



MACROSOCLINGUISTICS AND MINORITY LANGUAGES

2025 Том 3 № 2

DOI: 10.22363/2949-5997-2025-3-2

Сетевой научный журнал
Издается с 2023 г.

Издание зарегистрировано Федеральной службой по надзору в сфере связи,
информационных технологий и массовых коммуникаций (Роскомнадзор)

Свидетельство о регистрации: Эл № ФС 77-82546 от 30.12.2021

Учредитель: федеральное государственное автономное образовательное
учреждение высшего образования «Российский университет дружбы народов имени Патриса Лумумбы»

Главный редактор

**Москвичева
Светлана Алексеевна**

Кандидат филологических наук,
доцент, заведующая лабораторией
динамики языков в миноритарной
ситуации научно-образовательного
института современных языков,
межкультурной коммуникации
и миграций, Российский университет
дружбы народов,
Москва, Российская Федерация

E-mail: moskvicheva-sa@rudn.ru

Ответственный секретарь

**Борисова
Анна Степановна**

Кандидат филологических наук,
доцент, Российский университет
дружбы народов, Москва, Российская
Федерация

E-mail: borisova-as@rudn.ru

Почетный редактор Ален Вио, доктор лингвистики, директор исследований эмирита Национального центра научных исследований (CNRS), Бордо, Франция, e-mail: alain.viaut@orange.fr

Члены редакционной коллегии

Агранат Татьяна Борисовна, Институт языкознания РАН, Москва, Российская Федерация

Александрова Оксана Ивановна, Российский университет дружбы народов, Москва, Российская Федерация

Додыхудоева Лейли Рахимовна, Институт языкознания РАН, Москва, Российская Федерация

Замалетдинов Радиф Рифкатович, Казанский (Приволжский) федеральный университет, Казань, Российская Федерация

Илюха Ольга Павловна, Институт языка, литературы и истории Карельского научного центра Российской академии наук (ИЯЛИ КарНЦ РАН), Петрозаводск, Российская Федерация

Марцинковская Татьяна Давидовна, Институт психологии им. Л.С. Выготского, Российский государственный гуманитарный университет, Москва, Российская Федерация

Солнышкина Марина Ивановна, Казанский (Приволжский) федеральный университет, Казань, Российская Федерация

Эбзеева Юлия Николаевна, Российский университет дружбы народов, Москва, Российская Федерация

MACROSOCIOLINGUISTICS AND MINORITY LANGUAGES

ISSN 2949-5997 (Online)

2 выпуска в год

Языки: аварский, адыгейский, алтайский, английский, башкирский, бурятский, вепсский, даргинский, ингушский, испанский, итальянский, кабардино-черкесский, калмыцкий, карачаево-балкарский, карельский, коми, крымско-татарский, кумыкский, лакский, лезгинский, марийский, мордовский, ненецкий, немецкий, осетинский, русский, татарский, тувинский, удмуртский, французский, хакасский, чеченский, чувашский, якутский.

ЦЕЛИ И ТЕМАТИКА

Журнал *Macrosociolinguistics and Minority Languages* — периодическое международное рецензируемое научное издание в области социолингвистики и междисциплинарных исследований.

Актуальность и тематика

Узкоспециализированная тематика, связанная с дисциплинарным полем макросоциолингвистики, а также языкового права, психолингвистики, и предметной областью, включающей проблематику языковых контактов, языковых ситуаций, социолингвистической динамики, языковой политики и языкового планирования, социальных представлений о языке.

Цели и задачи журнала

- Развитие научного сотрудничества между российскими и зарубежными социолингвистами, а также специалистами смежных областей.
- Знакомство с новейшими российскими и международными исследованиями в области макросоциолингвистики.
- Публикация результатов оригинальных теоретических, прикладных и дескриптивных исследований в русле научной тематики журнала.
- Просветительство и распространение научного знания о различных языковых ситуациях для широкого круга читателей.

Специфика журнала

- Регулярные тематические выпуски с приглашенными российскими и зарубежными соредакторами.
- Постоянная рубрика, посвященная обзору научных работ «Макросоциолингвистика в России и мире».
- Постоянная просветительская рубрика «Просто о сложном», ориентированная на широкую читательскую аудиторию.
- Дискуссионная платформа для обсуждения вопросов, связанных с сохранением языков, языковой политикой и языковым планированием.
- Постоянная рубрика «Слово молодым».

Кроме исследовательских научных статей редколлегия принимает к рассмотрению научные обзоры, а также публикует рецензии на монографии, информацию о конференциях, научных проектах и т.д.

Перечень отраслей науки и групп специальностей научных работников в соответствии с номенклатурой ВАК РФ: Группа научных специальностей: 5.9. — Филология; 5.3. — Психология. Специальности научных работников: 5.9.8. — Теоретическая, прикладная и сравнительно-сопоставительная лингвистика, 5.9.5. — Русский язык. Языки народов России, 5.9.6. — Языки народов зарубежных стран, 5.3.1 — Общая психология, психология личности, история психологии.

Журнал строго придерживается международных стандартов публикационной этики, сформулированных в документе COPE (Committee on Publication Ethics): <http://publicationethics.org>

Правила оформления статей, архив и дополнительная информация размещены на сайте: <https://macrosociolinguistics.ru/MML/>

Электронный адрес: mmljrl@rudn.ru

Литературный редактор *О.В. Горячева*
Редактор англоязычных текстов *Ю.М. Алюнина*
Компьютерная верстка *М.В. Рогова*

Адрес редакции:

115419, Москва, Россия, ул. Орджоникидзе, д. 3
Тел.: +7 (495) 955-07-16; e-mail: publishing@rudn.ru

Почтовый адрес редакции

117198, Москва, Россия, ул. Миклухо-Маклая, д. 10/2
Тел.: +7 (495) 434-20-12; e-mail: mmljrl@rudn.ru

Подписано 19.12.2025. Выход в свет 22.12.2025. Формат 70×108/16.

Бумага офсетная. Печать офсетная. Гарнитуры Noto, Adelle.

Усл. печ. л. 11,55. Тираж 500 экз. Заказ № 781. Цена свободная.

Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы» (РУДН)

117198, Москва, Россия, ул. Миклухо-Маклая, д. 6

Отпечатано в типографии ИПК РУДН

115419, Москва, Россия, ул. Орджоникидзе, д. 3

Тел.: +7 (495) 955-08-61; publishing@rudn.ru



MACROSOIOLOGUISTICS AND MINORITY LANGUAGES

2025 Volume 3 № 2

DOI: 10.22363/2949-5997-2025-3-2

Published since 2023

by the Peoples' Friendship University of Russia named after Patrice Lumumba

EDITOR-IN-CHIEF

Dr. Svetlana A. Moskvitcheva

Institute of Modern Languages,
Intercultural Communication
and Migration (IM-LICaM),
RUDN University,
Moscow, Russian Federation

E-mail: moskvicheva-sa@rudn.ru

EXECUTIVE BOARD SECRETARY

Dr. Anna S. Borisova

RUDN University, *Moscow, Russian
Federation*

E-mail: borisova-as@rudn.ru

Honorary Editor Dr. Alain Viaut, French National Centre for Scientific Research (CNRS), Bordeaux, France,
e-mail: alain.viaut@orange.fr

Editorial board

Tatyana B. Agranat, Institute of Linguistics of the Russian Academy of Sciences, Moscow, Russian Federation

Oksana I. Aleksandrova, RUDN University, Moscow, Russian Federation

Leyli R. Dodykhudoeva, Institute of Linguistics of the Russian Academy of Sciences, Moscow, Russian Federation

Radif R. Zamaletdinov, Kazan Federal University, Kazan, Russian Federation

Olga P. Ilyukha, Institute of Linguistics, Literature and History of Karelian Research Centre, Russian Academy of Sciences,
Petrozavodsk, Russian Federation

Tatyana D. Martsinkovskaya, L.S. Vygotsky Institute of Psychology, Russian State University for the Humanities, Moscow,
Russian Federation

Marina I. Solnyshkina, Kazan Federal University, Kazan, Russian Federation

Yuliya N. Ebzeeva, RUDN University, Moscow, Russian Federation

MACROSOCIOLINGUISTICS AND MINORITY LANGUAGES

ISSN 2949-5997 (Online)

2 issues per year

Languages: Avar, Adyghe, Altaic, English, Bashkir, Buryat, Veps, Dargin, Ingush, Spanish, Italian, Kabardino-Circassian, Kalmyk, Karachay-Balkarian, Karelian, Komi, Crimean Tatar, Kumyk, Lak, Lezgin, Mari, Mordovian, Nenets, German, Ossetian, Russian, Tatar, Tuvan, Udmurt, French, Khakass, Chechen, Chuvash, Yakut.

AIMS AND SCOPE

Macrosociolinguistics and Minority Languages is an interdisciplinary peer-reviewed journal of international scope dealing with issues in sociolinguistics and related disciplines. The journal is international due to its editorial board, contributing authors and thematic focus of publications.

Focus Areas

The journal covers highly specialized research topics related to the disciplinary field of macrosociolinguistics, including linguistic law, psycholinguistics, issues of language contacts, language situations, sociolinguistic dynamics, language policy and language planning, social representations of language.

Aims and Scope

To promote expert and academic dialogue between Russian and international sociolinguists and specialists in other related areas of study.

- To disseminate the latest research in the field of macrosociolinguistics developed both in Russia and abroad.
- To publish results of original theoretical, applied and descriptive research within the journal focus areas.
- To advance knowledge of various language situations and educate a wide range of readers.

Distinctive Features of the Journal

- Regular guest-edited thematic issues.
- Regular section 'Macrosociolinguistics in Russia and abroad' dedicated to a survey of major works marking important theoretical and practical advances.
 - Regular section 'Let's Talk Science' intended for a general audience of readers.
 - Regular 'Notes and Discussion' section where views on important issues related to language preservation, language policy and language planning are exchanged.
 - regular section 'Call for Young Researches'.

In addition to research articles, the journal publishes editorials, book reviews, conference reports and research project announcements. The journal also considers responses and comments on manuscripts published in previous issues.

The Journal is published in accordance with the policies of COPE (Committee on Publication Ethics) <http://publicationethics.org>
The editors are open to thematic issue initiatives with guest editors.

Further information regarding notes for contributors, subscription, open access and back volumes is available at <https://macrosociolinguistics.ru/MML/>
E-mail: mmljrl@rudn.ru

Editor O.V. Goryacheva
English text editor Yu.M. Alyunina
Computer design M.V. Rogova

Address of the Editorial Board:

3 Ordzhonikidze St, 115419 Moscow, Russian Federation
Ph.: +7 (495) 952-04-41
e-mail: publishing@rudn.ru

Postal Address of the Editorial Board:

10/2 Miklukho-Maklaya St., Moscow, 117198, Russian Federation
Ph. +7 (495) 434-22-12; e-mail: mmljrl@rudn.ru

Printing run 500 copies. Open price

Peoples' Friendship University of Russia Named After Patrice Lumumba (RUDN University)
6 Miklukho-Maklaya St, 117198, Moscow, Russian Federation

Printed at RUDN Publishing House:

3 Ordzhonikidze St, 115419 Moscow, Russian Federation,
Ph.: +7 (495) 955-08-61; e-mail: publishing@rudn.ru

Содержание

Языки народов Российской Федерации: цифровые инструменты документирования и медиадоступность

- Кенешбекова Ш.Н.** Письменный корпус татарского языка: структура, состав и возможности использования 97
- Кириленко Е.** Электронный текстовый корпус лингвокультуры Северного Приангарья: структура, состав и возможности использования 117
- Пожидаев М.С., Теплых Е.С., Данилов С.И.** Производство аудиокниг на языках народов России с использованием синтезаторов речи: проблемы и перспективы 131

Язык и культура как символический капитал

- Alyunina Yu.M., Ganeeva E.R.** Social and cultural restrictions for movie titles translation: Why AI will never do it 146
- Кузнецова А.** Социолингвистические параметры образа СССР в романе У. Тевиса «The Queen's Gambit» и его экранизации: легитимация доминирования 166
- Роговец А.А.** Символический капитал и конструирование «другого»: цвет как маркер образа России в американском кинематографе 192

Научные события. Обзоры. Рецензии

- Сидорина И.Е., Сюзюмов А.А.** Геоинформационное картографирование в современных этнографических исследованиях 214

Contents

The Languages of the Peoples of Russian Federation: Digital Documentation Tools and Media Accessibility

- Keneshbekova Sh. N.** Corpus of written Tatar: Structure, composition and applications..... 97
- Kirilenko E.** Digital corpus of the linguoculture of the Northern Angara region: structure, composition and applications 117
- Pozhidaev M.S., Teplykh E.S., Danilov S.I.** Production of audiobooks in the languages of the peoples of Russia using speech synthesizers: problems and prospects..... 131

Language and Culture as a Form of Symbolic Capital

- Alyunina Yu.M., Ganeeva E.R.** Social and cultural restrictions for movie titles translation: Why AI will never do it 146
- Kuznetsova A.S.** Sociolinguistic Parameters of the image of the USSR in W. Tevis’s novel “The Queen’s Gambit” and its film adaptations: legitimization of dominance 166
- Rogovets A.A.** Symbolic capital and the construction of the “Other”: color as a marker of the image of Russia in American cinema 192

Scientific events. Briefs. Reviews

- Sidorina I.E., Siuziumov A.A.** GIS cartography in modern ethnography..... 214




**ЯЗЫКИ НАРОДОВ РОССИЙСКОЙ ФЕДЕРАЦИИ:
ЦИФРОВЫЕ ИНСТРУМЕНТЫ ДОКУМЕНТИРОВАНИЯ
И МЕДИАДОСТУПНОСТЬ**
**THE LANGUAGES OF THE PEOPLES OF RUSSIAN FEDERATION:
DIGITAL DOCUMENTATION TOOLS AND MEDIA ACCESSIBILITY**

DOI: 10.22363/2949-5997-2025-3-2-97-116

EDN HARJOQ

Научная статья / Research article

**Письменный корпус татарского языка:
структура, состав и возможности использования**

Шахдар Нурдиновна КЕНЕШБЕКОВА 

Российский университет дружбы народов, Москва, Российская Федерация

Российский новый университет, Москва, Российская Федерация

✉ keneshbekova.sh@yandex.ru

Аннотация. Исследование носит аналитический характер и посвящено изучению и описанию структуры, состава и функциональных возможностей Письменного корпуса татарского языка (Корпус) — одного из наиболее масштабных цифровых ресурсов для тюркских языков Российской Федерации. Рассмотрены этапы создания Корпуса, методология сбора и аннотирования текстов, включая метаразметку и морфологическую разметку с использованием системы *Apertium*. Особое внимание уделено прикладным аспектам: интеграции Корпуса в системы синтеза и распознавания речи, разработке различных лингвистических сервисов, а также его применению в образовательных и научных проектах. Проанализированы принципы устранения дублирующихся текстов и предложены перспективы дальнейшего развития, включая расширение жанрового разнообразия и внедрение международных стандартов аннотирования. Материалом исследования послужили сам Корпус и публикации, описывающие этапы его создания и применения. Методология представлена комплексом эмпирических методов и приемов, как наблюдение, анализ, описание, тестирование (функциональных возможностей корпуса и др.), а также графический метод для визуализации изучаемого материала. Отмечена научная и культурная значимость Письменного корпуса татарского языка в контексте цифровизации языков народов России, что соответствует задачам Международного десятилетия языков

© Кенешбекова Ш.Н., 2025

This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

коренных народов (2022–2032), инициированного Генеральной Ассамблей ООН и координируемого Организацией Объединенных Наций по вопросам образования, науки и культуры.

Ключевые слова: корпусная лингвистика, языки народов России, киберэтнография, цифровая гуманитаристика

Заявления о конфликте интересов. Автор заявляет об отсутствии конфликта интересов.

История статьи: получена 30 сентября 2025; принята в печать 14 октября 2025.

