tonal stitching along with two zip fastening pockets give a stylish touch, completed with the Michael Kors branding.*

(41) *Our senses help us experience the world around us. Harvard researchers are working to understand, improve, and replicate these sensory organs.*

5. Многие предложения в текстах рекламы выражены при помощи вопросительных высказываний. Они образованы по модели Subject + Verb (indicative mood) + (Object) + (Adverbial Modifier). Например:

(42) *What if your pet had the advantage of probiotics? Что, если ваша кожа будет стареть в 2 раза медленнее?*

(43) *Ваша любимая мебель со временем потеряла свой первоначальный внешний вид? На ней появились мелкие царапины и потертости?*

Примеры директивных речевых актов мы свели в таблицу. По проанализированным примерам можно сделать вывод о количественном значительном преобладании суггестивных РА над реквестивными и прескриптивными директивными речевыми актами в обоих языках.

Распределение лексических и грамматических средств выражения директивов
в русском и английском языках

The distribution of lexical and grammatical means of expressing directives
in Russian and English

Английский язык		Русский язык		
	Модель	Количество (100 % – 737 англ. примеров)	Модель	Количество (100 % – 737 англ. примеров)
1	2	3	4	5
суггестивы	(neg)Vimp + (Obj) + (Adv) <i>Stop shaving. Start living.</i>	47,6 %	(neg)Vimp + (Obj) + (Adv) <i>Начинайте свое путеше- ствие прямо сейчас, и вы обязательно найдете то, о чем всегда мечтали.</i>	48,8 %
	(No) + Obj /Ving + (Adv) <i>No Bending Over. No Touching Shoes. No Kidding.</i>	10,8 %	Obj + (Adv) <i>Офальмоферон капли глазные Лечение вирусных и аллер- гических заболеваний глаз у взрослых и детей</i>	9,9 %
	Welcome to + Obj <i>Welcome to The Brand Room, an exclusive collection showcasing an exquisite array of premium brands ...</i>	2,1 %	Добро пожаловать в (на) + Obj <i>Добро пожаловать в наш чудесный магазин детской одежды</i>	1,8 %

Продолжение табл.

1	2	3	4	5
	Subj + V (indicative mood) + (Obj) + (Adv) <i>Vitamin C corrects environmental damage caused by free radicals and stimulates collagen production to renew firm skin.</i>	26,1 %	4. Subj + V (indicative mood) + (Obj) + (Adv) <i>Профессиональную косметику KORA PROFESSIONAL с доставкой на дом можно приобрести через наш интернет-магазин</i>	25,1 %
	Subject + V (indicative mood) + (Obj) + (Adv)? <i>How do you efficiently deliver feed to a farmer who knows all his animals by name?</i>	2,3 %	5. Subject + V (indicative mood) + (Obj) + (Adv)? <i>Что, если ваша кожа будет стареть в 2 раза медленнее?</i>	2,6 %
	Всего суггестивов – 88,9 %		Всего суггестивов – 88,4 %	
реквестивы	please (neg) Vimp + (Obj) + (Adv) <i>Please visit www.astrazeneca.com</i>	7 %	пожалуйста (neg) Vimp + (Obj) + (Adv) <i>Пожалуйста, помни, что у тебя действительно много причин гордиться собой</i>	7,3 %
	Всего реквестивов – 7 %		Всего реквестивов – 7,3 %	
прескриптивы	Not + Adj/Past Participle <i>Not available to minors and good only in the USA</i>	4,1 %	–	0 %
	(neg) Vimp + (N) + (Adv) <i>Будь человеком! Не продавай и не передавай вейпы и сигареты несовершеннолетним!</i>	0 %	(neg) Vimp + (Obj) + (Adv) <i>Будь человеком! Не продавай и не передавай вейпы и сигареты несовершеннолетним!</i>	4,3 %
	Всего прескриптивов – 4,1 %		Всего прескриптивов – 4,3 %	

Для наглядности результаты распределения лексических и грамматических средств в русском и английском языках в процентном соотношении вынесены в диаграммы. Большой разницы в использовании определенных моделей суггестивных и реквестивных речевых актов в целом нет, кроме случаев использования прескриптивных моделей. Ниже приведены диаграммы распределения всех лексико-грамматических моделей в обоих языках.

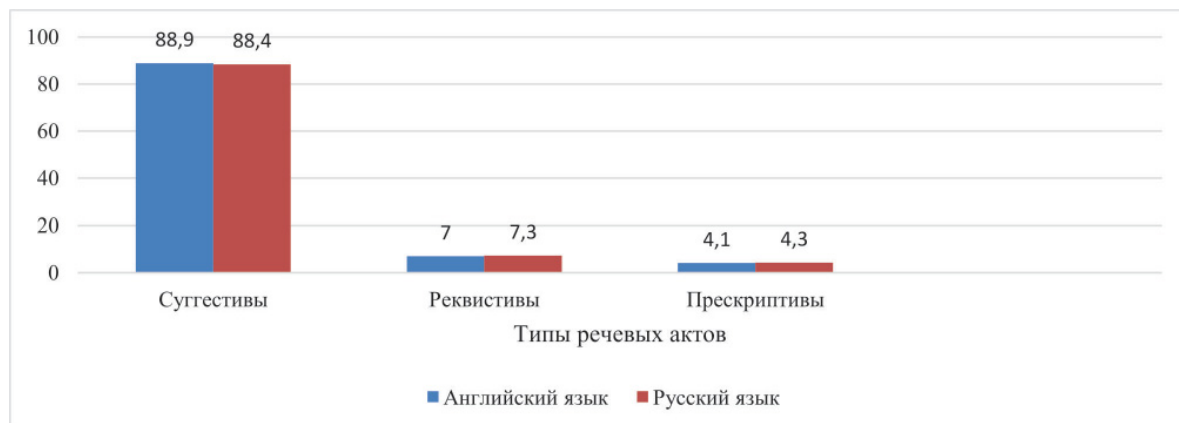


Рис. 1. Распределение видов речевых актов в английском и русском языках, %

Fig. 1. Distribution of speech act types in English and Russian, %

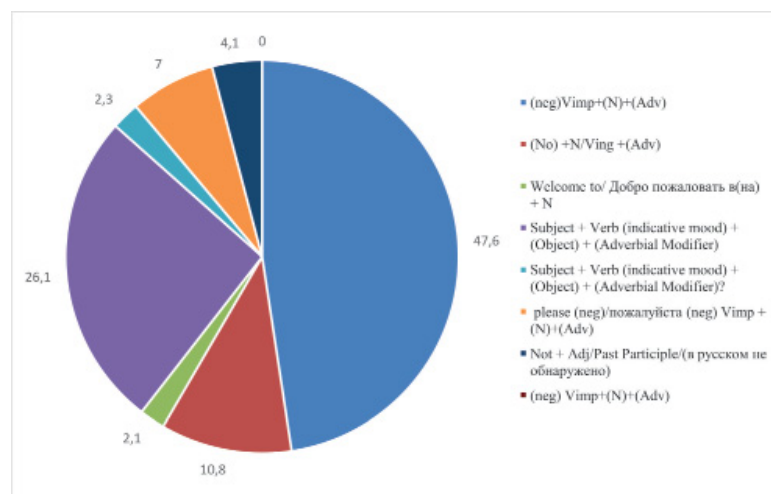


Рис. 2. Лексико-грамматические модели директивных речевых актов в английском языке