Для цитирования: Кенешбекова Ш.Н. Письменный корпус татарского языка: структура, состав и возможности использования // *Macrosociolinguistics and Minority Languages*. 2025. Т. 3. № 2. С. 97–116. <https://doi.org/10.22363/2949-5997-2025-3-2-97-116> EDN: HARJOQ

Corpus of written Tatar: structure, composition and applications

Shakhdar N. KENESHBEKOVA 

RUDN University, Moscow, Russian Federation
Russian New University, Moscow, Russian Federation
✉ keneshbekova.sh@yandex.ru

Abstract. The proposed study is analytical in nature and is devoted to examining and describing the structure, composition, and functional capabilities of the Written Corpus of the Tatar language, which represents one of the largest digital resources for the Turkic languages of the Russian Federation. The study discusses the stages of corpus creation and the methodology of text collection and annotation, including metadata annotation and morphological annotation using the *Apertium* system. Special attention is paid to applied aspects, such as the integration of the corpus into speech synthesis and speech recognition systems, the development of various linguistic services and its use in educational and research projects. The principles of eliminating duplicate texts are analyzed and prospects for further development are proposed, including the expansion of genre diversity and the introduction of international annotation standards. The material for this study comprises the corpus itself and publications describing the stages of its creation and application. The methodology is presented as a set of empirical methods and techniques, including observation, analysis, description, and testing (of the functional capabilities of the corpus, etc.), as well as the graphical method for visualizing the material under study. The study highlights the scientific and cultural significance of the Written Corpus of the Tatar Language in the context of the digitalization of the languages of the peoples of Russia, which corresponds to the objectives of the International Decade of Indigenous Languages (2022–2032), initiated by the United Nations General Assembly and coordinated by the United Nations Educational, Scientific and Cultural Organization.

Key words: corpus linguistics, languages of Russian Federation, Tatar language, cyberethnography, digital humanities

Conflict of interest. The author declares no conflicts of interest.

Article history: received 30 September 2025; accepted 14 October 2025.

For citation: Keneshbekova, Sh.N. (2025). Corpus of written Tatar: Structure, composition and applications. *Macrosociolinguistics and Minority Languages*, 3(2), 97–116. (In Russ.). <https://doi.org/110.22363/2949-5997-2025-3-2-97-116> EDN: HARJOQ

Введение

Разработка лингвистических корпусов для языков народов России занимает важное место в современной языковой политике и языковом планировании страны. Государство поддерживает подобные лингвистические инициативы, что подтверждается, например, Законом РФ «О языках народов Российской Федерации»¹ с изменениями от 13.06.2023 № 253-ФЗ, а также на Распоряжение Правительства Российской Федерации от 12.06.2024 № 1481-р². Первый признает языки народов России (ЯНР) национальным достоянием государства, определяя их место в системе образования, средствах массовой информации, производстве и других областях деятельности. Второй — определяет языковое многообразие РФ совокупностью «функционирующих языков, диалектов и говоров народов Российской Федерации»³, утверждая, что республики страны «наделены правом устанавливать свои государственные языки, функционирующие наряду с государственным»⁴. Регламентируя области функционирования ЯНР, данное Распоряжение обращает внимание на «содействие изданию литературы на языках народов Российской Федерации, <...> переводческой деятельности, финансирование научных исследований в области сохранения, изучения и развития языков народов Российской Федерации, создание условий для распространения через средства массовой информации сообщений и материалов на языках народов Российской Федерации, подготовку специалистов в указанной области, совершенствование системы образования в целях развития языков народов Российской Федерации»⁵. В таком контексте, с точки зрения государственной поддержки, разработка, развитие и научно-образовательная популяризация корпусов ЯНР видится актуальным и востребованным направлением деятельности современной фундаментальной и прикладной лингвистики.

¹ Закон РФ «О языках народов Российской Федерации» от 25.10.1001 № 1807-1 с изменениями от 13.06.2023 № 253-ФЗ. URL: https://www.consultant.ru/document/cons_doc_LAW_15524/ (дата обращения: 15.08.2025)

² Распоряжение Правительства Российской Федерации от 12.06.2024 № 1481-р. URL: <http://publication.pravo.gov.ru/document/0001202406140048?index=1> (дата обращения: 15.08.2025).

³ Распоряжение Правительства Российской Федерации от 12.06.2024 № 1481-р. С. 3. URL: <http://publication.pravo.gov.ru/document/0001202406140048?index=1> (дата обращения: 15.08.2025).

⁴ Распоряжение Правительства Российской Федерации от 12.06.2024 № 1481-р. С. 4. URL: <http://publication.pravo.gov.ru/document/0001202406140048?index=1> (дата обращения: 15.08.2025).

⁵ Распоряжение Правительства Российской Федерации от 12.06.2024 № 1481-р. С. 6. URL: <http://publication.pravo.gov.ru/document/0001202406140048?index=1> (дата обращения: 15.08.2025).

В собственно научной перспективе разработка корпусов для ЯНР встраивается в область интересов киберэтнографии в аспекте фиксации, упорядочивания и описания представленности того или иного языка в веб-среде. Эти задачи обусловлены большей «ориентированностью» киберэтнографии «на тексты <...>, а не на наблюдение и взаимодействие» с представителями изучаемого народа (Белорусова, 2021: 127).

Корпусные технологии позволяют не только сохранять и изучать языковое разнообразие, но и открывают новые возможности для цифровой обработки языка, машинного перевода, автоматического анализа текстов и других прикладных задач (Алюнина, 2025: 102–103). Так разработка лингвистических корпусов, особенно параллельных (Алюнина, 2025: 103–104), соответствует вектору развития программ машинного перевода, которые с начала 2020-х гг. активно пополняются ЯНР: *Переводчик Яндекс*⁶ (бурятский, коми, мокшанский, татарский, удмуртский, чувашский и др.), *Google Translate*⁷ (бурятский, коми, осетинский, татарский и др.), *Microsoft Bing*⁸ (башкирский, татарский и др.), *PROMT.One*⁹ (татарский) и др. Особую значимость деятельность по разработке подобных веб-ресурсов приобретает в условиях глобализации, когда многие языки, и не только миноритарные, сталкиваются с угрозой цифровой маргинализации — вытеснения из электронного пространства «языками-первопроходцами», такими как английский, китайский, русский, арабский, испанский, итальянский, немецкий, французский, итальянский, о чем свидетельствуют статистические данные разных лет, начиная с конца 1990-х до начала 2020-х гг.^{10, 11}

На сегодняшний день в России ведется работа по созданию корпусов для многих ЯНР, что способствует их документированию и интеграции в цифровую среду. Среди примеров можно назвать следующие проекты:

- Бурятский корпус¹² — проект Института монголоведения, буддологии и тибетологии Сибирского отделения РАН при участии специалистов Института востоковедения РАН, Лионского университета во Франции, Университета Гамбурга в Германии;

⁶ Переводчик Яндекс. URL: <https://translate.yandex.ru/translator/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9-%D0%A0%D1%83%D1%81%D1%81%D0%BA%D0%B8%D0%B9> (дата обращения: 16.07.2025)

⁷ Google Translate. URL: <https://translate.google.com/?hl=ru&tab=TT&sl=ru&tl=en&op=translate> (дата обращения: 16.07.2025).

⁸ Microsoft Bing. URL: <https://www.bing.com/translator?from=&to=ru&setlang=ru> (дата обращения: 16.07.2025).

⁹ PROMT.One. URL: <https://www.translate.ru/%D0%BF%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4> (дата обращения: 16.07.2025).

¹⁰ Usage statistics of content languages for websites. URL: https://w3techs.com/technologies/overview/content_language (дата обращения: 20.07.2025).

¹¹ Pimienta D., Prado D., Blanco A. Twelve years of measuring linguistic diversity in the Internet: balance and perspectives. Paris : UNESCO Publications for the World Summit on the Information Society, 2009. 58 p. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000187016> (дата обращения: 20.07.2025).

¹² Бурятский корпус. URL: http://web-corpora.net/BuryatCorpus/search/?interface_language=ru (дата обращения: 16.07.2025).

- Открытый корпус вепсского и карельского языков¹³ — разрабатывается сотрудниками Карельского научного центра РАН (Родионова, Пеллинен, 2024);
- Корпус мансийского языка¹⁴ — проект Института языкознания РАН при участии Института русского языка им. В.В. Виноградова, МГУ им. М.В. Ломоносова, Новгородского государственного университета им. Ярослава Мудрого, Обско-угорского института прикладных исследований и разработок);
- Корпус удмуртского языка¹⁵ (основной корпус, корпус соцсетей, звуковой корпус) — проект Школы лингвистики Научно-исследовательского университета «Высшая школа экономики» (далее — ВШЭ);
- Электронный корпус чувашского языка¹⁶ — разрабатывается Чувашским государственным институтом гуманитарных наук;
- Татарский национальный корпус «Туган тел»¹⁷ — проект, реализуемый Научно-исследовательским институтом «Прикладная семиотика» Академии наук Республики Татарстан, Казанским (Приволжским) федеральным университетом и ВШЭ;
- Письменный корпус татарского языка¹⁸ — проект Университета Иннополис в России, Университета Квебека в Канаде, Штутгартского университета в Германии, Университета Турку в Финляндии и др.

Среди существующих сегодня корпусных инициатив Письменный корпус татарского языка (ПКТЯ) является одним из наиболее развитых корпусов тюркских языков в России. Мы поставили цель: комплексно рассмотреть структуру, состав и функциональные возможности ПКТЯ и проанализировать его роль в современных лингвистических и прикладных исследованиях. Материалом для анализа служит ПКТЯ, включающий тексты различных жанров и периодов, а также его техническая и лингвистическая разметка.

Структура статьи воспроизводит последовательный анализ следующих аспектов ПКТЯ: 1) историю разработки и состав корпуса; 2) функциональные возможности, поисковый и аналитический потенциал; 3) применение в фундаментальных и прикладных исследованиях; 4) а также перспективы развития и использования функционала и наполнения ПКТЯ, что позволяет оценить его вклад в поддержку татарского языка, а также роль корпуса в развитии современных проектов в области цифровой гуманитаристики (Digital Humanities).

¹³ Открытый корпус вепсского и карельского языков. URL: <http://dictorpus.krc.karelia.ru/ru> (дата обращения: 16.07.2025).

¹⁴ Аннотированный корпус мансийских текстов. URL: <https://mansi.pro/corpus/> (дата обращения: 16.07.2025).

¹⁵ Корпуса удмуртского языка. URL: http://web-corpora.net/UdmurtCorpus/search/?interface_language=ru (дата обращения: 16.07.2025).

¹⁶ Электронный корпус чувашского языка. URL: <https://chuvkorpus.ru/> (дата обращения: 17.07.2025).

¹⁷ Татарский национальный корпус «Туган тел». URL: http://web-corpora.net/TatarCorpus/search/?interface_language=ru (дата обращения: 16.07.2025).

¹⁸ Письменный корпус татарского языка. URL: <https://www.corpus.tatar/> (дата обращения: 16.07.2025).

История разработки Письменного корпуса татарского языка и его состав

Инициатива создания ПКТЯ принадлежит коллективу исследователей, включающему специалистов из России, Финляндии, Германии и Канады, которые начали работу над проектом в 2010 г. Партнерами проекта стали редакция научно-информационного журнала «Фэн һәм Тел» — «Наука и язык»¹⁹, Республиканская специальная библиотека для слепых и слабовидящих в Татарстане²⁰, Институт языка, литературы и искусства имени Г.Ибрагимова Академии наук Республики Татарстан²¹, разработчики корпусного менеджера *Sketch Engine*²² и др. Таким образом, команда ПКТЯ представлена филологами, лингвистами, инженерами, специалистами сферы информационной и медиадоступности, что обеспечивало междисциплинарный подход к разработке рассматриваемого лингвистического ресурса.

Первый этап создания ПКТЯ включал проектирование его архитектуры, разработку поискового механизма и определение принципов выборки текстов. К марту 2012 г. была завершена базовая версия корпуса, включающая веб-интерфейс и систему поиска. Первая публичная версия корпуса, содержащая 45 млн словоупотреблений из 60 источников различных жанров и стилей, стала доступна в начале 2012 г. (Сайхунов, Хусаинов, Ибрагимов, 2018: 314).

В 2014 г. выпущена вторая версия корпуса, объем данных которого увеличился до 116 млн словоупотреблений из 2750 источников (Ибрагимов, Сайхунов, 2014: 261). Под версией или релизом корпуса в данном случае понимается обновление его текстовой базы. Для расширения функциональных возможностей ПКТЯ в период между 2014 и 2018 гг. в корпус была внедрена морфологическая разметка, в основу которой легла «система автоматической грамматической аннотации» (Сайхунов, Хусаинов, Ибрагимов, 2018: 314), разработанная *Apertium*²³ и адаптированная к большому количеству языков, в т.ч. к татарскому. Жанрово-стилистический состав корпуса в его второй итерации приведен на рис. 1.

После релиза второй версии ПКТЯ корпус был дополнен возможностью проверки правописания на татарском языке (2017 г.) и функцией синтеза татарской речи (2015 г.), в разработке которой участвовали специалисты Республиканской специальной библиотеки для слепых и слабовидящих²⁴.

¹⁹ «Фэн һәм тел» — «Наука и язык»: научно-информационный журнал. URL: <https://tatarica.org/ru/razdely/sredstva-massovoj-informacii/periodicheskie-izdaniya/fehn-hehm-tel> (дата обращения: 18.07.2025).

²⁰ ГБУК Республики Татарстан «Республиканская специальная библиотека для слепых и слабовидящих имени Ш.К. Еникеева». URL: <https://rsbrst.ru/> (дата обращения: 18.07.2025).

²¹ Институт языка, литературы и искусства имени Г. Ибрагимова Академии наук Республики Татарстан <https://www.antat.ru/ru/iyli/> (дата обращения: 19.07.2025).

²² Sketch Engine. URL: <https://www.sketchengine.eu/> (дата обращения: 19.07.2025).

²³ Apertium. URL: <https://www.apertium.org/index.rus.html#?dir=rus-bel&q=> (дата обращения: 17.08.2025).

²⁴ ГБУК Республики Татарстан «Республиканская специальная библиотека для слепых и слабовидящих имени Ш.К. Еникеева». URL: <https://rsbrst.ru/> (дата обращения: 18.07.2025).



Рис. 1. Жанрово-стилевой состав ПКТЯ по состоянию на 2014 г.

Источник: выполнила Ш.Н. Кенешбекова по материалам (Ибрагимов, Сайхуннов, 2014: 261).

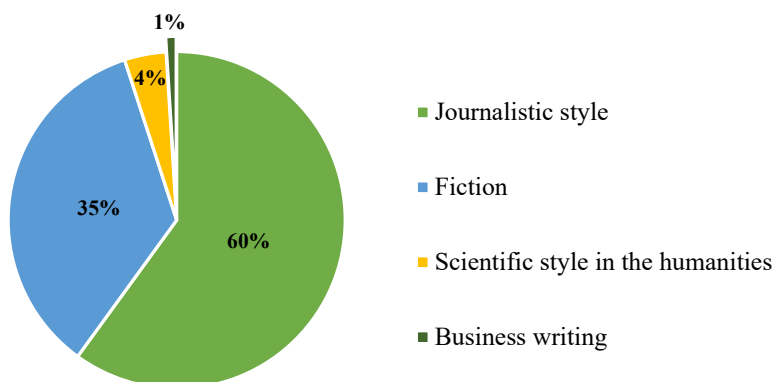


Fig. 1. Genre and style composition of the Corpus of Written Tatar as of 2014

Source: compiled by Sh.N. Keneshbekova after (Ibragimov, Saikhunov, 2014: 261).

Третья версия корпуса, выпущенная в конце 2018 г., почти втрое превысила объем предыдущей, — 356 млн слов (430 млн токенов) из 16 786 текстов (Сайхуннов, Хусаинов, Ибрагимов, 2019: 548); перечень функциональных возможностей пополнился разделом Тезаурус.

Последнее, четвертое обновление ПКТЯ, состоявшееся в конце 2019 г., увеличило его объем до 500 млн словоупотреблений (17 000 источников) и привнесло в корпус обновленную систему проверки правописания для татарского языка, улучшенную морфологическую разметку. Позже, в 2022 г., у ПКТЯ появился специализированный раздел «Личные имена», включающий перечень татарских антропонимов — имен, отчеств и фамилий. Пример результата выполнения поискового запроса по татарскому личному имени в ПКТЯ приведен на изображении далее (рис. 2).

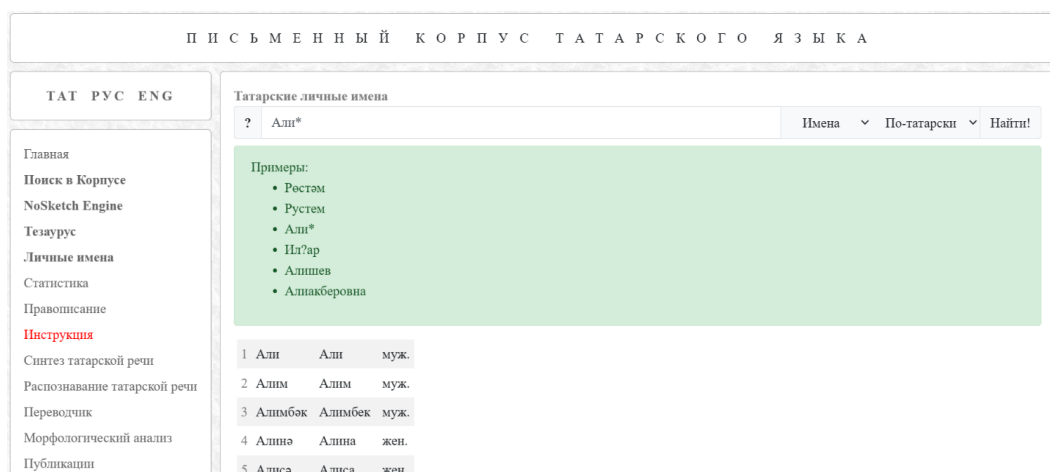


Рис. 2. Скриншот с результатом поиска татарского имени

Источник: Письменный корпус татарского языка. URL:

<https://www.corpus.tatar/index.php?of=search/names.php> (дата обращения: 12.08.2025).

Fig. 2. Screenshot with the search result for a Tatar name

Source: Corpus of written Tatar. Retrieved 12 August 2025, from:

<https://www.corpus.tatar/index.php?of=search/names.php>

Таким образом, основной целью разработки ПКТЯ стало создание инструментария для цифровой фиксации татарского языка и систематизации процесса его изучения. Также корпус может быть использован для преподавания татарского языка, разработки методических материалов, пополнения обучающих баз машинных переводчиков в языковой паре с татарским языком, автоматического синтеза и распознавания татарской речи, что делает вклад в развитие доступной среды. Помимо этого, корпус выполняет функцию важного ресурса для лингвистических исследований, предоставляя доступ к обширной и репрезентативной текстовой базе.

Процесс разработки ПКТЯ сопряжен с рядом технологических и организационных вызовов, которые включают отсутствие централизованного финансирования, необходимость адаптации программных решений под специфику татарского языка, а также привлечение экспертов в области корпусной лингвистики. Интерфейс корпуса доступен на трех языках (татарском, русском и английском), что способствует его интеграции в международное научное сообщество. Вместе с тем рост объема корпусов сопряжен с возрастающей сложностью в обеспечении уникальности содержащихся в них текстовых единиц. В процессе корпусной компиляции используются различные методологические подходы к устранению дублирования, основанные на лингвистическом анализе и алгоритмических решениях.

Функциональные возможности Письменного корпуса татарского языка

Письменный корпус татарского языка представляет собой универсальную цифровую платформу, предназначенную для комплексного анализа структуры и функционирования татарского языка. Корпус широко используется как в академических целях, так и для решения прикладных задач: разработка образовательных и методических материалов, лексикографических ресурсов разных типов и технологий автоматической обработки текста и речи, подготовка лингвистически аннотированных ресурсов для обучения и NLP-систем и др. (Кузнецов, 2023; Гатиатуллин и др., 2024; Galieva, Vavilova, Gafarova, 2017; Luutonen, Moisio, Daher, 2017).

Встроенные в корпус программные средства обеспечивают широкий спектр функциональных возможностей, охватывающих:

- поиск и количественный анализ лексических единиц (реализуется функцией Статистика);
- анализ лексической и грамматической сочетаемости, т.е. поиск типичных для языка коллокаций и коллигаций (реализуется, например, функциями n-грамм, KWIC — Key word in context — ключевое слово в контексте);
- поиск по морфологическим, семантическим и стилистическим признакам (реализуется с помощью настройки поискового запроса в интерфейсе ПКТЯ);
- проверку орфографии на татарском языке (функция Правописание);
- озвучивание письменного текста на татарском языке и транскрибирование татарской речи (функции Синтез татарской речи и Распознавание татарской речи).

Приведенные функции позволяют решать широкий спектр задач, а за их реализацией и достоверностью прогнозируемых результатов поиска стоит строгая система корпусного аннотирования или корпусной разметки, под которой понимается лингвистическая информация, «приписываемая всем единицам выбранного уровня: текст, предложение, словосочетание, словоформа» и др. (Алюнина, 2025: 96). В зависимости от степени обращения к программным ресурсам в аннотировании корпусных документов выделяют ручную и автоматическую разметку (Алюнина, 2025: 98).

Метаразметка и классификация собранных текстов

С технической точки зрения корпус создается с применением методов автоматической обработки текстов, включая морфологическую разметку и поисковые алгоритмы. Однако значительная часть работы, такая как верификация данных и категоризация текстов, выполняется вручную. Как было

показано на рис. 1, источниками для пополнения корпуса служат письменные тексты различного происхождения: художественная литература, публицистика, научные и официальные документы.

Метаразметка в составе ПКТЯ представляет собой систему аннотирования текстов с использованием описательных характеристик, которые обеспечивают упорядоченное хранение, систематизированный поиск и последующую аналитическую обработку данных.

В международной корпусной практике в качестве методологических ориентиров широко применяются стандарты *Text Encoding Initiative — Стандарт кодирования текстовых данных (TEI)*²⁵ и рекомендации *Expert Advisory Group on Language Engineering Standards — Экспертной группы по стандартизации структуры лингвистических ресурсов (EAGLES)*²⁶, на базе которых разработан *Corpus Encoding Standard — Перечень рекомендаций корпусной разметки (CES)*²⁷. Эти стандарты обеспечивают единообразие в представлении метаданных. Как правило, они реализуются в формате XML, что способствует совместимости корпусных ресурсов и их интероперабельности на межплатформенном уровне (Сайхунов, Хусаинов, Ибрагимов, 2019: 549).

В отличие от названных инструментов, ПКТЯ использует собственную, адаптированную к его специфике систему метаразметки. Формально она реализована в виде так называемых plain text-структур, не прибегающих к полноформатной XML-разметке. Подобное решение обусловлено рядом практических причин: текстовые файлы в «плоском» формате легче обрабатываются базовыми утилитами UNIX-подобных систем (такими как awk, sed, grep, sort, cut, paste и др.)²⁸, обладают высокой степенью читаемости и не требуют специализированного программного обеспечения для начального этапа анализа (Сайхунов, Хусаинов, Ибрагимов, 2019: 550).

Несмотря на функциональность существующей модели, дальнейшая эволюция корпуса предполагает постепенное внедрение унифицированных международных стандартов метаразметки. Переход к XML-структурам и расширенным схемам аннотации позволит достичь следующих целей:

- интеграции корпуса в многоязычные исследовательские платформы;
- повышения точности и детализации лингвистических исследований;
- обеспечения прозрачного обмена данными между научными учреждениями и проектами.

На скриншоте далее представлен пример интерфейса, где демонстрируется структура метаразметки текста в ПКТЯ (рис. 3).

²⁵ Text Encoding Initiative. URL: <https://tei-c.org/> (дата обращения: 23.08.2025).

²⁶ Expert Advisory Group on Language Engineering Standards. URL: <https://cordis.europa.eu/project/id/LE34244> (дата обращения: 23.08.2025).

²⁷ Corpus Encoding Standard. URL: <https://www.cs.vassar.edu/CES/> (дата обращения: 23.08.2025).

²⁸ 13 инструментов для обработки текста в командной оболочке // Хабр. URL: <https://habr.com/ru/companies/itsumma/articles/492932/> (дата обращения: 18.08.2025).

Количество совпадений: 153001

[Хасанов А.Б., Жәңбәт җәвәре: повестьлар, хикәяләр, публицистик язмалар - Нәшрият: Татарстан китап нәшрияты, 2007 ел.]

Мәскәүгә китап чыгарырга барган саен шунда кунак булам.

Абдулла Алиш. Кечкенә тоткын / Пьеса

Шул вакыт звонок тавышы ишетелә, Роза барып ишек ача, ишектән Ганс кайтып кәра, жылкәсендә китап сумкасы.

Ганс (китап сумкасын өстәлгә куеп).

Котлет пешерергә өйрәткән китап.

Абдулла Алиш. Куршәләр / Пьеса

Н у р ы й. Шулай булуну китап мактый.

Абдулла Алиш. Якты күл бие / Повесть (1931-1932)

- Син, егет, китап укый - укый бозылган.

Китап бит ул, иптәш, кешене бик боза_торган нәрсә.

Адлер Тимергалин. Айга ашкыну һәм "Алга! Марска таба!" / Публицистик язмалар һәм мәкаләләр

Өстәлемдә кул ясауы кадр_генә зурлыктагы гажәеп бер китап ята.

Алгы титулда бу русча китап "Айга очу" ("Полёт на Луну") дип атала һәм аның Мәскәүдә ярты гасыр элек - 1956 елда басылып чыгуы күрсәтелгән.

Чөнки бу китап, эгәр әйтергә яраса, "фәнни - фантастик публицистика" жанрына карый.

Рис. 3. Метаданные в Письменном корпусе татарского языка

Источник: Письменный корпус татарского языка.
URL: <https://www.corpus.tatar/ru> (дата обращения: 12.08.2025).

Fig. 3. Metadata in the Corpus of Written Tatar

Source: Corpus of written Tatar. Retrieved 12 August 2025,
from: <https://www.corpus.tatar/index.php?of=search/names.php>

Как показано на рис. 3, на текущем (по состоянию на август 2025 г.) этапе в корпус включены следующие параметры экстралингвистической метаразметки, которые обеспечивают базовую классификацию текстового материала (Сайхунов, Хусаинов, Ибрагимов, 2019: 550; Алюнина, 2025: 100):

- авторство — указание имени автора или авторского коллектива;
- название — заголовок произведения, статьи, книги, либо интернет-источника;
- хронологическая метка — сведения о времени создания или публикации текста;
- типологическая характеристика — сведения о стилистической принадлежности текста (художественный, публицистический, научный, официально-деловой, фольклорный и т.д.);
- жанровая принадлежность — сведения о жанровых характеристиках текста (роман, рассказ, статья, поэма, сказание и др.);
- источник — библиографическая ссылка на издание или цифровую платформу;
- URL-адрес — сведения об электронном источнике, которые указываются в случае включения в корпус текста с цифрового ресурса;

- дополнительные атрибуты — техническая или организационная информация (кодировка, версия разметки, ID источника, поставщик данных и пр.).

Морфологическая разметка

Под морфологической разметкой лингвистического корпуса понимается присвоение «каждой словоформе ее морфологических признаков» (Алюнина, 2025: 99). С утилитарной точки зрения этот тип разметки обеспечивает возможность поиска по морфологическим параметрам, например, поиск существительного мужского рода в форме множественного числа. В процессе морфологического аннотирования ПКТЯ применялся ресурс *Apertium*²⁹ — свободное программное обеспечение с полностью открытым исходным кодом, широко используемое в корпусной лингвистике благодаря своей модульности и адаптивности. *Apertium* работает как открытая платформа для машинного перевода. В числе поддерживаемых — языков России: татарский, крымскотатарский и русский.

Автоматическая обработка естественного языка неизбежно порождает множественные варианты интерпретации одной и той же лексической единицы, что связано с феноменом грамматической омонимии. Разрешение такой неоднозначности представляет собой одну из ключевых задач компьютерной лингвистики. Ввиду колоссального объема современных корпусов ручное устранение омонимии практически неосуществимо, что стимулирует разработку автоматизированных методов обработки. В этом направлении ведутся интенсивные исследования, ориентированные на создание специализированных программных комплексов, использующих предопределенные правила, статистические алгоритмы или предобученные нейронные модели для повышения точности разметки (Сайхунов, Хусаинов, Ибрагимов, 2017: 382).

Морфологический анализатор *Apertium*, интегрированный в ПКТЯ, обладает встроенным механизмом разрешения определенных типов грамматической неоднозначности на основе системы формализованных правил. Дополнительным преимуществом данного программного решения является использование унифицированного набора тегов, применимого к широкому спектру языков, особенно относящихся к одной языковой группе. Это способствует созданию корпусов со схожей структурой морфологической аннотации, что значительно упрощает проведение межъязыковых сопоставительных исследований и сравнительного анализа лексико-грамматических закономерностей.

²⁹ Apertium. URL: <https://www.apertium.org/index.rus.html#?dir=rus-bel&q=> (дата обращения: 17.08.2025).

Дополнительные сервисы

Как правило, при разработке определенного корпуса преследуется решение конкретных задач фундаментального или практического плана. Однако, наблюдения, сделанные в процессе изучения ПКТЯ, указывают на то, что структурированные корпусные данные могут быть адаптированы для решения весьма широкого спектра проблем. Наличие масштабных аннотированных массивов текстовой информации позволяет авторам ПКТЯ интегрировать его ресурсы в различные технологические разработки и лингвистические приложения. Ниже представлены некоторые направления практического приложения рассматриваемого корпусного инструмента.

1. Онлайн-система проверки правописания, основанная на корпусных данных, позволяющая автоматизировать процесс орфографического контроля и корректировки ошибок при обучении татарскому языку или при иной проверке текстов на татарском языке.

2. Автоматическая генерация разнообразных статистических данных, включая частотные распределения букв, биграмм, слов, лемм и n-грамм, что предоставляет ценные сведения для количественного анализа закономерностей в татарском языке.

3. Создание тезауруса на основе дистрибутивной семантики, включающего векторные представления лексических единиц, сформированные с использованием технологий машинного обучения, таких как word2vec. В данном разделе разработаны 3 типа поиска: а) семантически схожие лексем; б) аналогия лексем; в) сходства двух лексем по тем или иным признакам. Пример аналогии приведен на рис. 4.

Семантическое подобие слов в татарском языке

?	«Эт - хайван, карга -»	Аналогия слов	Найти!
Cosine similarity:	0.40039	ерткыч_кош	0.376883 кыргый_жэнлек
0.529546	кош	0.394989 дунгыз_асрау	0.376299 мэхлукат
0.452673	жэнлек_кош_корт	0.394723 бүдэнә	0.374626 каргыш
0.45135	хайван_кош_корт	0.388843 кыргый_хайван	0.371003 жэнлек_хайван
0.424303	кош_оя	0.387124 кош_жэнлек	0.368932 хайван_үсемлек
0.417403	күз_кар_чүкы	0.384505 чыпчык	0.368489 кош_корт
0.417002	киек_кош	0.384216 жэнлек	0.366938 үрдәк_каз
0.415236	кош_корт_хайван	0.382977 бытбылдык	0.365372 сьерчык
0.413669	жэнлек_жанвар	0.380001 күчмә_кош	0.365242 мифик_зат
0.408859	бөжәк	0.379483 карр_карр	0.363836 кош_корт_жэнлек
0.407776	тереклек_ия	0.379233 зарарлы_бөжәк	0.361987 хайван_бөжәк
0.403236	жэнлек_кош	0.379104 ерткыч_хайван	0.361716 үсемлек

Рис. 4. Тезаурус. Пример типа поиска «аналогия слов»

Источник: Письменный корпус татарского языка.

URL: <https://www.corpus.tatar/index.php?of=search/names.php> (дата обращения: 12.08.2025).

Fig. 4. Thesaurus. Example of the search type ‘Word analogy’

Source: Corpus of written Tatar.

Retrieved 12 August 2025, from: <https://www.corpus.tatar/index.php?of=search/names.php>

4. Разработка системы синтеза татарской речи, предназначенной для создания различных тифлоинструментов и реализованной в Республиканской специальной библиотеке для слепых и слабовидящих³⁰ (рис. 5).