Fig. 2. Lexical and grammatical models of directive speech acts in English

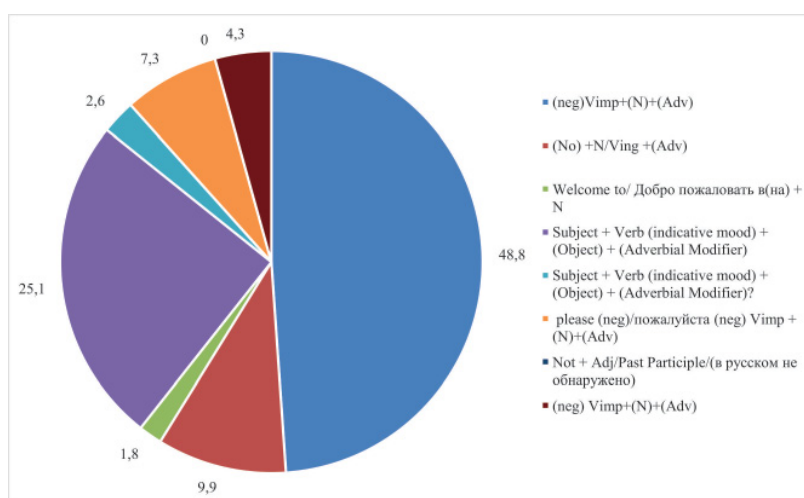


Рис. 3. Лексико-грамматические модели директивных речевых актов в русском языке

Fig. 3. Lexical and grammatical models of directive speech acts in Russian

Заключение

Директивный анализ рекламных текстов предоставляет уникальную возможность для глубокого понимания целей и побуждений, которые лежат в основе коммуникативных действий говорящего. Исследование директивных речевых актов (ДРА) открывает многогранное представление о том, как они функционируют в различных контекстах и какое значение имеют в языковой коммуникации. Классификация ДРА по уровню обязательности и вежливости подчеркивает их разнообразие.

В данной работе мы рассмотрели суггестивные, реквестивные и прескриптивные речевые акты. Примеры директивных речевых актов были показаны в таблице и на рис. 1. По проанализированным примерам можно сделать вывод о значительном количественном преобладании суггестивных РА над реквестивными и прескриптивными директивными речевыми актами в обоих языках.

Далее, внутри каждой группы был проведен анализ лексико-грамматических форм. В результате были выведены формулы, характерные для обоих языков. Для наглядности результаты распределения лексических, грамматических и синтаксических средств в русском и английском языках в процентном соотношении были также вынесены в диаграммы (см. рис. 2, 3). Подсчет показал, что нет большой разницы в использовании определенных моделей суггестивных и реквестивных речевых актов в целом, кроме случаев использования прескриптивных моделей: в русском не наблюдается использование данной формы в силу отсутствия таковой. В английском не проявили себя отрицательные формы императива.

Таким образом, директивные речевые акты являются неотъемлемым элементом рекламного дискурса, цель которого всегда заключается в изменении поведения покупателя. Так как идеи представляются в виде совета, предложения, рекомендации, просьбы и пр., происходит навязывание продукта (товара, услуги) за счет убеждения в существовании у него уникальных свойств. Понимание таких характеристик текста позволяет рекламистам более эффективно использовать РА в рекламных текстах, а покупателю – быть критичнее и уметь анализировать преподносимую информацию.

Список литературы

- Анищенко А. В., Ларина Т. С. Директивные речевые акты как лингвокультурный феномен // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2021, вып. 11 (853). С. 9–21.
- Ветрова Э. С. Средства выражения вежливости в английском и русском языках // *Studia Germanica, Romanica et Comparatistica*. 2019. Т. 15, вып. 3 (45). С. 80–90.
- Воейкова А. А. Стилистическая прагматика англоязычных рекламных текстов (на примере директивных речевых актов) // Филологические науки. Вопросы теории и практики. 2016. № 10-1 (64). С. 45–48.
- Воейкова А. А. Лингвистическая специфика вопросительных предложений в текстах русской рекламы // Международный журнал прикладных и фундаментальных исследований. 2016. № 2–4. С. 580–583.
- Глазкова С. Н. «Ленивый директив» в русской языковой картине мира // Вестник Челябинского государственного университета. 2011. № 13 (228). С. 38–43.
- Дьякова Е. Ю. Средства языковой компрессии торговых наименований в рекламном дискурсе сферы образовательных услуг // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2010. № 1. С. 94–97.
- Ермишина Е. С. Директивы в составе цитаты в современном японском языке. // *Рема. Rhema*. 2023. № 3. С. 9–32.

- Костюшко Е. Т.** Директивные речевые акты в рекламном тексте (на примере модульной рекламы в газете «СБ. Беларусь сегодня») // Славянский мир и национальная речевая культура в современной коммуникации: сб. науч. ст., посвященный 160-летию со дня рождения академика Е. Ф. Карского. Ч. 2. Гродно: Гродненский гос. ун-т им. Янки Купалы, 2020. С. 94–105.
- Малышев В. С.** Возвращение. Рассказы, стихи. М.: Вече, 2022. 304 с.
- Массальская Ю. В., Николаева А. В.** Аномалии в рекламном тексте (на примере английского языка) // Современные исследования социальных проблем. 2023. Т. 15, № 2. С. 151–167. DOI 10.12731/2077-1770-2023-15-2-151-167
- Омельченко О. В.** Директивные речевые акты предписания (разрешения) и запрета в религиозном дискурсе // Вопросы журналистики, педагогики, языкознания. 2020. № 39 (3). С. 477–486.
- Роз М. А.** Круги по воде. Новосибирск: Автограф, 2022. 147 с.
- Тер-Минасова С. Г.** Язык и межкультурная коммуникация: учеб. пособие. М.: Слово/Slovo, 2000. 261 с.
- Чижикова С. Н.** Вербальные и невербальные компоненты в креолизованных текстах (на примере мультимедийных презентаций) // Научно-методический электронный журнал «Концепт». 2016. № 7. С. 64–69.
- Шорина Л. В.** Публичный директив и его место в системе речевых актов (на материале американского варианта английского языка и русского языка): специальность 10.02.04 «Германские языки»: дис. ... канд. филол. наук. СПб., 2007. 143 с.
- Austin J. L.** How To Do Things With Words. URL: <https://archive.org/details/HowToDoThingsWithWordsAUSTIN/page/n3/mode/2up> (дата обращения: 12.07.2024).

References

- Anishchenko A. V., Larina T. S.** Directive speech acts as a linguocultural phenomenon. *Bulletin of the Moscow State Linguistic University. Humanities*, 2021, iss. 11 (853), pp. 9–21. (in Russ.)
- Austin J. L.** How to do Things with Words. URL: <https://archive.org/details/HowToDoThingsWithWordsAUSTIN/page/n3/mode/2up> (accessed: 12.07.2024).
- Chizhikova S. N.** Verbal and non-verbal components in creolized texts (using the example of highlighted presentations). *Scientific and methodological electronic journal "Concept"*, 2016, no. S7, pp. 64–69. (in Russ.)
- D'yakova E. Yu.** The means of linguistic compression of trade names in the advertising discourse of the educational services sector. *Bulletin of Voronezh State University. Series: Linguistics and Intercultural communication*, 2010, no. 1, pp. 94–97. (in Russ.)
- Ermishina E. S.** Directives as part of a quotation in modern Japanese. *Rema. Rhema*, 2023, no. 3, pp. 9–32. (in Russ.)
- Glazkova S. N.** "Lazy directives" in the Russian language picture of the world. *Bulletin of the Chelyabinsk State University*, 2011, no. 13(228), pp. 38–43. (in Russ.)
- Kostyushko E. T.** Directive speech acts in advertising text (on the example of modular advertising in the newspaper "SB. Belarus Today"). *Slavic world and national speech culture in modern communication: Collection of scientific articles dedicated to 160- anniversary of the birth of Academician E. F. Karsky. In 2 parts / Editorial Board: T. A. Pivovarchik (chief editor) [and others]. Part 2.* Grodno: Grodno State University named after Yanka Kupala, 2020, pp. 94–105. (in Russ.)
- Malyshev V. S.** Return. Stories, poems. M.: Veche, 2022. 304 p. (in Russ.)
- Massal'skaya Yu. V., Nikolaeva A. V.** Anomalies in advertising texts (based on English) / Yu. V. Massal'skaya, *Modern Studies of Social Issues*. 2023, vol. 15, no 2, pp. 151–167. DOI 10.12731/2077-1770-2023-15-2-151-167 (in Russ.)

- Omelchenko O. V.** Directive speech acts of prescription (permission) and prohibition in religious discourse. *Issues of journalism, pedagogy, linguistics*, 2020, vol. 39(3), pp. 477–486. (in Russ.)
- Roz M. A.** Circles on the water. Novosibirsk: Autograph publ., 2022. 147 p. (in Russ.)
- Shorina L. V.** Public directive and its place in the system of speech acts (based on the material of the American version of English and the Russian language): specialty 10.02.04 “Germanic languages”: dissertation for the degree of candidate of philological sciences. St. Petersburg, 2007. 143 p. (in Russ.)
- Ter-Minasova S. G.** Language and intercultural communication: Textbook. Moscow: Slovo/Slovo publ., 2000. 261 p. (in Russ.)
- Vetrova E. S.** Means of expressing politeness in English and Russian languages. *Studia Germanica, Romanica et Comparatistica*, 2019, vol. 15, iss. 3 (45), pp. 80–90. (in Russ.)
- Voeikova A. A.** Stylistic pragmatics of English-language advertising texts (on the example of directive speech acts). *Philological Sciences. Questions of theory and practice*, 2016, no. 10-1(64), pp. 45–48. (in Russ.)
- Voeykova A. A.** Linguistic specificity of interrogative sentences in the texts of Russian advertising. *International Journal of Applied and Fundamental Research*, 2016, no. 2-4, pp. 580–583. (in Russ.)

Информация об авторах

- Николаева Анастасия Владимировна**, кандидат филологических наук, доцент кафедры дошкольного образования Государственного университета просвещения
- Шорина Любовь Валентиновна**, кандидат филологических наук, доцент кафедры иностранных языков и межкультурной коммуникации Финансового университета при Правительстве РФ
- Чижикова Светлана Николаевна**, кандидат филологических наук, доцент кафедры иностранных и русского языков Российского государственного аграрного университета – МСХА им. К. А. Тимирязева

Information about the Authors

- Anastasiya V. Nikolaeva**, Candidate of Sciences (Philology), Associate Professor at the Department of Preschool Education
- Liubov V. Shorina**, Candidate of Sciences (Philology), Associate Professor at the Department of Foreign Languages and Intercultural Communication, The Financial University under the Government of the Russian Federation
- Svetlana N. Chizhikova**, Candidate of Sciences (Philology), Associate Professor at the Department of Foreign and Russian languages, Russian State Agrarian University – Moscow Timiryazev Agricultural Academy

Статья поступила в редакцию 02.10.2024;
одобрена после рецензирования 15.10.2025; принята к публикации 24.10.2025

The article was submitted 02.10.2024;
approved after reviewing 15.10.2025; accepted for publication 24.10.2025

Научная статья

УДК 81.255

DOI 10.25205/1818-7935-2026-24-2-160-172

Особенности перевода русских причастий на малайский язык

Александра Сергеевна Черных

Московский государственный университет им. М. В. Ломоносова
Москва, Россия

chernykh@my.msu.ru, <https://orcid.org/0000-0002-8126-3624>

Аннотация

В статье рассматривается проблема перевода русских причастий на малайский язык, обусловленная отсутствием прямого эквивалента данной морфологической формы в малайской грамматике. Целью исследования является разработка эффективных стратегий перевода с учетом двойственной природы русских причастий, совмещающих в себе глагольные (время, наклонение, вид) и адъективные (число, род, падеж) признаки. Анализируются различные типы русских причастий и их семантические функции в тексте, на основе чего предлагаются альтернативные стратегии, направленные на максимально точную передачу значения и грамматических особенностей. Исследование опирается на методы сопоставительного анализа языковых систем русского и малайского языков, а также контекстуальный анализ, позволяющий учитывать влияние окружающего текста на выбор оптимального варианта перевода. Результаты показывают, что прямое калькирование русских причастий в малайском языке невозможно и для достижения адекватного перевода необходимо применение различных трансформаций. В частности, предлагается замена причастий глагольными конструкциями, относительными предложениями или другими описательными оборотами, а также использование видовременных показателей и служебного слова *yang*. Особое внимание уделяется необходимости учета контекста и семантики каждого причастия, а также стремлению к стилистически корректному оформлению текста на малайском языке. Полученные результаты могут быть использованы как практическое руководство для переводчиков, работающих с парой русский – малайский, а также найти применение в образовательном процессе, способствуя развитию русской филологии в Малайзии и изучению малайского языка в России.

Ключевые слова

причастие, русский язык, малайский язык, перевод, семантика, грамматика

Для цитирования

Черных А. С. Особенности перевода русских причастий на малайский язык // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2026. Т. 24, № 2. С. 160–172. DOI 10.25205/1818-7935-2026-24-2-160-172

Analysis of Translating Russian Participles into Malay

Alexandra S. Chernykh

Lomonosov Moscow State University
Moscow, Russian Federation

chernykh@my.msu.ru, <https://orcid.org/0000-0002-8126-3624>

Abstract

This article examines the challenges inherent in translating Russian participles into Malay, given the absence of a direct grammatical equivalent in the latter. The research aims to formulate effective translation strategies that address the dual verbal and adjectival nature of Russian participles, encompassing Tense, Mood, Aspect, Number, Gender and Case. The study analyzes diverse types of Russian participles and their semantic functions, proposing alternative translation methods to accurately render both meaning and grammatical features. Employing comparative analysis of Russian and Malay language systems, alongside contextual analysis to determine optimal translation choices, the research demonstrates the infeasibility of direct calquing. Instead, it advocates for various transformations, including substituting participles for verbal constructions, relative clauses, or descriptive phrases, and utilizing aspect markers as well as the Malay particle *yang*. Careful consideration of context and a semantic nuance of each participle is emphasized to ensure stylistically appropriate Malay renderings. The findings provide a practical framework for translators working with the Russian-Malay language pair and offer implications for language acquisition, potentially advancing Russian philology in Malaysia and Malay studies in Russia.