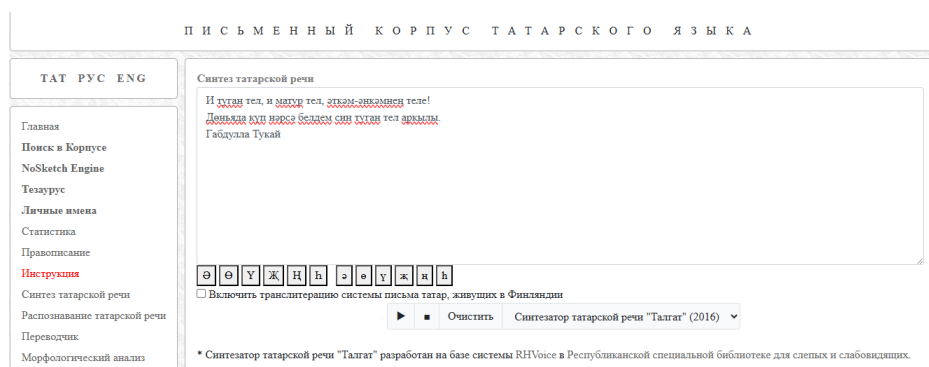


Рис. 5. Синтез татарской речи

Источник: Письменный корпус татарского языка.

URL: <https://www.corpus.tatar/index.php?of=search/names.php> (дата обращения: 12.08.2025).

Fig. 5. Tatar Text-To-Speech

Source: Corpus of written Tatar.

Retrieved 12 August 2025, from: <https://www.corpus.tatar/index.php?of=search/names.php>

5. Интеграция корпуса в систему распознавания татарской речи *Common Voice*³¹, способствующей совершенствованию технологий автоматической транскрипции и взаимодействия с голосовыми интерфейсами (рис. 6).

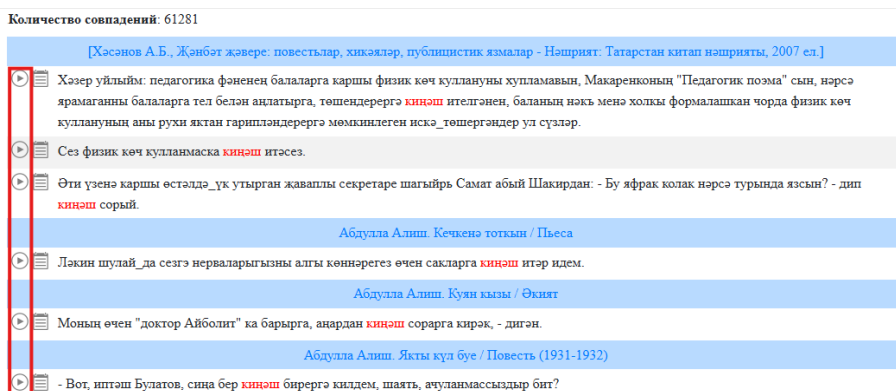


Рис. 6. Голосовой интерфейс

Источник: Письменный корпус татарского языка. URL: <https://www.corpus.tatar/index.php?of=search/names.php> (дата обращения: 12.08.2025).

Fig. 6. Voice interface

Source: Corpus of written Tatar. Retrieved 12 August 2025, from: <https://www.corpus.tatar/index.php?of=search/names.php>

³⁰ ГБУК Республики Татарстан «Республиканская специальная библиотека для слепых и слабовидящих имени Ш.К. Еникеева». URL: <https://rsbsrt.ru/> (дата обращения: 18.07.2025).

³¹ Common Voice. URL: <https://commonvoice.mozilla.org/tt> (дата обращения: 19.07.2025).

6. Оценка покрываемости морфологического анализатора *Apertium*, включающая выявление нерегулярных словоформ, автоматическое пополнение словарных ресурсов, а также оптимизацию системы правил для обучения технологий анализа естественного языка.

Некоторые из перечисленных проектов реализованы в формате веб-сервисов, что способствует расширению пользовательской аудитории и повышает прикладную значимость корпусных исследований, способствуя развитию цифровых лингвистических технологий.

Вклад Письменного корпуса татарского языка в фундаментальные и прикладные исследования

Письменный корпус татарского языка представляет собой не только технически значимый цифровой ресурс, но и платформу, имеющую фундаментальную ценность, которая активно используется в широком спектре лингвистических и междисциплинарных исследований.

На материале корпуса выполняются исследования в следующих направлениях:

- морфологический анализ и автоматическая обработка текста, включая разработку алгоритмов морфологического разборщика и анализ продуктивных словообразовательных моделей (Сайхунов, Хусаинов, Ибрагимов, 2018);
- изучение лексики и грамматики татарского языка, включая частотный анализ, синтаксические конструкции, словообразовательные модели, синонимию и фонологию (Ибрагимов, Сэйхунов, 2016; Ибрагимова, 2020; Кузнецов, 2023; Galieva, Vavilova, Gafarova, 2017; Luutonen, Moisio, Daher, 2017);
- исследование татарского фольклора и литературы, охватывающее жанровую эволюцию, особенности художественной речи и национально-культурные коды (Сайхунов, Ибрагимов, Галиуллин, 2018);
- социолингвистические исследования, посвященные анализу языковой ситуации, сохранению и трансформации татарской идентичности, а также вопросам языковой политики в условиях многоязычия (Ибрагимов, Сайхунов, 2015);
- цифровые и прикладные проекты, включая участие в создании систем распознавания и синтеза татарской речи на базе корпуса (Сайхунов, 2010);
- методология корпусной лингвистики, охватывающая вопросы создания, структурирования и использования национальных корпусов, в т.ч. на ЯНР (Ибрагимов, Сайхунов, Салимзянов, 2012; Гатиатуллин, Мухамедшин, Прокопьев, Сулейманов, 2024: 543).

Научная и практическая значимость названных исследований заключается в том, что они обеспечивают: 1) формирование репрезентативной эмпирической

базы для описания и нормализации татарского языка; 2) развитие технологий автоматической обработки тюркских языков, ранее недостаточно представленных в цифровом пространстве; 3) поддержку процессов цифровизации языков народов России; 4) интеграцию татарского языка в международные исследовательские проекты в области компьютерной лингвистики.

Перспективы развития и использования Письменного корпуса татарского языка

Развитие ПКТЯ имеет стратегическое значение как для лингвистической науки, так и для государственной языковой политики, направленной на сохранение и цифровую поддержку языков народов России. С учетом вызовов времени и международных инициатив в сфере языкового многообразия дальнейшая работа над корпусом открывает широкие перспективы.

Основные направления развития корпуса могут включать:

- расширение текстовой базы, включая добавление произведений современной художественной литературы, научных публикаций в разных областях знаний, публицистики и официальных документов, что повысит тематическую и стилевую репрезентативность корпуса;
- развитие лингвистической разметки, в частности — совершенствование морфологического и синтаксического анализа, а также внедрение элементов семантической и прагматической аннотации;
- создание специализированных подкорпусов, отражающих территориальные диалекты, возрастные и гендерные особенности, профессиональные и жанровые различия употребления языка;
- разработка электронных лингвистических ресурсов, таких как частотные словари, тезаурусы и обучающие базы, предназначенные для преподавания татарского языка и автоматической обработки текстов (Ибрагимов, Сайхунов, 2018: 73).

Перспективы применения корпуса в будущем видятся в следующих плоскостях:

- цифровизация образования, включая интеграцию корпуса в онлайн-платформы для обучения татарскому языку;
- машинный перевод и распознавание речи, где корпус будет служить основой для обучения моделей искусственного интеллекта;
- лексикографические и грамматические исследования, способствующие созданию новых нормативных справочников татарского языка;
- поддержка культурной политики, включая анализ языковой динамики и адаптацию языка к условиям цифровой среды;
- проектная деятельность в области Digital Humanities, ориентированная на разработку мультимодальных и межъязыковых ресурсов.

Деятельность по развитию ПКТЯ полностью соответствует задачам, поставленным в рамках Международного десятилетия языков коренных народов (2022–2032)³², инициированного Генеральной Ассамблей ООН и координируемого ЮНЕСКО. Инициатива направлена на сохранение, развитие и распространение языков, находящихся, в т. ч. под угрозой исчезновения.

Заключение

ПКТЯ представляет собой масштабный лингвистический ресурс, структурированный в соответствии с современными требованиями, стандартами и традициями корпусной лингвистики. Он включает тексты различных жанров, стилей и исторических периодов, охватывая широкий спектр письменной речевой практики. Корпус аннотирован с применением морфологического анализатора *Apertium*, что обеспечивает высокую точность разметки и возможность формального анализа грамматических структур.

Организационная система корпуса обеспечивается метаразметкой по ключевым экстралингвистическим параметрам: авторство, дата, жанровая и стилистическая принадлежность, источник и формат публикации. Несмотря на использование собственного формата представления данных, структура корпуса позволяет эффективно решать широкий круг исследовательских задач в области лексикографии, грамматического анализа, исторической лингвистики, межъязыкового сопоставления и обработки естественного языка.

Благодаря открытой архитектуре и машиночитаемому формату ПКТЯ интегрирован в ряд прикладных сервисов: автоматическую орфографическую проверку, генерацию статистических параметров, построение тезаурусов на основе *word2vec*, синтез и распознавание татарской речи. Эти разработки способствуют цифровизации татарского языка и его активному использованию в новых технологических контекстах.

Корпус продолжает развиваться. Планируется расширение жанрового охвата, улучшение морфологической аннотации, переход к стандартизированным форматам хранения и создание подкорпусов для региональных вариантов языка. Вектор развития согласуется с международной повесткой, связанной с сохранением языкового многообразия, в т. ч. в рамках Международного десятилетия языков коренных народов (2022–2032). Письменный корпус татарского языка представляет собой не только научный, но и культурно-стратегический проект, направленный на поддержку и продвижение национального языка в цифровую эпоху.

³² Международное десятилетие языков коренных народов (2022–2032) // UNESCO. URL: <https://www.unesco.org/ru/decades/indigenous-languages> (дата обращения: 15.08.2025).

Список литературы

- Алюнина Ю.М. Цифровые технологии в переводе. СПб. : Лань, 2025. 144 с.
- Белорусова С.Ю. Киберэтнография: методология и технология // Этнография. 2021. № 3 (13). С. 123–145. [https://doi.org/10.31250/2618-8600-2021-3\(13\)-123-145](https://doi.org/10.31250/2618-8600-2021-3(13)-123-145) EDN: DXBUUJ
- Богородицкий В.А. Введение в татарское языкознание в связи с другими тюркскими языками / под ред. Н.К. Дмитриева. Казань : Татаргосиздат, 1953. 220 с.
- Гатиатуллин А.Р., Мухамедишин Д.Р., Прокопьев Н.А., Сулейманов Д.Ш. Электронный корпус татарского языка на базе модели лингвистических графов знаний // Онтология проектирования. 2024. Т. 14. № 4 (54). С. 542–554. <https://doi.org/10.18287/2223-9537-2024-14-4-542-554> EDN: FXVFET
- Ибрагимов Т.И., Сайхунов М.Р. Письменный корпус татарского языка: структурные и функциональные характеристики // Актуальные проблемы диалектологии языков народов России : материалы XIV Всеросс. науч. конф., Уфа, 20–22 ноября 2014 г. / отв. ред. Ф.Г. Хисамитдинова. Уфа : Институт истории, языка и литературы Уфимского научного центра РАН, 2014. С. 261–264. EDN: TIEBGB
- Ибрагимов Т.И., Сайхунов М.Р. Языковое состояние этнической общности на материале Письменного корпуса татарского языка // Tatarica. 2015. № 1 (4). С. 22–27. EDN: UNYXIL
- Ибрагимов Т.И., Сайхунов М.Р. Хэзерге татар сөйләм теле: сузык авазлар составы // Фэнни Татарстан. 2016. № 3. С. 35–47. EDN: XRPKMН
- Ибрагимова Э.Р. К вопросу о референтном и атрибутивном употреблении наименований лица в предложениях тождества в английском и татарском языках // Филология и культура. 2020. № 4 (62). С. 36–42. <https://doi.org/10.26907/2074-0239-2020-62-4-36-42> EDN: RXAZHS
- Кузнецов М.Ю. Алгоритм нахождения основы татарского глагола по его инфинитиву (орфографическо-грамматические аспекты) // Многоязычие в образовательном пространстве. 2023. Т. 15. № 2 (17). С. 192–202. <https://doi.org/10.35634/2500-0748-2023-15-2-192-202> EDN: YUGYOJ
- Родионова А.П., Пеллинен Н.А. Корпусная лингвистика и марафон записей вепсской и карельской речи как инструмент популяризации прибалтийско-финских языков Карелии // *Macrosociolinguistics and Minority Languages*. 2024. Т. 2. № 2. С. 115–130. <https://doi.org/10.22363/2312-797X2024-2-2-115-130> EDN: IPVMMVI
- Сайхунов М.Р. Ритмо-темпоральные характеристики татарского языка в плане автоматического синтеза речи : автореф. дис. ... канд. филол. наук. Казань, 2010. 26 с. EDN: QGVUUR
- Сайхунов М.Р., Ибрагимов Т.И., Галиуллин К.Р. Корпус татарской художественной литературы // Традиционная культура народов Поволжья : материалы IV Всеросс. науч.-практ. конф. с междунар. участием. Казань : ИХЛАС, 2018. С. 370–377. EDN: USSRQS
- Сайхунов М.Р., Хусаинов Р.Р., Ибрагимов Т.И. Система сложного морфологического поиска в Письменном корпусе татарского языка // Традиционная культура тюркских народов в изменяющемся мире : материалы I Междунар. науч. конф. Казань : Ак Буре, 2017. С. 382–385. EDN: YSRGAQ
- Сайхунов М.Р., Хусаинов Р.Р., Ибрагимов Т.И. Сложности при создании текстового корпуса объемом более 400 млн токенов // Финно-угорский мир в полиэтничном пространстве России: культурное наследие и новые вызовы : сб. статей по материалам VI Всеросс. науч. конф. финно-угроведов, Ижевск, 04–07 июня 2019 г. Ижевск : Изд-во Анны Зелениной, 2019. С. 548–554. EDN: LSYZBT
- Сайхунов М.Р., Хусаинов Р.Р., Ибрагимов Т.И. Эволюция систем поиска в Письменном корпусе татарского языка // Языковые контакты народов Поволжья и Урала : сб. статей XI Междунар. симпозиума, Чебоксары, 21–24 мая 2018 г. Чебоксары : Чувашский государственный университет им. И.Н. Ульянова, 2018. С. 314–319. EDN: XVTCWD
- Galieva A., Vavilova Z., Gafarova V. Developing Tatar corpus-based dictionaries for educational purposes // INTED2017 Proceedings. Valencia : INTED, 2017. P. 9014–9022. <https://doi.org/10.21125/inted.2017.2131>
- Luutonen J., Moiois A., Daher O. Finnish Tatars and the trilingual Tatar-Finnish dictionary // *Turcic Languages*. 2017. Vol. 21. № 2. P. 266–280. <https://doi.org/10.13173/TL/2017/2/266>

References

- Alyunina, Yu.M. (2025). *Tsifrovye tekhnologii v perevode [Digital technologies in translation]*. Lan' Publ. (In Russ.).
- Belorussova, S. Yu. (2021). Cyberethnography: Methodology and technology. *Etnografia*, (3), 123–145. (In Russ.). [https://doi.org/10.31250/2618-8600-2021-3\(13\)-123-145](https://doi.org/10.31250/2618-8600-2021-3(13)-123-145) EDN: DXBUUJ