Keywords

participle, Russian language, Malay language, translation, semantics, grammar

For citation

Chernykh A. S. Analysis of translating Russian participles into Malay. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2026, vol. 24, no. 2, pp. 160–172. DOI 10.25205/1818-7935-2026-24-2-160-172

Введение

Перевод как один из древнейших видов человеческой деятельности играет важнейшую роль в обеспечении межкультурной коммуникации и, как следствие, в развитии мировой цивилизации [Комиссаров, 1990. С. 6]. В лингвистической науке термин «перевод» охватывает как сам процесс межъязыковой трансформации, так и его результат – переведенный текст. Лингвистический подход акцентирует внимание на закономерностях и операциях, посредством которых происходит межъязыковое преобразование, рассматривая перевод как отношение между двумя языковыми единицами [Бархударов, 1975. С. 2–3]. При этом одним из ключевых требований к качественному переводу является сохранение соответствия между исходным и переводным текстами по содержанию, семантике, стилю и функциональности, причем первостепенное значение имеет адекватная передача семантической информации [Виноградов, 2001. С. 20]. В целом, стратегию переводческой деятельности определяет признание уникальности каждой языковой системы.

В связи с этим существенные грамматические различия между языками создают дополнительные трудности в процессе перевода. В частности, русский язык (РЯ), принадлежащий к восточнославянской группе индоевропейской языковой семьи, с точки зрения морфологической типологии характеризуется как флективный. Данный тип языкового строя предполагает наличие многообразия морфем в составе слова, нечеткость границ между ними и полисемантность каждой морфемы [Журинская, 1990].

В отличие от русского малайский язык¹ (МЯ), относящийся к австронезийской языковой семье, демонстрирует агглютинацию индонезийского типа. Спецификой данной разновидности

¹ Малайский язык насчитывает около 220 млн носителей [WorldData.info], имеет статус государственного в Малайзии, Индонезии, Брунее и Сингапуре и распространен в ряде регионов Юго-Восточной Азии, где проявляется в виде локальных диалектов, демонстрирующих заметное влияние местных языков. В Малайзии официальным термином для государственного языка является «малайзийский» [M(T) 10766/7/2, 2007], однако термин

сти агглютинации является отсутствие строгого разграничения деривационных и словоизменительных аффиксальных моделей, что приводит к образованию синкретических моделей, в которых аффиксы совмещают словообразовательные и грамматические функции. Кроме того, агглютинация индонезийского типа характеризуется фузией, регулярно проявляющейся в виде назального элемента на стыке префиксальной и корневой морфем. Несмотря на обязательность и регулярность, данная фузия не приводит к слиянию или переразложению морфем, сохраняя их отождествление, что является характерной чертой агглютинативных языков [Алиева, 2008. С. 112–113].

Особую сложность при переводе с РЯ на МЯ представляет отсутствие прямых эквивалентов для ряда грамматических явлений. К числу таковых относится русское причастие – глагольная форма, совмещающая признаки глагола и прилагательного, и выражающая действие или состояние, характеризующее объект. Глагольные характеристики причастия проявляются в категориях залога и вида, а также в моделях управления и примыкания. Признаки прилагательного проявляются в категориях рода, числа, падежа и способности выполнять функцию определения [Козинцева, 1990]. Данное явление отсутствует в МЯ, для которого характерно отсутствие изменений форм существительных и прилагательных в зависимости от рода и падежа, а также отсутствие личных форм и временных категорий у глаголов. Отсутствие аналогичной грамматической формы в МЯ обуславливает необходимость поиска альтернативных стратегий перевода, поскольку буквальная передача причастия посредством существующих слов и структур не представляется возможной.

Цель настоящего исследования заключается в разработке классификации русских причастий и предложении оптимальных методов их перевода на МЯ, обеспечивающих максимальную семантическую и грамматическую адекватность.

Для достижения поставленной цели в рамках данной статьи решаются следующие задачи:

- 1) определение основных типов русских причастий, учитывая их морфологические, синтаксические и семантические характеристики;
- 2) разработка стратегий перевода русских причастий на МЯ;
- 3) оценка точности и эффективности предложенных приемов с точки зрения адекватности и стилистической уместности перевода.

Предполагается, что результаты данного исследования внесут вклад в развитие теории и практики перевода, а также будут полезны для преподавателей и переводчиков, работающих с парой РЯ – МЯ.

Классификация русских причастий

В РЯ причастие представляет собой глагольную форму, характеризующую существительное (человека, животное, предмет или абстрактное понятие) и обозначающую признак, связанный с реализуемым или реализованным действием. С грамматической точки зрения русское причастие сочетает в себе категории, свойственные как глаголу, так и прилагательному. Глагольные характеристики проявляются в категориях залога (действительный или страдательный) и вида (совершенный или несовершенный), а также времени (настоящее или прошедшее). Образование русских причастий напрямую связано с категорией вида: глаголы несовершенного вида служат основой для образования причастий настоящего и прошедшего времени, в то время как глаголы совершенного вида образуют причастия исключительно прошедшего времени. Наряду с этим причастия, подобно прилагательным, изменяются по числам (единственное и множественное), падежам (именительный, родительный, винительный, творительный и т. д.) и родам (женский, мужской и средний) [Козинцева, 1990].

«малайский язык» лингвистически более корректен, учитывая полиэтнический и многоязычный характер страны, где, помимо малайского, распространены китайские идиомы, тамильский, английский и языки коренных народов, что создает дополнительную смысловую нагрузку для термина «малайзийский».

Предваряя анализ переводческих стратегий, необходимо подчеркнуть важность понимания морфологических и синтаксических особенностей русских причастий. Ниже будут рассмотрены морфологические формы и синтаксические функции русских причастий, что является необходимым условием для разработки эффективных методов их перевода на МЯ.

1.1. Морфологическая классификация русских причастий

В РЯ причастия подразделяются на две основные категории: действительные и страдательные.

Действительные причастия обозначают признак субъекта, возникший в результате его собственного действия. Таким образом, субъект является и актором действия, и носителем признака одновременно. В качестве иллюстрации действительного причастия настоящего времени рассмотрим следующий пример:

- (1) *чита-ющ-ий*² ученик
 читать-PART.PRES-NOM.SG.MASC ученик-NOM.SG.MASC
 «читающий ученик»

В примере (1) причастие *читающий* демонстрирует согласование с определяемым словом *ученик* в грамматических категориях падежа, числа и рода (именительный падеж, единственное число, мужской род).

Примером действительного причастия прошедшего времени является:

- (2) *пришед-ш-ими* ученик-ами
 прийти-PART.PST-INS.PL ученик-INS.PL
 «пришедшими учениками»

В конструкции (2) форма причастия *пришедшими* согласуется с определяемым словом *учениками* в творительном падеже и множественном числе. Важно отметить, что категория рода не выражается в причастиях, употребляемых во множественном числе.

Страдательные причастия обозначают признак субъекта, возникший в результате действия, производимого другим субъектом. Важно отметить, что страдательные причастия не образуются от непереходных глаголов (таких как *спать*, *хотеть*, *бежать* и т. п.). В русском языке выделяют полные и краткие формы страдательных причастий, при этом краткие формы образуются от полных.

В качестве иллюстрации страдательного причастия настоящего времени приведем следующее:

- (3) *чита-ем-ые* книг-и
 читать-PARTP.PRES-NOM.PL книг-NOM.PL
 «читаемые книги»

Морфологические характеристики причастия *читаемые* в примере (3) (множественное число и именительный падеж) определяются характеристиками определяемого слова *книги*.

Рассмотрим также пример страдательного причастия прошедшего времени:

- (4) *прочита-нн-ой* книг-и
 прочитать-PARTP.PST-GEN.SG.F книг-GEN.SG.F
 «прочитанной книги»

В примере (4) определяемое слово *книги* обуславливает форму причастия *прочитанной* в единственном числе, женском роде и родительном падеже.

Краткие формы страдательных причастий характеризуются ограниченным набором морфологических признаков. Рассмотрим пример:

- (5) *дверь* *закры-т-а*
дверь.NOM.SG.F *закрыть*-PARTP.PST-SG.F
 «дверь закрыта»

² Для объяснения грамматических форм русских причастий и, при необходимости, определяемого слова здесь и далее будет использоваться морфемное глоссирование (см. список сокращений).

перевод с применением трансформации русской причастной конструкции в соответствующее придаточное предложение на МЯ.

2.1.1. Трансформация в определительное придаточное предложение с относительным местоимением *yang*

Русская причастная конструкция может быть трансформирована в определительное придаточное предложение с использованием относительного местоимения *yang*, характерного для МЯ. Рассмотрим следующие примеры:

- (14) Он видел *игра-ющ-их* детей.
он видеть играть-PART.PRES-ACC.PL дети
→ РЯ: «Он видел детей, которые играли».
→ МЯ: **Dia melihat kanak-kanak yang sedang³ bermain.**
он видеть дети отн. мест. гл. показ. играть
- (15) Пётр слышал *пе-вш-ую* сестру.
Пётр слышать петь-PART.PST-ACC.SG.F сестра
→ РЯ: «Пётр слышал сестру, которая пела».
→ МЯ: **Petr dengar kakak yang sedang menyanyi.**
Пётр слышать сестра отн. мест. гл. показ. петь

В ряде случаев трансформация в определительное придаточное предложение может сопровождаться преобразованием активной конструкции в пассивную на МЯ:

- (16) *Заполня-ем-ый* папой бланк упал.
заполнять-PARTP.PRES-NOM.SG.MASC папа бланк падать
→ РЯ: «Бланк, который заполняет папа, упал».
→ МЯ: **Borang yang di-isi oleh ayah jatuh.**
бланк отн. мест. PASS-заполнять предлог отец падать
- (17) *Купле-нн-ая* мной книга интересная.
купить-PARTP.PST-NOM.SG.F мной книга интересный
→ РЯ: «Книга, которую я купила, интересная».
→ МЯ: **Buku yang saya beli menarik.**
книга отн. мест. я купить интересный

2.1.2. Трансформация в изъяснительное придаточное предложение с подчинительным союзом *bahawa*

Русская конструкция с причастием также может быть преобразована в изъяснительное придаточное предложение с подчинительным союзом *bahawa* («что»). В этом случае предложение (14) может быть переведено следующим образом:

- (14a) Он видел *игра-ющ-их* детей.
он видеть играть-PART.PRES-ACC.PL дети
→ РЯ: «Он видел, что дети играют».
→ МЯ: **Dia melihat bahawa kanak-kanak sedang bermain.**
он видеть что дети гл. показ. играть
- Предложение (15) может быть преобразовано аналогичным образом:
(15a) Пётр слышал *пе-вш-ую* сестру.
Пётр слышать петь-PART.PST-ACC.SG.F сестра
→ РЯ: «Пётр слышал, что сестра пела».
→ МЯ: **Petr dengar bahawa kakak sedang menyanyi.**
Пётр слышать что сестра гл. показ. петь

³ *Sedang* – видовременный глагольный показатель настоящего времени или процесса.

Следует отметить, что в малайских изъяснительных придаточных предложениях союз *bahawa* часто опускается без изменения смысла и структуры предложения [Nik Safiah Karim et al., 2010. P. 511]. В частности, вышеприведенные целевые предложения (14a) и (15a) могут быть представлены в виде: ***Dia melihat kanak-kanak bermain.*** (досл. «Он видел дети играют») и ***Petr mendengar kakaknya menyanyi.*** (досл. «Пётр слышал сестра поет»).

Однако данный способ трансформации не универсален. Он не применим к переводу русских предложений, содержащих страдательные причастные конструкции. Таким образом, примеры предложений (16) и (17) не могут быть преобразованы в изъяснительные придаточные предложения с малайским союзом *bahawa*. Кроме того, не все предложения с русскими действительными причастиями допускают такую трансформацию. Применимость данной стратегии во многом определяется семантикой предложения и, как правило, ограничивается случаями, когда глагол в главном предложении выражает перцептивное действие, как в примерах (14a–15a).

2.2. Перевод субстантивированных причастий: стратегии замены

В соответствии с положениями, изложенными в разделе 1.2, русские причастия проявляют способность к субстантивации, приобретая семантику и синтаксические функции, характерные для существительных, и занимают позиции подлежащего или дополнения. В контексте перевода подобных причастий на МЯ предлагается использование стратегий замены, представленных ниже.

2.2.1. Замена причастий конструкцией с местоимением *yang*

Данная стратегия подразумевает замену русского субстантивированного причастия конструкцией, включающей малайское местоимение *yang* («тот, кто; тот, что»), что позволяет передать значение субъекта или объекта действия:

- (18) *Чита-ющ-ие* довольны.
 читать-PART.PRES-NOM.PL довольный
 → РЯ: «Те, кто читают, довольны».
 → МЯ: ***Yang sedang membaca puas hati.***
 тот, кто глаг. показ. читать довольный
- (19) *Написа-вш-ий* эти слова умён.
 писать-PART.PST-NOM.SG.MASC этот слово умный
 → РЯ: «Тот, кто написал эти слова умён».
 → МЯ: ***Yang menulis perkataan ini bijak.***
 тот, кто писать слово этот умный
- (20) Мне не нравится *покупа-ем-ое* мамой.
 я не нравиться покупать-PARTP.PRES-ACC.SG.NTR мама
 → РЯ: «Мне не нравится то, что покупает мама».
 → МЯ: ***Saya tidak suka yang dibeli oleh emak.***
 я не нравиться то, что покупать предлог мама
- (21) *Украд-енн-ое* лежит в коробке.
 красть-PARTP.PST-NOM.SG.NTR лежать в коробка
 → РЯ: «То, что украли, лежит в коробке».
 → МЯ: ***Yang dicuri terletak di dalam kotak.***
 то, что красть находиться внутри коробка

2.2.2. Замена действительного причастия существительным с префиксом *peN-* (агентивным дериватом)

Для перевода русских действительных причастий, обозначающих деятеля (как настоящего, так и прошедшего времени), возможно использование малайских существительных, образованных с помощью префикса *peN-*, который служит для образования агентивных дериватов, т. е. существительных, обозначающих лицо, совершающее действие.

Рассмотрим пример (18) с действительным причастием настоящего времени:

- (18a) *Чита-ющ-ие* довольны.
 читать-PART.PRES-NOM.PL довольный
 → РЯ: «Читатели довольны».
 → МЯ: Pembaca puas hati.
 читатель довольный

Аналогично для действительного причастия прошедшего времени:

- (22) *Прочита-вш-ие* довольны.
 читать- PARTP.PST-NOM.PL довольный
 → РЯ: «Читатели довольны».
 → МЯ: Pembaca puas hati.
 читатель довольный

Важно отметить, что оба примера русских предложений (18а, 22) транслируются на МЯ идентично. Следовательно, при использовании данного подхода временная характеристика, присущая русскому причастию, нейтрализуется в переводе на МЯ.