- Bogoroditsky, V.A., Dmitriev, N., ed. (1953). *Vvedenie v tatarskoe yazykoznanie v svyazi s drugimi tyurkskimi yazykami [Introduction to Tatar Linguistics in Relation to Other Turkic Languages]*. Kazan: Tatgosizdat publ. (In Russ.).
- Galieva, A., Vavilova, Z., & Gafarova, V. (2017). Developing Tatar corpus-based dictionaries for educational purposes. *INTED2017 Proceedings*, 9014–9022. <https://doi.org/10.21125/inted.2017.2131>
- Gatiatullin, A.R., Mukhamedshin, D.R., Prokopyev, N.A., & Suleymanov, D. Sh. (2024). Electronic corpus of the Tatar language based on the model of linguistic knowledge graphs. *Ontology of Designing*, 14(4), 542–554. (In Russ.). <https://doi.org/10.18287/2223-9537-2024-14-4-542-554> EDN: FXVFET
- Ibragimov, T.I., & Saikhunov, M.R. (2014). The written corpus of the Tatar language: Structural and functional characteristics. In F.G. Khisamitdinova (ed.). *Proceedings of the 14th All-Russian scientific conference «Current issues in the dialectology of the languages of the peoples of Russia»*, 261–264. Ufa: Federal State Budgetary Institution of Science, Institute of History, Language and Literature, Ufa Scientific Center of the Russian Academy of Sciences. (In Russ.). EDN: TIEBGB
- Ibragimov, T.I., & Saikhunov, M.R. (2015). The language status of an ethnic community (on the material of the Tatar language written corpus). *Tatarica*, (1), 22–27. (In Russ.). EDN: UNYXIL
- Ibrahimov, T.I., & Saykhunov, M.R. (2016). The modern Tatar spoken language: The structure of vowel sounds. *Fanni Tatarstan*, (3), 35–47. (In Tatar). EDN: XRPKMH
- Ibragimova, E.R. (2020). On the reference and attributive use of personal names in English and Tatar identical utterances. *Philology and Culture*, (4), 36–42. (In Russ.). <https://doi.org/10.26907/2074-0239-2020-62-4-36-42> EDN: RXAZHS
- Kuznetsov, M. Yu. (2023). Algorithm for finding the base of the Tatar verb by its infinitive (orthographic-grammatical aspects). *Russian Journal of Multilingualism and Education*, 15(2), 192–202. (In Russ.). <https://doi.org/10.35634/2500-0748-2023-15-2-192-202> EDN: YUGYOJ
- Luutonen, J., Moisio, A., & Daher, O. (2017). Finnish Tatars and the trilingual Tatar-Finnish dictionary. *Turcic Languages*, 21(2), 266–280. <https://doi.org/10.13173/TL/2017/2/266>
- Rodionova, A.P., & Pellinen, N.A. (2024). The corpus linguistics and the marathon of recordings of Vepsian and Karelian speech as a tool for popularizing the Baltic-Finnish languages of Karelia. *Macrosociolinguistics and Minority Languages*, 2(2), 115–130. (In Russ.). <https://doi.org/10.22363/2312-797X2024-2-2-115-130> EDN: IPVMVI
- Saikhunov, M.R. (2010). *Ritmo-temporal'nye kharakteristiki tatarskogo yazyka v plane avtomaticheskogo sinteza rechi [Rhythmic and temporal characteristics of the Tatar language in the context of automatic speech synthesis] [Dissertation abstract]*. Kazan: Kazan State University. (In Russ.). EDN: QGVUYR
- Saikhunov, M.R., Ibragimov, T.I., & Galiullin, K.R. (2018). The Corpus of Tatar fiction literature. *Traditional culture of the peoples of the Volga Region: Proceedings of the 4th All-Russian scientific and practical conference with international participation*, p. 370–377. Kazan: Ikhlas publ. (In Russ.). EDN: USSRQS
- Saikhunov, M.R., Khusainov, R.R., & Ibragimov, T.I. (2017). The system of advanced morphological search in the written corpus of the Tatar language. *Traditional culture of Turkic peoples in a changing world: Proceedings of the 1st international scientific conference*, p. 382–385. Kazan: Ak Bure publ. (In Russ.). EDN: YSRGAQ
- Saikhunov, M.R., Khusainov, R.R., & Ibragimov, T.I. (2018). Search systems evolution in the corpus of written Tatar language. *Language contacts of the peoples of the Volga Region and the Urals: Proceedings of the 11th international symposium*, p. 314–319. Chuvash State University named after N. Ulyanov. (In Russ.). EDN: XVTCWD
- Saikhunov, M.R., Khusainov, R.R., & Ibragimov, T.I. (2019). Challenges in creating a text corpus exceeding 400 million tokens. *The Finno-Ugric World in the Multiethnic Space of Russia: Cultural Heritage and New Challenges: Proceedings of the 6th All-Russian conference on Finno-Ugric studies*, p. 548–554. Izhevsk, Anna Zelenina publ. (In Russ.). EDN: LSYZBT

Сведения об авторе:

КЕНЕШБЕКОВА Шахдар Нурдиновна, аспирант кафедры социальной педагогики Института иностранных языков, Российский университет дружбы народов, Российская Федерация, 117198, г. Москва, ул. Миклухо-Маклая, д. 6; старший преподаватель кафедры лингвистики и межкультурной коммуникации Гуманитарного института, Российский новый университет, Российская Федерация, 105005, Москва, ул. Радио, д. 22. *Научные интересы:* корпусная лингвистика, компьютерная лингвистика, тюркские языки, сопоставительные исследования, методика преподавания иностранных языков.

E-mail: keneshbekova.sh@yandex.ru

ORCID: 0009-0002-3599-2065

SPIN-код: 8204-2172

Bionote:

Shakhdar N. KENESHBEKOVA, a PhD student at the Department of Social Pedagogy, Institute of Foreign Languages, RUDN University, 6 Miklukho-Maklaya St., Moscow, 117198, Russian Federation; Senior Lecturer at the Department of Linguistics and Intercultural Communication, Institute of Humanities, Russian New University, 22 Radio St., Moscow, 105005, Russian Federation. *Research interests:* corpus linguistics, computational linguistics, Turkic languages, contrastive studies, methods of teaching foreign languages.

E-mail: keneshbekova.sh@yandex.ru

ORCID: 0009-0002-3599-2065

SPIN-code: 8204-2172



DOI: 10.22363/2949-5997-2025-3-2-117-130

EDN HDFRDB

Научная статья / Research article

Электронный текстовый корпус лингвокультуры Северного Приангарья: структура, состав и возможности использования

Елена КИРИЛЕНКО 

Российский университет дружбы народов, Москва, Российская Федерация

✉ kirilenko-e@rudn.ru

Аннотация. Развитие цифровых технологий привело к появлению новых и удобных инструментов документирования и сохранения находящихся под угрозой исчезновения идиомов. Под последними могут пониматься как языки, так и диалекты и иные варианты языков, стандартные формы которых обладают высокой витальностью. Исследование посвящено комплексному анализу структуры, состава и функционального потенциала Электронного текстового корпуса лингвокультуры Северного Приангарья (КЛСП). Актуальность исследования обусловлена необходимостью фиксации исчезающих региональных вариантов русского языка и связанных с ними культурных практик в условиях глобализации и масштабных социально-экономических изменений. Цель исследования — представить детальное описание корпуса как междисциплинарного инструмента для гуманитарных исследований. Материалом исследования послужили данные, представленные на официальном сайте корпуса, включая его структуру, метаразметку и описательные разделы. Основными методами выступили описательный анализ, корпусная лингвистика и элементы контент-анализа. Результаты исследования включают: детализацию трехуровневой структуры корпуса (диалектный, фольклорный, мультимедийный подкорпуса); описание уникальной системы многоаспектной ручной разметки (ареальная, временная, жанровая, тематическая, концептуальная, сюжетно-мотивная); анализ функциональных возможностей онлайн-платформы для сложного поиска; оценку научной значимости корпуса на основе уже выполненных исследований в области коммуникативной диалектологии, этнолингвистики и фольклористики. Подчеркивается роль КЛСП как ключевого ресурса для сохранения нематериального культурного наследия Северного Приангарья и потенциал корпуса в образовательных, лексикографических и технологических проектах.

Ключевые слова: корпусная лингвистика, документирование языка, русский язык, диалект, витальность идиома, сохранение языка

© Кириленко Е., 2025

This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

Заявления о конфликте интересов. Автор заявляет об отсутствии конфликта интересов.

История статьи: получена 20 июня 2025; принята в печать 15 августа 2025.

Для цитирования: Кириленко Е. Электронный текстовый корпус лингвокультуры Северного Приангарья: структура, состав и возможности использования // *Macrosociolinguistics and Minority Languages*. 2025. Т. 3. № 2. С. 117–130. <https://doi.org/10.22363/2949-5997-2025-3-2-117-130> EDN: HDFRDB

Digital corpus of the linguoculture of the Northern Angara region: structure, composition and applications

Elena KIRILENKO 

RUDN University, Moscow, Russian Federation

✉ kirilenko-e@rudn.ru

Abstract. The development of digital technology has led to the emergence of new and convenient tools for documenting and preserving endangered idioms. These include languages, dialects, and other language variants whose standard forms are highly active. The study presents a comprehensive analysis of the Electronic Text Corpus of the Linguistic Culture of the Northern Angara Region (CLCNA), focusing on its structure, composition, and functional potential. The relevance of this study stems from the need to preserve regional variants of the Russian language and associated cultural practices in a rapidly changing world. The purpose of this study is to describe the corpus as a valuable tool for humanities research, providing a detailed overview of its structure and features. The study relies on data available on the official corpus website, including information about its structure, metadata, and descriptive elements. Using descriptive analysis, corpus linguistics, and content analysis techniques, the author provides a thorough description of the CLCNA's three-tiered structure, including dialectal, folklore, and multimedia subsets. The findings of this research contribute to a better understanding of the corpus's capabilities and potential applications in various fields of study; a description of the unique system of multi-dimensional manual annotation (spatial, temporal, genre, thematic, conceptual, and plot motif); an analysis of the functional capabilities of the online platform for complex search; and an assessment of the scientific significance of the corpus based on previous research in the fields of communicative dialectology, ethnolinguistics, and folklore studies. The study concludes by emphasizing the role of CLCNA as a key resource for preserving the intangible cultural heritage of the Northern Angara Region and its potential in educational, lexicographic, and technological projects.

Key words: corpus linguistics, language documentation, Russian language, dialect, idiom vitality, language preservation

Conflict of interest. The author declares no conflicts of interest.

Article history: received 20 June 2025; accepted 15 August 2025.

For citation: Kirilenko, E. (2025). Digital corpus of the linguoculture of the Northern Angara region: Structure, composition and applications. *Macrosociolinguistics and Minority Languages*, 3(2), 117–130. (In Russ.). <https://doi.org/110.22363/2949-5997-2025-3-2-117-130> EDN: HDFRDB

Введение

Сохранение языкового и культурного разнообразия, особенно в контексте исчезающих языков и региональных вариантов больших языков, является одной из актуальных задач современной гуманитарной науки, входит в число приоритетных направлений развития языковой политики России¹. Создание электронных корпусов выступает ключевым инструментом для документирования, исследования и актуализации таких форм нематериального наследия. Примерами корпусов языков коренных народов Российской Федерации могут служить Корпус вепсского языка (ВепКар)² (Родионова, Пеллинен, 2024), Корпус мансийского языка³, Корпус текстов на языках коренных малочисленных народов Севера, Сибири и Дальнего Востока РФ. Эти ресурсы обеспечивают базу для лингвистических, антропологических и культурологических исследований, а также способствуют сохранению уникальных идиомов в условиях глобализации.

Не менее значимы усилия по документированию и сохранению региональных вариантов русского языка. Русский язык безусловно относится к мажоритарным языкам и является одним из наиболее функционально развитых языков мира. Однако, как любой язык, он подвержен диатопическому варьированию, что приводит к появлению территориальных вариантов-диалектов. Диалект — это не просто подвид коммуникативной системы языка. Обслуживая определенные языковые коллективы, диалект отражает особенности и идентичность его носителей, обусловленные культурным укладом, типом хозяйственной деятельности и иными факторами экстралингвистического порядка. В этом смысле региональный диалект схож с любым отдельным языком, а его изучение, в т. ч. и документирование, служит делу сохранения культурного наследия русского народа, как это происходит в случае с документированием языков народов России. Электронный текстовый корпус лингвокультуры Северного Приангарья (КЛСП)⁴

¹ Закон РФ «О языках народов Российской Федерации» от 25.10.1001 № 1807-1 с изменениями от 13.06.2023 № 253-ФЗ. URL: https://www.consultant.ru/document/cons_doc_LAW_15524/ (дата обращения: 24.07.2025). Распоряжение Правительства Российской Федерации от 12.06.2024 № 1481-р. URL: <http://publication.pravo.gov.ru/document/0001202406140048?index=1> (дата обращения: 15.08.2025). Указ Президента Российской Федерации от 8 мая 2024 г. № 314 «Об утверждении Основ государственной политики Российской Федерации в области исторического просвещения». URL: <http://www.kremlin.ru/acts/bank/50534> (дата обращения: 24.07.2025).

² Корпус вепсского языка (ВепКар). URL: <http://dictorpus.krc.karelia.ru/ru> (дата обращения: 10.06.2025).

³ Корпус мансийского языка. URL: <https://mansi.pro/corpus/> (дата обращения: 10.06.2025).

⁴ Корпус лингвокультуры Северного Приангарья (КЛСП). URL: <https://angara.sfu-kras.ru/> (дата обращения: 10.06.2025).

тоже выполняет эту задачу, фокусируясь на документации и исследовании русского языка и культуры в его сибирском региональном воплощении — ангарской лингвокультуре.

Русское население Северного Приангарья сформировалось в результате освоения бассейна реки Ангары выходцами из европейской части России в XVII–XVIII вв., в период продвижения русских поселенцев в Восточную Сибирь. Расселение осуществлялось преимущественно вдоль речных путей и было связано с основанием острогов, развитием промысловой, торговой и земледельческой деятельности. В последующие столетия население региона пополнялось за счет как добровольных миграций, так и административных переселений. Русские жители Северного Приангарья проживают в населенных пунктах Иркутской области и Красноярского края, расположенных вдоль среднего и нижнего течения Ангары. Длительное существование локальных сообществ в условиях относительной территориальной и культурной изолированности способствовало формированию устойчивых региональных языковых и культурных особенностей, отраженных в материалах корпуса.

Цель исследования — комплексный анализ и описание структуры, состава и функциональных возможностей КЛСП как междисциплинарного цифрового ресурса, а также оценка его роли в сохранении, изучении и актуализации лингвокультурного наследия Северного Приангарья. Материалом исследования стал сам корпус КЛСП, его структура, система разметки, контент и технологические решения, представленные на официальной платформе проекта и в сопутствующих научных публикациях.

Актуальность работы определяется двумя ключевыми факторами. Во-первых, необходимостью документирования исчезающих языков, диалектов и традиций Северного Приангарья в условиях глобализации. Во-вторых, растущей востребованностью корпусных методов в цифровой лингвистике и искусственном интеллекте. КЛСП не только фиксирует языковые особенности региона, но и воссоздает культурный контекст через фольклорные тексты, аудио- и видеозаписи, что делает его незаменимым для этнолингвистических исследований, цифровых гуманитарных проектов и разработки образовательных курсов.

Исследована трехкомпонентная структура корпуса, а именно диалектный, фольклорный и мультимедийный подкорпусы. Особое внимание уделено системе разметки данных, включая ареальную, жанровую и сюжетно-мотивную классификацию. Приведены примеры использования ресурса в лингвистическом анализе (например, при изучении региолектных вариаций), создании учебных модулей (курсы по диалектологии) и разработке алгоритмов обработки поликодовых текстов (синхронизация аудио- и видеоматериалов с текстовыми данными). Рассмотрена также роль корпуса в сохранении нематериального

культурного наследия и формировании цифровой идентичности коренных жителей региона, что подчеркивает его значение для современных гуманитарных исследований.

История разработки электронного текстового КЛСП и его состав

Проект КЛСП был инициирован в 2012 г. научным коллективом Сибирского федерального университета (СФУ, Красноярск) под руководством специалиста в области коммуникативной диалектологии и лексикографии д-ра филол. наук, проф. О.В. Фельде (Фельде, 2021) и реализуется в рамках научно-исследовательской деятельности СФУ⁵. Основная цель проекта — создание комплексного электронного архива, фиксирующего современное состояние русской лингвокультуры Северного Приангарья через призму ее языкового выражения и культурных практик.