2.2.3. Замена страдательных причастий существительными с суффиксом *-an* (результативным дериватом)

В качестве альтернативной стратегии для перевода русских страдательных причастий предлагается использование малайских существительных, образованных путем добавления суффикса *-an* к глагольной основе. Такие существительные обозначают предмет или вещь, являющуюся результатом действия, выраженного глагольной основой. Например, предложение (20) со страдательным причастием настоящего времени может быть переведено следующим образом:

- (20a) Мне не нравится *покупа-ем-ое* мамой.
 я не нравится покупать-PARTP.PRES-ACC.SG.NTR мама
 → РЯ: «Мне не нравится покупка мамы».
 → МЯ: Saya tidak suka belian emak.
 я не нравится покупка мама

Аналогично для предложения (21) со страдательным причастием прошедшего времени:

- (21a) *Украд-енн-ое* лежит в коробке.
 красть-PARTP.PST-NOM.SG.NTR лежать в коробка
 → РЯ: «Украденное лежит в коробке».
 → МЯ: Curian ada di dalam kotak.
 то, что украли находится внутри коробка

Следует учитывать, что данная стратегия не подходит для случаев, когда необходимо эксплицитно выразить на МЯ временную характеристику действия, заключенного в причастии. Примеры (20а–21а) демонстрируют потерю временной информации в результате такого перевода на МЯ.

Предложенная стратегия применима не только для перевода субстантивированных форм страдательных причастий, но и для перевода полных причастий, выступающих в роли определения, что рассматривалось в разделе 2.1:

- (23) Мы купили машину, сделанную в Китае.
 Мы купить машина делать– PARTP.PST-NOM.SG.F в Китай
 → РЯ: «Мы купили машину китайского производства».
 → МЯ: **Kita membeli kereta buatan Cina.**
 мы купить машина изделие Китай

2.3. Перевод кратких причастий в функции сказуемого: стратегии пассивизации и результативности

В РЯ краткие формы причастий обладают способностью функционировать в качестве именной части составного сказуемого. В контексте перевода на МЯ предлагаются две основные стратегии, основанные на семантических особенностях русских кратких причастий.

2.3.1. Использование пассивного залога с глаголами, оформленными префиксом *di-* или личными местоимениями первого/второго лица (пассивизация агента)

Данная стратегия направлена на передачу пассивного значения, указывающего на то, что объект подвергается воздействию со стороны агента. В МЯ это достигается использованием глаголов с префиксом *di-* (маркером пассивного залога при агенте 3-го лица) или пассивных конструкций с личными местоимениями 1-го или 2-го лица.

- (24) Книга купл-ен-а братом в магазине.
 книга купить–PARTP.PST-SG.F брат в магазин
 → РЯ: «Книга была куплена братом в этом магазине».
 → МЯ: **Buku dibeli oleh abang di kedai.**
 книга купить предл. брат в магазин
- (25) Книга купл-ен-а мной в этом магазине.
 книга купить–PARTP.PST-SG.F я в этот магазин
 → РЯ: «Книга была куплена мной в магазине».
 → МЯ: **Buku saya beli di kedai ini.**
 книга я купить в магазин этот

2.3.2. Замена пассивным глаголом с префиксом *ter-* (выражение результативности)

В случаях, когда акцент делается на состоянии завершенности действия и готовности объекта, целесообразно использовать малайские пассивные глаголы с префиксом *ter-*. В таких конструкциях агент действия, как правило, не указывается:

- (26) Рис приготовл-ен.
 Рис готовить–PARTP.PST-SG.MASC
 → РЯ: «Рис готов».
 → МЯ: **Nasi tersedia.**
 рис готовый
- (27) Журнал уже прочита-н.
 журнал уже читать–PARTP.PST-SG.MASC
 → РЯ: «Журнал уже прочитан».
 → МЯ: **Majalah sudah terbaca.**
 журнал уже читать

Таким образом, несмотря на отсутствие прямых словарных соответствий русским причастиям в МЯ, существуют разнообразные грамматические конструкции, позволяющие адекватно передать их значение. Успешный перевод требует от переводчика глубокого понимания морфологических, синтаксических и семантических особенностей русских причастий, а также хорошего владения МЯ.

Заключение

В заключение проведенного исследования можно сформулировать следующие ключевые выводы о стратегиях перевода русских причастий на МЯ:

1. При переводе русских действительных причастий целесообразно использовать малайские конструкции с действительными глаголами, а страдательные причастия заменять структурами с пассивными глаголами.

2. В РЯ категория времени является грамматически значимой и должна отражаться в формах глаголов и причастий. В МЯ категория времени отсутствует, однако существует категория аспекта, позволяющая передавать временные нюансы. Для точного перевода русских причастий с сохранением темпоральных характеристик рекомендуется использовать вспомогательные слова – видовременные показатели, указывающие на аспект действия (завершенность, процесс, длительность). При этом следует учитывать, что эксплицитное выражение времени не всегда необходимо для понимания смысла на МЯ, а в ряде случаев лаконичные малайские конструкции без вспомогательных слов оказываются предпочтительнее.

3. Наиболее распространенной стратегией перевода русских причастий является использование конструкций со служебным словом *yang*, выполняющим функцию относительного местоимения в определительных конструкциях и роль местоимения при субстантивации.

4. В МЯ отсутствуют грамматические категории числа, рода и падежа, свойственные РЯ. Хотя в большинстве случаев эти категории нерелевантны для перевода, при необходимости сохранения семантической точности (например, при указании числа или пола субъекта действия) рекомендуется использовать лексические средства, такие как удвоение существительных или использование неопределенно-количественных слов (много, мало, несколько и т. п.).

5. Для достижения адекватного и точного перевода необходимо тщательно анализировать исходный текст и учитывать семантику каждого причастия. Переводчик должен стремиться к созданию стилистически и грамматически корректного текста на МЯ, обеспечивающего полное и ясное понимание содержания исходного текста целевой аудиторией.

В заключение следует подчеркнуть, что при переводе русских причастий на МЯ не всегда целесообразно стремиться к буквальному воспроизведению всех грамматических категорий, выраженных в исходном тексте. Переводчику следует ориентироваться на семантическую релевантность этих категорий для понимания смысла оригинала, а также руководствоваться предложенными в настоящем исследовании стратегиями.

Полученные результаты могут быть использованы как практическое руководство для переводчиков, специализирующихся на переводе в паре РЯ – МЯ, а также внести вклад в развитие теоретических и практических аспектов данной области, особенно в части, касающейся передачи причастных конструкций – важного элемента РЯ, не имеющего прямых эквивалентов в МЯ. Кроме того, материалы исследования могут быть интегрированы в образовательный процесс при обучении РЯ, а также способствовать расширению научного знания в области русской филологии в Малайзии и изучения МЯ в России.

Список сокращений

МЯ – малайский язык; РЯ – русский язык; → трансформация; АСС – винительный падеж; F – женский род; GEN – родительный падеж; INS – творительный падеж; MASC – мужской род; NOM – именительный падеж; NTR – средний род; PART – действительное причастие; PARTP – страдательное причастие; PASS – пассивный залог; PL – множественное число; PRES – настоящее время; PST – прошедшее время; SG – единственное число; гл. показ. – глагольный показатель; отн. мест. – относительное местоимение.