Ключевые задачи проекта:

- фиксация устных текстов (современная лингвокультура, исчезающие ангарские говоры, локальное просторечие, ангарский региолект) с использованием аудио- и видеоаппаратуры;
- корпусная разметка материала и представление его в виде электронной системы, включающей тексты разных жанров, включая поликодовые и аудиовизуальные;
- создание и наполнение трех специализированных подкорпусов с соответствующей разметкой (диалектный, фольклорный, мультимедийный);
- научное обоснование информативности материала с позиций коммуникативной диалектологии, лингвокультурологии и фольклористики.

Материалом для корпуса служат полевые записи, собранные экспедициями СФУ. Корпус основан преимущественно на устных текстах различного характера (интервью, беседы, рассказы, фольклорные произведения), а также включает мультимедийные данные (аудио-, видео-, фото-), синхронизированные с текстовыми расшифровками. Таким образом, КЛСП является по своей сути устно-речевым поликодовым корпусом⁶, он имеет трехуровневую структуру, охватывающую текстовые и мультимедийные материалы.

Диалектный подкорпус. Включает 1445 текстов (441 тыс. слов), фиксирующих особенности ангарских говоров, региолектов и просторечий. Тексты снабжены ареальной, жанровой и тематической разметкой. Например, указывается место записи (Кежемский, Богучанский, Мотыгинский районы), тип речи (монолог, диалог), жанр (устный рассказ, интервью) и лингвокультурные концепты («медведь», «домовой»). Ниже приведен пример одного из текстов.

⁵ Корпус лингвокультуры Северного Приангарья (КЛСП). URL: <https://angara.sfu-kras.ru/> (дата обращения: 10.06.2025).

Место записи: Красноярский край, г. Красноярск

Год записи: 2021

*Кстати, про рыбу. Мы не поговорили с вами про пескарей. Это кежемская тема. Пескарь в Кежме считался рыбой неприличной. Не в смысле пошлой и хамской, а которую есть стыдно, ловить стыдно, вот. И более того, в Кежме он носил название **мокчён**. Назывались они **мокчёны**. И мокчёнов можно было ловить только детям для своих животных: для кошки, для собаки, вот. Самому есть **мокчёнов** — это стыд и позор, значит, это нельзя, вот. Ну и за счёт того, что их не ловил никто, обычно в берег заходишь, они прямо вот такими стаями к ногам, к ногам. Было очень много, за счёт того, что их никто не ловил, вот. И эта история, она приобрела большой символ (Красноярск, 2021 < Кежма).*

Фольклорный подкорпус предоставляет доступ к уникальным образцам устного народного творчества Северного Приангарья. На момент анализа (май 2025 г.) подкорпус содержит 438 произведений (76 тыс. слов), охватывающих широкий спектр жанров, что позволяет исследовать как обрядовую, так и необрядовую традицию региона. Количественный анализ, выполненный с помощью встроенных фильтров платформ⁶, выявляет следующую структуру материалов (табл.).

Количество материалов в фольклорном подкорпусе электронного текстового корпуса лингвокультуры Северного Приангарья / The number of materials in the folklore subcorpus of the Electronic text corpus of linguoculture of the Northern Angara region

Вид материала / Type of material	Количество /Number
Былички и демонологические / Supernatural tale	11
Рассказы / Stories	8
Пословицы и поговорки / Proverbs and sayings	156
Приметы и поверья / Omens and sayings	29
Календарные обряды и календарная поэзия / Calendar rites and poetry	41
Детский фольклор / Children's folklore	27
Частушки / Ditties	17
Легенды / Legends	18

⁶ Корпус лингвокультуры Северного Приангарья (КЛСП). URL: <https://angara.sfu-kras.ru/> (дата обращения: 10.06.2025).

Окончание табл. / Ending table

Вид материала / Type of material	Количество / Number
Лирические необрядовые песни / Lyrical non-ritual songs	11
Заговоры / Spells	7
Предания / Lore	8
Анекдоты / Anecdote	4
Сказки / Tales	2

Источник: составлено Е. Кириленко / Source: compiled by E. Kirilenko.

Такое распределение (см. табл.) свидетельствует о преобладании малых прозаических жанров (былички, пословицы), связанных с повседневным опытом и мифологическим сознанием, что характерно для современной фольклорной традиции старожильческого населения Сибири. Значительный блок календарных обрядов (41 единица) позволяет реконструировать аграрный и праздничный циклы ангарцев. В их числе представлены как общераспространенные обряды (например, святочные гадания, масленичные обычаи), так и локальные практики, связанные с началом посевной, сенокоса или охотничьего сезона.

Материалы классифицированы не только по жанрам, но и по сюжетно-мотивным признакам, что позволяет проводить тематический и сравнительный анализ. Например, детальная разметка календарных обрядов дает возможность изучать их региональную специфику, вариативность исполнения и связанные с ними вербальные тексты (заклички, приговоры). Приведем пример.

Место: Богучанский район, п. Артюгино

Год записи: 2017

Жанр уст. текста: Календарные обряды и календарная поэзия

— Б.В.А.: Вот на Троицу берёзку наряжали. После Троицы всегда вот ставили берёзку в дома. А потом её в понедельник, эту берёзку наряжали, спускали по речке, по реке с ленточками. Ну, тогда чо было — ленточки, тряпочки, цветочки. Отправляли по теченью, по Ангаре ту берёзку... Ну, загадывали желанья, они сбывались... здоровье⁷.

Мультимедийный подкорпус. Объединяет аудио-, видео- и фотоматериалы, синхронизированные с текстами. Раздел включает документальные фильмы, записи фольклорных обрядов, интервью с носителями диалектов

⁷ Корпус лингвокультуры Северного Приангарья (КЛСП). URL: <https://angara.sfu-kras.ru/> (дата обращения: 10.06.2025).

и изображения этнографических артефактов. Далее приведен пример формата в мультимедийном подкорпусе.

Легенда об открывшемся небе

Информант: Рукоуев Владимир Александрович, родился в 1951 г.

в д. Яркино Кежемского р-на; проживает в пгт. Богучаны

Красноярского края.

Оператор: Васильев В.К.

Видеозапись сделана в пгт. Богучаны в июле 2017 г., доступна по ссылке на сайте⁸.

Пополнение корпуса осуществляется вручную в ходе ежегодных полевых исследований экспедиционных групп Сибирского федерального университета. Собранные аудио- и видеозаписи подвергаются комплексной обработке: лингвистической транскрипции, этнографическому описанию и многоуровневой аннотации командой специалистов (лингвистов, фольклористов, этнографов). Ключевой технологической особенностью проекта является поликодовая разметка.

Несмотря на отсутствие регулярного целевого финансирования, проект развивается благодаря усилиям научного коллектива и институциональной поддержке СФУ, регулярно пополняясь новыми материалами и совершенствуя функционал платформы.

Функциональные возможности электронного текстового КЛСП

Функциональные особенности КЛСП обеспечиваются детальной системой разметки и развитым поисковым интерфейсом, включающим следующие ключевые аспекты.

Ареальная разметка (Метаданные). Данный уровень разметки фиксирует пространственно-временные и персональные параметры материала: географическое происхождение (районы Кежемский, Богучанский, Мотыгинский; конкретный населенный пункт), год записи, ФИО собирателя. Информация об информантах (ФИО, место рождения, социальные характеристики) вынесена в отдельный список с обязательным указанием согласия на использование персональных данных (рис. 1).

Жанрово-типологическая разметка. Система реализует дифференцированный подход к классификации текстов:

- диалектный подкорпус: учитывает вид речи (монолог, диалог, полилог), тип текста для монологов (повествование, описание, рассуждение) и жанровую принадлежность (устный рассказ [нарратив, меморат, фабулат, предание,

⁸ Корпус лингвокультуры Северного Приангарья (КЛСП). URL: <https://angara.sfu-kras.ru/> (дата обращения: 10.06.2025).

Место записи:	Мотыгинский район, пгт.Мотыгино
Год записи:	2017
<p>О том, что такое Родина Вот, спрашивают: «Чё, вот, где Родина, где у тебя Родина? Чё ты сюда приехал». А тот отвечает: «Там где пупок обрезали – там и Родина». Это, по-моему, очень точное (Мт.: Мотыгино, 2017).</p>	
Подробнее	
Информант:	Аннонен Виктор Иоганнович, родился в 1951 г. в п. Южно-Енисейск Удереийского р-на Красноярского края, в н. вр. проживает в пгт. Мотыгино Мотыгинского р-на.
Собиратель:	Смирнов Е.С.

Рис. 1. Метаданные Электронного текстового корпуса лингвокультуры Северного Приангарья
 Источник: КЛСП.

URL: <https://angara.sfu-kras.ru/> (дата обращения: 25.07.2025).

Fig. 1. Metadata in the Electronic text corpus of the linguoculture of the Northern Angara region

Source: Digital corpus of the linguoculture of the Northern Angara region.

Retrieved 25 July 2025, from: <https://angara.sfu-kras.ru/>

быличка], разговор/беседа [социально-бытовой, информативный и др.], интервью, реплика);

- фольклорный подкорпус: классифицирует материалы по жанрам (обряд, гадание, заговор, загадка, предание, быличка, бывальщина, легенда, частушка и т.д.) (рис. 2).

Слово или фраза:	<input type="text"/>
Место записи текста:	Выбрать
Время записи текста:	2014 <input type="text"/> 2025 <input type="text"/>
Фольклорный жанр:	Былички и бывальщины / Демонологические рассказы <input type="text"/>
Сюжет/Мотив:	Водяной боится грома <input type="text"/>
Искать	

Рис. 2. Система поиска в фольклорном подкорпусе Электронного текстового корпуса лингвокультуры Северного Приангарья

Источник: КЛСП. URL: <https://angara.sfu-kras.ru/> (дата обращения: 25.07.2025).

Fig. 2. The search system in the folklore subcorp of the Electronic text corpus of the linguoculture of the Northern Angara region

Source: Digital corpus of the linguoculture of the Northern Angara region.

Retrieved 25 July 2025, from: <https://angara.sfu-kras.ru/>

Тематическая и концептная разметка (Диалектный подкорпус). Данный уровень обеспечивает глубину семантического анализа:

- тематика текста — маркируются ключевые темы (например, «охота», «рыбалка», «переселение», «быт»), причем один текст может быть отнесен к нескольким темам;

- лингвокультурный концепт — идентифицируются и отмечаются ключевые культурные концепты, эксплицируемые в тексте («домовой», «леший», «медведь», «река Ангара», «тайга»);

- диалектизмы / просторечия — фиксируются специфические слова или выражения.

Сюжетно-мотивная разметка (Фольклорный подкорпус). Специализированная разметка, применяемая преимущественно к прозаическим фольклорным жанрам (былички, бывальщины). Она классифицирует тексты по мотивам встреч с мифологическими персонажами (домовой, леший, водяной) или сверхъестественными явлениями.

Мультимедийная атрибуция. Этот уровень обеспечивает детальное описание невербальных материалов, среди которых можно назвать следующие:

- жанр записи (документальный фильм, видеосюжет, фольклорное исполнение, интервью, аудиозапись речи, фотография);

- точные параметры: место и время съемки/записи;

- участники: информанты, исполнители, собиратели, операторы.

Данная атрибуция гарантирует корректную интерпретацию и контекстуализацию мультимедийных ресурсов корпуса.

Научная значимость электронного текстового КЛСП

Электронный текстовый КЛСП стал основой для исследований, демонстрирующих его междисциплинарный потенциал. Благодаря детальной разметке и разнообразию данных, ресурс активно используется в лингвистике, фольклористике, этнографии и культурологии.

КЛСП служит для изучения диалектных и региолектных особенностей русского языка. Ю.В. Уткина (Уткина, 2024: 123) исследует региональные вариации лексики, связанные с мифологическими персонажами, что демонстрирует роль корпуса в выявлении локальных культурно-языковых феноменов. А.Н. Сперанская (Сперанская, 2017: 392) анализирует речевые стратегии, выявляя, как диалогические элементы структурируют устные рассказы. Е.С. Смирнов (Смирнов, 2018) изучает этнолингвистические аспекты взаимодействия ангарцев с коренными народами Сибири, опираясь на интервью и диалоги из корпуса.

Фольклорный подкорпус КЛСП стал основой для реконструкции традиционной культуры. В.К. Васильев (Васильев, 2018: 150) исследует сказки и легенды, сопоставляя их с аудиозаписями обрядовых песен. К.В. Кайзер (Кайзер, 2018) анализирует демонологические рассказы, выявляя связь между языковыми

маркерами и этнографическими артефактами. Особый интерес представляет статья О.В. Вязовкиной (Вязовикина, 2020: 137), где охотничьи рассказы из корпуса рассматриваются как источник знаний о традиционном укладе жизни.

Корпус также используется как инструмент для изучения исторических процессов. Таким образом, КЛСП не только фиксирует региональное наследие, но и стимулирует новые направления в гуманитарных науках, подтверждая актуальность корпусных методов в цифровую эпоху.

Перспективы развития и использования электронного текстового КЛСП

Несмотря на существующие ограничения, КЛСП обладает значительным потенциалом развития в научной, образовательной, прикладной и культурно-сохранной сферах. Дальнейшее расширение и устойчивое функционирование корпуса во многом зависят от наличия целевого финансирования, в т.ч. в рамках грантовых программ, поддерживающих цифровые гуманитарные исследования, сохранение нематериального культурного наследия и развитие региональных научно-образовательных инициатив. Перспективными источниками поддержки проекта могут выступать государственные и ведомственные гранты, программы университетского развития, а также проекты, реализуемые в рамках национальных и региональных стратегий цифровизации культуры и образования. Одним из таких источников поддержки был грант Российского фонда фундаментальных исследований, Правительства Красноярского края, Красноярского краевого фонда поддержки научной и научно-технической деятельности.

Научные перспективы включают планомерное пополнение подкорпусов новыми полевыми данными, что обеспечит репрезентативность и актуальность ресурса. Ключевое направление развития — углубление разметки, потенциально включающее внедрение морфологической или синтаксической аннотации для репрезентативных текстов, что откроет новые возможности для системного лингвистического анализа. Повышение исследовательской ценности платформы связано с развитием встроенных инструментов анализа, таких как конкордансеры и генераторы частотных словарей. Усиление связей между подкорпусами позволит создавать комплексные «досье» по информантам и темам, интегрирующие текстовые, аудио- и видеоматериалы. Корпус также обладает уникальным потенциалом для сравнительных исследований с другими региональными корпусами русского языка и языков коренных народов Сибири.

В образовательной сфере КЛСП может стать основой для создания интерактивных модулей и онлайн-курсов по диалектологии, лингвокультурологии и фольклористике Сибири. Его материалы пригодны для интеграции в регио-

нальный компонент образования школ и вузов Красноярского края. Разработка просветительских цифровых проектов (виртуальные выставки, цифровые истории) будет способствовать популяризации культурного наследия Приангарья.

Прикладное значение ресурса заключается в возможности разработки электронных словарей и атласов ангарских говоров. Массив аудиозаписей представляет ценность для создания речевых технологий (ASR — Automatic Speech Recognition, TTS — Text-to-Speech), адаптированных к региональным особенностям русской речи. Корпус может служить основой для создания музеев и центров сохранения культурного наследия в районах Северного Приангарья.

Культурно-сохранная роль КЛСП соответствует глобальным и национальным инициативам: целям Международного десятилетия языков коренных народов (2022–2032) ЮНЕСКО⁹ по документированию исчезающих идиом (к которым относятся ангарские говоры) и Стратегии государственной национальной политики РФ, поддерживающей сохранение культурного наследия. КЛСП выступает действенным инструментом фиксации НКН русского старожильческого населения Сибири и формирования его цифровой идентичности.

Реализация этого потенциала зависит от решения ключевых задач: привлечения стабильного финансирования и разработки полуавтоматических методов обработки данных для снижения трудоемкости транскрипции и разметки.

Заключение

Электронный текстовый КЛСП представляет собой значимый и уникальный цифровой ресурс, созданный усилиями ученых Сибирского федерального университета. Его трехкомпонентная структура (диалектный, фольклорный, мультимедийный подкорпуса) обеспечивает комплексное представление об ангарской лингвокультуре через аутентичные устные тексты, фольклорные произведения и мультимедийные материалы. Детальная система разметки (арёальная, жанрово-типологическая, тематическая, концептная, сюжетно-мотивная) является ключевым достоинством корпуса, предоставляя исследователям мощный инструмент для формирования релевантных выборок и проведения междисциплинарных изысканий в области лингвистики, фольклористики, этнографии и культурологии.

Несмотря на отсутствие регулярного финансирования, проект демонстрирует устойчивость и продолжает развиваться. КЛСП уже доказал свою научную значимость, став основой для ряда исследований особенностей ангарской речи, регионального фольклора и культурных практик. Потенциал КЛСП далеко не исчерпан, а перспективы развития связаны с пополнением корпуса, углублением разметки, улучшением функционала и интеграцией в более ши-

⁹ Международное десятилетие языков коренных народов (2022–2032) // UNESCO. URL: <https://www.unesco.org/ru/decades/indigenous-languages> (дата обращения: 19.07.2025).

рокий контекст цифровой документации культурного наследия России, что соответствует целям государственной культурной политики и международных инициатив, таких как Десятилетие языков коренных народов ЮНЕСКО.

Будучи инструментом сохранения исчезающих диалектных черт и традиционных культурных практик в эпоху глобализации, КЛСП не только служит научным целям, но и вносит вклад в поддержание региональной идентичности жителей Северного Приангарья. Его развитие и использование в образовательных и культурно-просветительских проектах представляется важным направлением для будущего.

Список литературы

- Васильев В.К. Об архетипическом подходе к анализу женских образов-характеров (письменный и устный текст) // Сибирский филологический журнал. 2018. № 3. С. 142–153. <https://doi.org/10.17223/18137083/64/14> EDN: XZEUXR
- Вязовикина О.В. Ангарский охотничий нарратив как объект лингвистического исследования // Филологические науки. Вопросы теории и практики. 2020. Т. 13. № 1. С. 135–138. <https://doi.org/10.30853/filnauki.2020.1.26> EDN: RBFQFW
- Кайзер К.В. Концепт «знахарка» в ангарской лингвокультуре // Высокие технологии, наука и образование: актуальные вопросы, достижения и инновации : сб. статей II Междунар. науч.-практ. конф., Пенза, 10 ноября 2018 г. Пенза : МЦНС «Наука и Просвещение», 2018. С. 164–166. EDN: YQZBMD
- Родионова А.П., Пеллинен Н.А. Корпусная лингвистика и марафон записей вепсской и карельской речи как инструмент популяризации прибалтийско-финских языков Карелии // *Macrosociolinguistics and Minority Languages*. 2024. Т. 2. № 2. С. 115–130. <https://doi.org/10.22363/2312-797X2024-2-2-115-130> EDN: IPVMMVI
- Смирнов Е.С. Ценностный слой концепта «эвенки» в ангарской лингвокультуре // МНСК-2018: Языкознание : материалы 56-й Междунар. науч. студ. конф. Новосибирск : Новосибирский национальный исследовательский государственный университет, 2018. С. 76–77. EDN: XVMNOP
- Сперанская А.Н. Вставные реплики о родителях в нарративах ангарцев // *Общественные науки*. 2017. № 5. С. 390–399. EDN: XOQCST
- Уткина Ю.В. Репрезентация концепта-мифологемы «домовой» в русской лингвокультуре (на материале региональных быличек и городского фольклора) // Человек и текст : материалы VI Междунар. лингвокультуролог. конф., посв. 225-летию со дня рожд. А.С. Пушкина, Ульяновск, 24 апреля 2024 г. Ульяновск : Ульяновский государственный университет, 2024. С. 118–125. EDN: JVLJJO
- Фельде О.В. Электронный текстовый корпус лингвокультуры Северного Приангарья: принципы создания, структура и возможности использования // *Вестник Кемеровского государственного университета*. 2021. Т. 23. № 4 (88). С. 1086–1095. <https://doi.org/10.21603/2078-8975-2021-23-4-1086-1095> EDN: YBZRHW

References

- Felde, O.V. (2021). Electronic corpus of linguoculture of the Northern Angara region: Foundations, structure, and application. *The Bulletin of Kemerovo State University*, 23(4), 1086–1095. (In Russ.). <https://doi.org/10.21603/2078-8975-2021-23-4-1086-1095> EDN: YBZRHW
- Kaiser, K.V. (2018). The concept 'Healer' in the linguistic culture of Angara region. *Proceedings of the 2nd international scientific and practical conference 'High technologies, science and education: current issues, achievements and innovations'*, 164–166. Nauka i prosveshchenie publ. (In Russ.). EDN: YQZBMD
- Rodionova, A.P., & Pellinen, N.A. (2024). The corpus linguistics and the marathon of recordings of Vepsian and Karelian speech as a tool for popularizing the Baltic-Finnish languages of Karelia. *Macrosociolinguistics and Minority Languages*, 2(2), 115–130. (In Russ.). <https://doi.org/10.22363/2312-797X2024-2-2-115-130> EDN: IPVMMVI
- Smirnov, E.S. (2018). The value layer of the 'Evenki' concept in Angarsk linguoculture. *Proceedings of the 56th international scientific student conference*, 76–77. Novosibirsk State University publ. (In Russ.). EDN: XVMNOP

- Speranskaya, A.N. (2017). Inserted utterances about parents in narratives of the Angara river area inhabitants. *Social Sciences*, (5), 390–399. (In Russ.). EDN: XOQCST
- Utkina, Yu.V. (2024). Representation of the concept-mythologeme 'Brownie' in Russian linguoculture (based on the material of regional and urban folklore). *Chelovek i tekst* [Man and text], 118–125. (In Russ.). EDN: JVLLJO
- Vasiliev, V.K. (2018). On the archetypical approach to the analysis of female images-characters (written and oral text). *The Siberian Journal of Philology*, (3), 142–153. (In Russ.). <https://doi.org/10.17223/18137083/64/14> EDN: XZEUXR
- Vyazovikina, O.V. (2020). Angara hunters' narrative as an object of linguistic studies. *Philology. Theory & Practice*, 13(1), 135–138. (In Russ.). EDN: RBFQFW

Сведения об авторе:

КИРИЛЕНКО Елена, аспирант, лаборант кафедры иностранных языков филологического факультета, Российский университет дружбы народов, Российская Федерация, 117198, Москва, ул. Миклухо-Маклая, д. 6. *Научные интересы:* миноритарные языки, цифровая гуманитаристика, креолизованные тексты, немецкий язык.

E-mail: vshokolade01@yandex.ru

ORCID: 0009-0003-1238-2923

SPIN-код: 6450-4415

Bionote:

Elena KIRILENKO, post-graduate student, Laboratory assistant at the Department of Foreign Languages, Faculty of Philology, RUDN University, 6 Miklukho-Maklaya st., Moscow, 117198, Russian Federation. *Research interests:* minority languages, digital humanities, creolized texts, German.

E-mail: vshokolade01@yandex.ru

ORCID: 0009-0003-1238-2923





SPIN-code: 6450-4415

DOI: 10.22363/2949-5997-2025-3-2-131-145


EDN HHYPZI

Research article / Научная статья

Производство аудиокниг на языках народов России с использованием синтезаторов речи: проблемы и перспективы

Михаил Сергеевич ПОЖИДАЕВ¹  , Елена Сергеевна ТЕПЛЫХ¹ ,
Сергей Ильич ДАНИЛОВ² 

¹Национальный исследовательский Томский государственный университет, Томск, Российская Федерация

²Российский университет дружбы народов, Москва, Российская Федерация
 mosp@luwrain.org

Аннотация. Создание аудиокниг на языках народов России с применением синтезаторов речи — научно и социально значимая задача. Актуальность исследования обусловлена развитием речевых технологий и государственной политикой поддержки языкового разнообразия в т.ч. в цифровом пространстве. Рассмотрен типовой алгоритм создания аудиокниги, выделены инвариантные и лингво-специфичные этапы разработки. Отмечено, что основные сложности связаны с этапами, требующими языковой адаптации текста к озвучиванию синтезатором речи: аннотированием, расшифровкой аббревиатур и сокращений. Для малоресурсных языков особую проблему представляют задачи сегментации, токенизации и контекстного аннотирования, включая обработку омографов и фонетических особенностей конкретных языков. Сделан вывод о невозможности полной автоматизации процесса создания аудиокниг на языках народов России с использованием синтезаторов речи на данном этапе развития этой технологии. Создание аудиокниг на таких языках требует предварительной разработки специализированных лингвистических ресурсов. Необходимым условием является формирование параллельного корпуса текстов и аудиозаписей, созданных носителями языка. Таким образом, успешная реализация подобных проектов требует значительных предварительных работ по сбору обучающих датасетов и адаптации алгоритмов под специфику конкретного языка.

Ключевые слова: миноритарные языки, малоресурсные языки, машинное обучение, распознавание текста, синтезирование речи, рекуррентные нейронные сети

Вклад авторов: Пожидаев М.С. — разработка концепции, формулировка и развитие ключевых целей и задач, проведение исследования, утверждение окончательного варианта результатов

© Пожидаев М.С., Теплых Е.С., Данилов С.И., 2025



This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

исследования; Теплых Е.С. — подготовка и редактирование текста для публикации; Данилов С.И. — подготовка языковых примеров и лингвистическое редактирование текста. Все авторы одобрили окончательную версию статьи.

Заявления о конфликте интересов. Авторы заявляют об отсутствии конфликта интересов.

История статьи: получена 29 июня 2025; принята в печать 10 сентября 2025.


Для цитирования: Пожидаев М.С., Теплых Е.С., Данилов С.И. Производство аудиокниг на языках народов России с использованием синтезаторов речи: проблемы и перспективы // *Macrosociolinguistics and Minority Languages*. 2025. Т. 3. № 2. С. 131–145. <https://doi.org/110.22363/2949-5997-2025-3-2-131-145> EDN: ННYPZI

Production of audiobooks in the languages of the peoples of Russia using speech synthesizers: problems and prospects

Mikhail S. POZHIDAEV¹  , Elena S. TEPLYKH¹ ,
Sergey I. DANILOV² 

¹National Research Tomsk State University, *Tomsk, Russian Federation*

²RUDN University, *Moscow, Russian Federation*

 msp@luwrain.org

Abstract. The creation of audiobooks in the languages of the peoples of Russia using speech synthesizers is a scientifically and socially significant task. The relevance of the research is driven by the development of speech technologies and state policies supporting linguistic diversity, including in the digital space. The study examines a standard algorithm for audiobook creation, distinguishing between invariant and language-specific development stages. The study notes that the main difficulties are associated with the stages requiring linguistic adaptation of the text for speech synthesis: annotation and the expansion of abbreviations and acronyms. For low-resource languages, tasks such as segmentation, tokenization, and contextual annotation, including the processing of homographs and specific phonetic features, pose particular challenges. In conclusion, it is argued that full automation of audiobook creation for the languages of Russia's peoples using current speech synthesis technology is currently unfeasible. Developing audiobooks in such languages requires the prior creation of specialized linguistic resources. A necessary condition is the formation of a parallel corpus of texts and audio recordings produced by native speakers. Therefore, the successful implementation of such projects demands significant preliminary work on compiling training datasets and adapting algorithms to the specific features of each language.

Key words: minority languages, low-resource languages, machine learning, text recognition, speech synthesis, recurrent neural networks

Authors' contribution: Pozhidaev M.S. — developing the vision, defining and establishing key goals and objectives, and conducting the study, approving the final version of the research results; Teplykh E.S. — preparing and editing the text; Danilov S.I. — preparing the language examples and linguistic editing of the text. All authors approved the final version of the article.

Conflict of interest. The authors declare no conflicts of interest.

Article history: received 29 June 2025; accepted 10 September 2025.

For citation: Pozhidaev, M.S., Teplykh, E.S., & Danilov, S.I. (2025). Production of audiobooks in the languages of the peoples of Russia using speech synthesizers: problems and prospects. *Macrosociolinguistics and Minority Languages*, 3(2), 131–145. (In Russ.). <https://doi.org/10.22363/2949-5997-2025-3-2-131-145> EDN: HHYPZI

Введение

Развитие рекуррентных нейронных сетей, а с 2017 г. и моделей на базе Трансформера (Tosun, Dincer, 2018) привело к появлению новых синтезаторов речи, обеспечивающих качество аудиозаписей, все больше похожих на чтение диктором-человеком (Li et al., 2019). Под синтезатором речи понимается программа, «которая преобразует печатный текст в звучащую речь»¹. Примерами синтезаторов речи служат такие программы, как *Zvukogram*², *Yandex SpeechKit*³, *Natural Reader*⁴, *Eleven Labs*⁵, *Minimax*⁶ и др. На сегодняшний день данная технология используется во многих сферах: в разработке голосовых помощников и навигационных систем, в озвучивании видеоигр, мобильных приложений, рекламы и инструкций, в записывании автоматических ответов на телефонные звонки в различных сервисах и магазинах, в создании аудиокниг, электронных курсов (в создании диалоговых тренажеров для имитации общения с виртуальным собеседником, например, в конструкторе курсов *iSpring Suite*⁷), систем автоматизированного и машинного перевода, тифломаршрутов (Алюнина, 2021: 13; Алюнина, 2025: 67; Zheng et al., 2020; Arulprakash et al., 2023; Воркунова, Кисиева, Наумова, 2025: 116) и др. Несмотря на высокое технологическое качество, современные синтезаторы речи пока не способны стать полноценной заменой естественной речи. Тем не менее, у них есть ряд несомненных преимуществ, к числу которых относятся, например, возможность выбора желаемых голоса и языка, более высокая производительность, значительно превосходящая работоспособность диктора-человека. Также в долгосрочной перспективе можно проследить экономичность использования синтезаторов речи, разработанных под определенные задачи для автоматического создания аудиоряда без необходимости нанимать дикторов и актеров озвучивания.

¹ Синтез речи, или Text-to-Speech (TTS). URL: <https://www.logrusit.com/ru/services-and-solutions/creative-services/design-and-multimedia/text-to-speech/> (дата обращения: 23.08.2025).

² Zvukogram. URL: <https://zvukogram.com/> (дата обращения: 23.08.2025).

³ Yandex SpeechKit. URL: https://yandex.cloud/ru/services/speechkit?utm_referrer=https%3A%2F%2Fwww.google.com%2F (дата обращения: 23.08.2025).

⁴ Natural Reader. URL: <https://www.naturalreaders.com/online/> (дата обращения: 23.08.2025).

⁵ Eleven Labs. URL: <https://unitool.ai/ru/elevenlabs> (дата обращения: 23.08.2025).

⁶ Minimax. URL: <https://www.minimax.io/audio/text-to-speech> (дата обращения: 24.08.2025).

⁷ iSpring Suite. URL: <https://www.ispring.ru/ispring-suite> (дата обращения: 24.08.2025).

Сложившаяся ситуация фактически означает появление новой модели работы по созданию аудиокниг, включая книги на редких языках, к которым относятся языки народов России (ЯНР). Само по себе создание аудиокниг на ЯНР является актуальной практикой на сегодняшний день. Примерами тому служат следующие аудиоиздания и инициативы:

- аудиокнига «Nel'ĭ dölod» (рус. «Четыре ветра») ⁸ — сказки для детей на вепском языке;
- аудиокниги на чувашском языке, среди которых притча «Вёсекен кўлĕ» (рус. «Летающее озеро»), рассказ «Кăтра хёвелсаврăнăш» (рус. «Кудрявый подсолнушек»), аудиоиздание «Ачасем валли Тăван сёршывăн аслă вăрси сĕнчен вулатпăр» (рус. «Читаем детям о Великой Отечественной войне») ⁹;
- стихи на чувашском языке «Хаваслă карусель» (рус. «Веселая карусель»), «Кивĕ пушмак» (рус. «Старый ботинок»), «Кам-ши тётĕ пăлтăрта?» (рус. «Кто там в темноте»), «Хăюллă Якур» (рус. «Смелый Егорка») ¹⁰ и др.;
- аудиокниги для незрячих на 10 языках коренных народов Башкортостана ¹¹;
- подкасты на удмуртском языке «Удмурт литература», «Ныло-пиё» (подкаст о родительстве на удмуртском языке) и на татарском «Икенче дэүләт теле — Второй государственный» (о татарском языке как втором родном) ¹²;
- аудиокниги на бурятском языке, среди которых эпос «Шоно баатар» — легенда о Шоно Баторе, рассказы Г.Д. Дамбаева «Эжын хоёр» (рус. «Двое у матери») и Д.О. Батожабая «Төөригдэһэн хуби заяан» (рус. «Похищенное счастье»), переводы произведений А.С. Пушкина на бурятский язык («Санаартан ба тэрэнэй хүлһэшэн Балдаа тухай үльгэр» — русс. «История о попе и работнике его Балде»; «Загаһашан ба загаһан тухай үльгэр» — рус. «Сказка о рыбаке и рыбке»), стихи бурятских поэтов и детские сказки ¹³, а также романы Ж. Тумунова «Нойrhoо һэриһэн тала» (рус. «Степь проснулась») и Д.О. Батожабая «Уулын бүргэдүүд» (рус. «Горные орлы») ¹⁴.