Список литературы

- Алиева Н. Ф. Структурная эволюция малайского языка. Соотношение средств синтетизма и аналитизма в разных текстах // Маклаевский сборник. СПб.: МАЭ РАН, 2008. Т. 1. С. 48–58.
- Бархударов Л. С. Язык и перевод (Вопросы общей и частной теории перевода). М.: Международные отношения, 1975. 240 с.
- Виноградов В. С. Введение в переводоведение (общие и лексические вопросы). М.: Изд-во Ин-та общего среднего образования РАО, 2001. 224 с.
- Журинская М. А. [Электронный ресурс]. URL: <https://tapemark.narod.ru/les/511d.html> (дата обращения: 06.05.2025).
- Козинцева Н. А. [Электронный ресурс]. URL: <https://tapemark.narod.ru/les/399e.html> (дата обращения: 06.05.2025).
- Комиссаров В. Н. Теория перевода (лингвистические аспекты): Учебник для ин-тов и фак. иностр. яз. М.: Высш. шк., 1990. 253 с.
- Nik Safiah Karim, Farid M. Onn, Hashim Haji Musa. Tatabahasa Dewan (3rd ed.). Dewan Bahasa dan Pustaka, 2010.
- PM(T) 10766/7/2 «Penggunaan Terminologi Bahasa Kebangsaan yang Tepat», 2007.
- WorldData.info [Электронный ресурс]. URL: <https://www.worlddata.info/languages/malay.php> (дата обращения: 31.07.2024).

References

- Alieva N. F. Structural evolution of the Malay language. Correlation of means of synthesisism and analytism in various texts. St. Petersburg: MAE RAN, 2008, vol. 1, pp. 48–58. (in Russ.)
- Barkhudarov L. S. *Language and translation (Questions of general and specific translation theory)*. Moscow: Mezhd. otn.publ., 1975. (in Russ.)
- Vinogradov V. S. *Introduction to translation studies (general and lexical issues)*. Moscow: Izd. instit. obshch. sr. obraz. RAO, 2001. (in Russ.)
- Zhurinskaya M. A. [Online]. URL: <https://tapemark.narod.ru/les/511d.html> (accessed on: 06/05/2025) (in Russ.)
- Kozintseva N. A. [Online]. URL: <https://tapemark.narod.ru/les/399e.html> (accessed on: 06/05/2025) (in Russ.)
- Komissarov, V. N. Translation theory (linguistic aspects): Textbook for institutes and faculties of foreign languages. Moscow: Vyssh. shkola publ., 1990. (in Russ.)
- Nik Safiah Karim, Farid M. Onn, Hashim Haji Musa. *Tatabahasa Dewan (3rd ed.)*. Dewan Bahasa dan Pustaka, 2010.
- PM(T) 10766/7/2 “Penggunaan Terminologi Bahasa Kebangsaan yang Tepat” [The use of precise terminology in the national language], 2007.
- WorldData.info [Online]. – URL: <https://www.worlddata.info/languages/malay.php> (accessed on: 06/05/2025).

Информация об авторе

Черных Александра Сергеевна, старший преподаватель ИСАА МГУ им. М. В. Ломоносова

Information about the Author

Alexandra S. Chernykh, Senior Lecturer, Institute of Asian and African Studies, Lomonosov Moscow State University

*Статья поступила в редакцию 07.05.2025;
одобрена после рецензирования 24.07.2025; принята к публикации 26.07.2025*

*The article was submitted 07.05.2025;
approved after reviewing 24.07.2025; accepted for publication 26.07.2025*

Правила оформления текста рукописи

Авторы представляют статьи на русском языке объемом от 0,5 печатного листа (20 тыс. знаков, шрифт 14, межстрочный интервал 1,5) до 1 авторского листа (40 тыс. знаков) включая иллюстрации (1 иллюстрация форматом 190×270 мм = $\frac{1}{6}$ авторского листа, или 6,7 тыс. знаков). Публикации, превышающие указанный объем, допускаются к рассмотрению только после индивидуального согласования с ответственным редактором.

Требования к оформлению основного текста и иллюстративных материалов

К рукописи необходимо приложить сведения об ученой степени, ученом звании, должности и месте работы, а также контактный телефон, электронный и почтовый адрес автора.

Обязательным требованием является наличие индекса УДК (Универсальной десятичной классификации), резюме статьи на русском и английском языках (до 300 слов), а также авторский перевод названия статьи на английский язык, ключевые слова (до 10 слов) на двух языках, сведения о финансовой поддержке.

Образец оформления статьи

УДК 81 + 811.131.1 + 811.161.1

Русская и итальянская абстрактная адъективная метафоризация

Иван Иванович Иванов

Новосибирский государственный университет
Новосибирск, Россия

ivan@mail.ru, <https://orcid.org/xxxx-xxxx-xxxx-xxxx>

Аннотация

Ключевые слова

Финансирование

Исследование проведено за счет гранта Российского научного фонда (проект № 14-50-00036)

The Russian and Italian Abstract Adjectival Metaphorization

Ivan I. Ivanov

Novosibirsk State University
Novosibirsk, Russian Federation

ivan@mail.ru, <https://orcid.org/xxxx-xxxx-xxxx-xxxx>

Abstract

Keywords

Funding

The work was supported by the Russian Science Foundation, project 14-50-00036

Основной текст статьи

Список литературы

Список словарей
Список источников
References
List of Dictionaries
List of Sources
Информация об авторах/Information about the Authors

Подпись автора (авторов)

Библиографические ссылки: в тексте в квадратных скобках указывается фамилия автора, год издания. Например: [Розен, 1969. С. 5]. В конце статьи помещается список литературы в алфавитном порядке без нумерации. Библиографическое описание публикации включает: фамилии и инициалы авторов (всех, независимо от их числа), полное название работы, а также издания, в котором опубликована (для статей), город, название издательства, год издания, том (для многотомных изданий), номер, выпуск (для периодических изданий), объем публикации (количество страниц – для монографии, первая и последняя страницы – для статьи). Ссылки на архивные документы оформляются в виде сноски (текст сноски располагается внизу страницы). Библиографические источники не нумеруются!

Образцы составления библиографического описания

Авторская монография:

Ильиш Б. А. История английского языка. М.: Лит. на иностр. яз., 1958. 366 с.

Коллективная монография

(все авторы должны быть упомянуты):

Суперанская А. В., Сталтмане В. Э., Подольская Н. В., Султанов А. Х. Теория и методика ономастических исследований: моногр. М.: Наука, 1986. 298 с.

Статья в сборнике:

Черкасова Г. А. Русский ассоциативный тезаурус: компьютерная технология создания и издания // Этнокультурная специфика языкового сознания: сб. науч. ст. / под ред. Н. В. Уфимцевой. М., 2003. С. 181–190.

Статья в ученых записках (ученых трудах):

Скрипка А. С. К датировке некоторых типов сарматского оружия // Учен. зап. Ленингр. пед. ин-та. 1977. Вып. 2. С. 60–77.

Статья в журнале:

Кириллов Д. А. Образ выборов в языковом сознании российской молодежи. Опыт сравнительного исследования на материале РАС и свободного ассоциативного эксперимента // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2008. Т. 6, № 2. С. 17–24.

Автореферат:

Яньшин П. В. Психосемантический анализ категоризации цвета в структуре сознания субъекта: автореф. дис. ... д-ра психол. наук. М., 2001. 42 с.

Рецензия:

Панин Л. Г. [Рецензия] // Сибирский филологический журнал. 2003. № 3–4. С. 245–247. Рец. на кн.: Турбин Г. А., Шулежкова С. Г. Старославянский язык: учеб. пособие. М.: Изд-во Моск. гос. ун-та, 2002. 145 с.

При подготовке иллюстративного материала просим учесть, что рисунки следует предоставлять в форматах .gif, .jpg, .tif отдельными файлами.

Допускается создание таблиц и диаграмм в WinWord и Excel (обязательно приложить исходный файл в формате .xls), обязательно прилагать файлы используемых (нестандартных)

шрифтов (.ttf), кегль шрифта в надписях не должен быть меньше 9. Максимальное поле изображения 190 × 270 мм.

Все вопросы, связанные с изменением и уточнением текста в процессе редакторской правки, должны сниматься авторами в ходе переписки по электронной почте в строго определенных для этого редколлегий сроки. Нарушение сроков ведет к отказу в опубликовании статьи. *Переписка традиционной почтой не осуществляется.*

Требования к содержанию публикуемых материалов

Требования к теме исследования, заявленной в статье. Как правило, выбор аспирантом темы публикации соотносится с темой его кандидатского диссертационного исследования. Прежде чем приступить к написанию статьи, необходимо выяснить, является ли разрабатываемая аспирантом тема актуальной и новой для филологических наук. В решении этого вопроса аспиранту не следует полагаться только на мнение научного руководителя и сотрудников своего НИИ или вуза. Необходимо тщательно проверить самому, какое количество работ уже было выполнено по данной теме отечественными и зарубежными исследователями. Для уточнения ситуации с актуальностью и новизной темы нужно обратиться к авторитетным библиографическим ресурсам (rsl.ru, rnl.ru, elibrary.ru, loc.gov и т. д.) и с помощью всех ключевых слов выявить круг уже выполненных исследований. Если, например, соматическим компонентам в составе английских фразеологизмов, в том числе социолектных, уже посвящено более 150 структурно-семантических исследований, то соискателям ученых степеней стоит переключиться на другие темы.

Требования к объектной и предметной новизне исследования. Новизна лингвистического исследования может быть объектной и предметной.

Требования к описанию объекта исследования. Автор статьи в результате выполненного им исследования должен иметь наиболее полное и системное представление об изученном им объекте. Границы объекта должны быть предельно четко очерчены в статье. Например, если речь идет о языковых единицах, то должно быть указано, к какому языку или языкам какого исторического периода или периодов относятся данные единицы, каково их общее количество, из каких именно источников, письменных или устных, они были взяты и каким методом были собраны. При отсутствии у объекта точных квалификативных характеристик выводы исследования нельзя считать репрезентативными.

Требования к характеристике предмета исследования. Об объекте, даже хорошо изученном в ряде направлений, можно получить ценную новую информацию при новом подходе к его изучению. Из содержания статьи специалистам должно быть ясно, в чем именно состоит предметная новизна выполненной работы.

Требования к формулировке названий статьи. Формулировка названия статьи должна четко отражать объектные и предметные характеристики исследования. Слишком широкие формулировки, не соответствующие содержанию работы (например «Морфологические категории: коммуникативный аспект интерпретации»), вводят читателей в заблуждение.

Требования к обоснованности и достоверности научных положений и выводов, представленных в статье. Степень обоснованности и достоверности научных положений и выводов любой лингвистической работы обуславливается репрезентативностью исследовательской картотеки и применением необходимых методов исследования. В настоящее время репрезентативность большинства лингвистических исследовательских картотек должна подкрепляться компьютерными корпусными данными (коллекции ссылок на лингвистические корпусы можно посмотреть, например, на сайте <http://www.uow.edu.au/~dlee/CBLLinks.htm>). Перечни методов лингвистических исследований представлены в учебной и специальной литературе.

Требования к соотношенности полученных в исследовании новых выводов с целью и задачами, заявленными в статье. Последовательно перечисленные в конце статьи выводы исследова-

дования (в порядке убывания их значимости) должны быть скоррелированы с заявленными в начале статьи целью и задачами.

Требования к точному и последовательному использованию терминов в научном тексте. Системность – отличительная черта научного знания. Субъективный эссеизм, компилятивность, слишком вольное обращение с терминами или полное игнорирование специальной терминологии свидетельствуют о непрофессионализме автора. Все необходимые для изложения термины должны быть системно сгруппированы и употреблены автором статьи только однозначно. В статьях, посвященных металингвистическим проблемам (истории лингвистической терминологии, неоднозначности толкования терминов в различных научных школах, фиксации терминов в специальных словарях и т. д.), должны быть представлены только новые для отечественных специалистов сведения. Введение оригинальной авторской терминологии должно быть объективно обосновано.

Требования к использованию цитат. Все приведенные в тексте статьи явные и скрытые цитаты должны иметь ссылки. Реферативность изложения, обилие цитат и другие признаки «вторичного текста» не позволяют рассматривать некоторые статьи как оригинальные и самостоятельные произведения научного стиля.

Требования к отражению в статье информации о личном участии автора в исследовании. Из содержания статьи должно быть понятно, в чем именно состоит личное участие автора или авторского коллектива в получении научных результатов. В теоретических исследованиях автор (или авторы) не должен «компоновать» конспекты чужих трудов в соответствии с замыслом своей работы, а должен выходить на качественно новый уровень самостоятельной рефлексии.

Требования к отражению в статье прикладного значения полученных результатов исследования. В тексте статьи должны быть указаны сферы использования результатов, полученных автором.

Соответствие текста статьи требованиям к научному стилю и оформлению. Текст статьи должен соответствовать требованиям, предъявляемым к текстам этого жанра. Информацию о жанровой специфике научных статей можно почерпнуть из учебных пособий по научному стилю речи. Требования к оформлению статей в журнале «Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация» публикуются на соответствующих страницах в каждом номере журнала. Просим авторов обращать внимание на объем присылаемых рукописей (не менее 20 000 знаков) и недопустимость тезисного изложения материала.

Передавая рукопись статьи (произведение) в редколлегию журнала, автор тем самым предоставляет ей право использования передаваемых материалов в составе журнала следующими способами: обнародование, воспроизведение, распространение, доведение произведения до всеобщего сведения путем размещения в сети Интернет, публичный показ, а также перевод на иностранные языки, включая те же действия относительно переведенного произведения, на территории всех государств, где произведение подлежит правовой охране.

Доставка материалов

Представляемые в редакцию материалы можно передать лично (комната 1269, новый корпус НГУ) или переслать по электронной почте.

Адрес редакционной коллегии
серии «Лингвистика и межкультурная коммуникация»

Кафедра истории и типологии языков и культур
ул. Пирогова, 1, Новосибирск, 630090, Россия

Тел.: (383) 363 42 23

E-mail: lingua@vestnik.nsu.ru

Вестник Новосибирского государственного университета
Серия: Лингвистика и межкультурная коммуникация

Научный журнал

2026. Т. 24, № 2

Учредитель
Новосибирский государственный университет

Адрес учредителя, издателя и типографии
ул. Пирогова, 2, Новосибирск, 630090, Россия

Главный редактор
доктор филологических наук М. К. Тимофеева

Адрес редакции
ул. Пирогова, 1, Новосибирск, 630090, Россия

Дата выхода в свет 07.09.2026

Тираж 30 экз.

Цена 1135 руб.