Данная тенденция отвечает концепции государственной политики Российской Федерации в области исторического просвещения и сфере поддержки и популяризации языков и культур народов страны, в т.ч. благодаря созданию

⁸ «Nel'ĭ dölod» (рус. «Четыре ветра»). URL: <https://arctic-children.com/article/itti-totti-i-ne-tolko/> (дата обращения: 08.09.2025).

⁹ Аудиокниги Национальной библиотеки Чувашской Республики. URL: http://www.nbchr.ru/index.php?option=com_content&view=article&id=13730&Itemid=484 (дата обращения: 08.09.2025).

¹⁰ Стихи на чувашском языке. URL: http://nbchr.ru/virt_potomkam/gordeeva.htm (дата обращения: 08.09.2025).

¹¹ Аудиокниги для незрячих на языках коренных народов выпускают в Башкортостане. URL: <https://nazaccent.ru/content/30317-audioknigi-dlya-nezryachih-na-yazykah-korennyh/> (дата обращения: 09.09.2025).

¹² Как устроены подкасты на языках народов России. URL: <https://mastery.academy/local-podcasts/> (дата обращения: 09.09.2025).

¹³ Аудиокниги на бурятском языке на сайте Soyol.Ru (виртуальное пространство о культуре, искусстве и жизни Республики Бурятия): URL: <https://soyol.ru/tag/?q=Аудиобиблиотека> (дата обращения: 10.09.2025).

¹⁴ Проект «Родные голоса» ООО «Агинское библиотечное общество» и ГУК «Агинская краевая библиотека им. Ц. Жамцарано». URL: <https://agalibr.ru/rodnye-golosa/> (дата обращения: 15.09.2025).

произведений культуры на ЯНР (Указ Президента РФ от 8 мая 2024 г. № 314¹⁵, 2024: 5; Распоряжение Правительства РФ от 12 июня 2024 г. № 1481-р¹⁶, 2024: 19).

Несомненно, что цифровизация книг на ЯНР с использованием синтезаторов речи положительно влияет на распространение образовательного и просветительского контента на ЯНР, а благодаря автоматизации ряда этапов обработки текста снижаются временные затраты на их производство. Известно также, что создание аудиокниги с использованием синтезатора речи проходит ряд этапов. Их набор постепенно стабилизируется, и в нем можно выделить как инвариантные, не зависящие от языка исходного материала, так и требующие участия человека и поиска отдельных специфических решений для каждого нового языка операции. Ко вторым можно отнести разработку датасета для обучения алгоритма озвучивать тексты на том или ином языке, что, в свою очередь, требует предварительного составления параллельного корпуса размеченных письменных текстов и их аудиовариантов, озвученных носителями соответствующего языка. Это существенно затрудняет полную автоматизацию изготовления аудиокниг особенно на так называемых малоресурсных языках, под которыми принято понимать языки «с небольшими корпусами данных, ограниченными или отсутствующими аннотированными данными, малым количеством носителей языка, а также языки, находящиеся под угрозой исчезновения, и языки с нестабильной орфографией на ранних стадиях своего развития» (Дрожащих, Ефимова, 2025: 304).

Мы поставили задачу проанализировать основные этапы создания аудиокниг с использованием синтезаторов речи с целью выделения трудностей, возникающих при создании аудиоконтента на языке, который по разным причинам может быть отнесен к категории малоресурсного.

Типовой алгоритм разработки аудиокниги

В процессе синтезирования аудиофайла, к которому относится аудиокнига, выполняется несколько этапов (Mache, Baheti, Namrata Mahender, 2015: 54; Tan et al., 2021: 4–5), среди которых можно выделить следующие:

- 1) предварительная обработка исходных материалов;
- 2) определение иерархической структуры издания;
- 3) аннотирование и расшифровка текста;
- 4) синтезирование речи;
- 5) упаковка материала и оформление комплектов.

Рассмотрим подробнее названные выше этапы создания аудиокниги с использованием синтезаторов речи.

¹⁵ Указ Президента Российской Федерации от 8 мая 2024 г. № 314 «Об утверждении Основ государственной политики Российской Федерации в области исторического просвещения». URL: <http://www.kremlin.ru/acts/bank/50534> (дата обращения: 13.09.2025).

¹⁶ Распоряжение Правительства Российской Федерации от 12 июня 2024 № 1481-р «Об утверждении Концепции государственной языковой политики Российской Федерации». URL: <http://publication.pravo.gov.ru/document/0001202406140048> (дата обращения: 13.09.2025).

Инвариантные этапы разработки аудиокниги

Этапы 1, 2 и 5 предполагают проведение преимущественно технических работ, содержащих очень небольшое количество операций, связанных с культурными и языковыми особенностями печатного исходника и итогового аудиального издания. Предварительная обработка исходных материалов (этап 1) требуется в силу широкого разнообразия форматов текста для будущей аудиокниги. Если исходные материалы представлены файлами типа doc. или docx., то из них должны быть удалены сноски, гиперссылки, подписи к иллюстрациям и таблицам, любые примечания (особенно со ссылками на номера страниц) и подобные операции. Неудобен и неудачен формат PDF, который содержит текст в виде последовательности символов с указанием их координат. Это приводит к утрате информации о структуре документа (разделение на абзацы, выделение нумерованных и ненумерованных списков и т.д.). Из подобного файла можно получить материал только путем его обработки с использованием различных технологий распознавания текста¹⁷, что нередко приводит к появлению многочисленных ошибок и требует тщательной ручной проверки.

Не все тексты могут быть одинаково успешно переведены в формат аудиокниги. Очевидно, что художественная литература и учебные материалы по гуманитарным дисциплинам значительно лучше приспособлены к восприятию на слух, чем по физико-математическому профилю. Обилие иллюстраций, таблиц, уравнений, формул, графиков, схем и иных элементов, предназначенных для сугубо визуального восприятия, понижает потенциал издания для получения аудиокниги на его основе.

Указанные проблемы характерны для всех аудиокниг, вне зависимости от языка, алфавита и других национальных особенностей. Поэтому при выборе книги для перевода в аудиоиздание необходимо принимать во внимание ее исходный текстовый формат, способы передачи информации в ней (наличие таблиц, иллюстраций, графиков и под.), их информационную нагрузку и значимость, а также цель заказчика и потребности целевой аудитории, которые определяют утилитарность того или иного аудиоиздания.

Определение иерархической структуры издания

Следующий этап — определение правильной рубрикации издания. Определение рубрикации представляет собой не самую очевидную сложность, которая, тем не менее, требует пояснения. Она связана с тем, что названия заголовков содержат краткие формы нумерации структурных элементов печатного изда-

¹⁷ Наиболее простым примером распознавания текста в PDF-документе является распознавание с помощью различных нейросетей, как GigaChat (<https://giga.chat/>), который по соответствующему промпту перепечатывает текст с PDF-страницы и выдает его в виде предложений, доступных для ручного редактирования.

ния. Примером могут быть заголовки: на русском языке «11. Древняя история»; башкирском — «1. Баксасы белешмәһе» (рус. *садоводство*); якутском — «II. Айылҕа харыстабыла» (рус. *охрана природы*); бурятском языке — «3. Ургамал» (рус. *растения*). Для читателя использование нумерации структурных элементов в тексте привычно и интуитивно понятно, но синтезатору речи нужно «объяснить», как должен быть прочитан тот или иной номер, та или иная цифра. Например, заголовок «11. Древняя история» может быть прочитан синтезатором речи, как «Одиннадцать. Древняя история», «Одиннадцатая глава. Древняя история», «Глава одиннадцать. Древняя история». Помимо глав в русском языке в зависимости от структурирования издания допускается использование таких слов, как *книга, том, раздел, подраздел, стих, параграф, часть* и др. Разделы могут быть вложенными, причем встречаются издания, в которых в разных частях книги используется разный уровень вложенности. Используются как арабские, так и римские цифры. Таким образом, определение типов структурных элементов издания и установление между ними иерархических отношений происходит после целостного анализа материала, подлежащего озвучиванию.

С учетом названных особенностей определение правильной рубрикации и достаточно трудно автоматизировать. В этой связи на данном этапе видится затруднительным отказаться от ручного предредактирования текста для его подготовки к машинному озвучиванию с использованием синтезатора речи.

Аннотирование текста и расшифровка аббревиатур

В использовании термина *аннотирование* мы будем отталкиваться от корпусной лингвистики и понимать его как приписывание вспомогательной лингвистической информации «всем единицам выбранного уровня» (Алюнина, 2025: 96). В корпусной лингвистике к такой информации относятся: морфологические признаки определенной лексической единицы (род, число, падеж, время); жанрово-стилистические параметры текста; просодическая разметка, как в Акцентологическом корпусе¹⁸ Национального корпуса русского языка; прагматические маркеры определенной реплики, как, например, в корпусе устной речи Один речевой день¹⁹ (маркеры аппроксиматор, рефлексив, хезитатив, ксенопоказатель, метакоммуникатив, ритмообразующий маркер и др.). В нашем случае, говоря о расшифровках, мы будем иметь в виду расшифровку аббревиатур и сокращений с учетом их положения в контексте и, следовательно, корректного прочтения в соответствии с морфосинтаксическими и грамматическими нормами языка. Проиллюстрировать суть этого преобразования можно следующими примерами.

¹⁸ Акцентологический корпус // Национальный корпус русского языка. URL: <https://ruscorpora.ru/corpus/accent> (дата обращения: 12.09.2025).

¹⁹ Один речевой день. URL: <https://ord.spbu.ru/> (дата обращения: 13.09.2025).

Сокращение не может быть прочитано отдельно стоящими буквами, как оно представлено в письменной/печатной форме. В случае с аббревиатурами: *ВТО, МКС, ДЗ, ВВП* и др. — орфоэпической нормой является [вэ тэ о], [эм ка эс], [дэ зэ], [вэ вэ нэ], а не *вто, мкс, дз, вви*, как это может быть прочитано неподготовленным синтезатором речи.

Здесь важно подчеркнуть, что аббревиатуры должны быть расшифрованы и записаны для синтезатора речи теми же алфавитными символами, которые встречаются в текстах, использованных для составления обучающего датасета. Так, в алфавите русского языка используются только кириллические символы, а в транскрипции встречаются элементы латиницы, последние не должны применяться в тексте, адаптированном для синтезатора речи, если данные элементы латиницы не были включены в обучающую базу алгоритма для озвучивания. Иными словами, при наличии в русскоязычном тексте, предназначенном для озвучивания синтезатором речи, аббревиатуры *BBC* от англ. *British Broadcasting Corporation* (*Британская вещательная корпорация*), которая формально не отличается от *BBC* от рус. *Военно-воздушные силы*, то первая аббревиатура должна быть вручную расшифрована для машинного считывания и записана как [би би си], чтобы исключить озвучивание как [вэ вэ эс], не подходящее контексту.

Такая запись, как «в X в. до н.э.», должна читаться как «в десятом веке до нашей эры». Когда аудиокнигу озвучивает человек, он выполняет необходимые преобразования самостоятельно. Однако у человека могут возникнуть трудности, если озвучиваемый текст принадлежит специализированной области знания. Эта проблема актуальна, например, для книг, в которых приводятся первоисточники исторических документов или фразы на иностранных языках, химические и математические формулы. Если подобный материал без предварительной ручной обработки озвучивает синтезатор речи, он читает текст буквально. Даже ударение не всегда проставляется верно, особенно если в работу попадает поэтический текст, в котором постановка ударения может намеренно отличаться от принятой нормы для соблюдения рифмы, как в примерах далее.

Пример 1

*Отдых напрасен. Дорога крута.
Вечер прекрасен. Стучу в ворота*²⁰.

Пример 2

*Под ружьём в глубоком сне,
И на спящем спит коне
Перед ней хорунжий сам;
Неподвижно по стенам
Мухи сонные сидят;
У ворот собаки спят;*²¹

²⁰ Александр Блок. Отдых напрасен. Дорога крута. URL: <https://www.culture.ru/poems/1924/otdykh-naprasen-doroga-kruta> (дата обращения: 15.08.2025).

²¹ Василий Жуковский. Спящая царевна (Сказка). URL: <https://www.culture.ru/poems/17875/spyashaya-carevna-skazka> (дата обращения: 15.08.2025).

Таким образом, любой текст перед озвучиванием синтезатором речи должен быть обработан вручную для дополнения специальными аннотирующими метками и раскрытия сокращений.

Построение необходимых аннотаций, дополняющих текст информацией для корректного машинного чтения, подразумевает обращение к методам компьютерной лингвистики, включая алгоритмы на основе искусственных нейронных сетей. Фактически все этапы обработки, приведенные ниже, в значительной степени обусловлены спецификой языка, на котором создается аудиокнига.

Обработка начинается с традиционных для компьютерной лингвистики этапов: сегментации и токенизации текста (Tan et al., 2021: 6–7, 16). Для обоих этапов требуется принимать во внимание языковые особенности. Сегментация — это разделение текста на синтаксически независимые фрагменты, между которыми допускаются семантические связи, но не синтаксические (Tan et al., 2021: 7). Токенизация заключается в разделении каждого такого сегмента на токены. Токеном является «минимальная единица вхождения в корпус, которая обычно считается от пробела до пробела» (Алюнина, 2025: 98). Токеном могут быть слова, знаки препинания, пробелы, числа и пр. На любом из двух этапов разбиение может проводиться различными способами, которые зависят от особенностей конкретного языка. К примеру, в русском языке символы «.», «!» и «?», являясь необходимыми маркерами границы предложений, не всегда используются в функции знаков препинания конца предложения. Например, при записи инициалом имени *В.С. Высоцкий* появление точек не подразумевает разделения на отдельные предложения. Неоднозначен и вопрос о том, считать ли последовательность «В.» одним токеном или двумя отдельными токенами — буква «В» и знак «.».

Подобные трудности, препятствующие ясному проведению сегментации и токенизации, могут варьироваться в зависимости от специфики языка, что обязательно следует учитывать на этапе планирования материалов для автоматизированного озвучивания.

На определенных этапах аннотирования возможны и трудности иного порядка, не связанные с сегментацией текста. Так, например, для русского языка выделим следующие проблемные категории для аннотирования:

- простейшие фиксированные сокращения;
- сокращения, которые могут читаться по-разному в зависимости от контекста;
- числа, записанные римскими и арабскими цифрами;
- определение положения ударения в омографах.

Первые три категории в целом универсальны и присутствуют в текстах на всех языках, в т.ч. и на ЯНР. Хотя, конечно, особенности, обусловленные как собственно структурной спецификой языка, так и его тяготением к тем или иным цивилизационным и культурным ареалам, неизбежны.

Четвертый тип обозначенных выше трудностей, связанный с определением ударения в омографах, должен представлять проблему для языков с нефиксированным разноместным ударением, к которому в полной мере принадлежит русский. Для языков с фиксированным ударением этот вопрос не столь актуален, если речь идет об исконной лексике. Так, бурятский язык (монгольская языковая семья) относится к языкам агглютинативного типа, в которых практически отсутствуют омографы (кроме заимствованных слов). Но у бурятского языка есть ряд фонетических особенностей, способных влиять на качество синтезированной речи, например: наличие долгих и кратких гласных, а также дифтонгов; слабая редукция гласных не первых слогов и проявление лабиального сингармонизма. Среди перечисленных особенностей наличие долгих и кратких гласных, а также дифтонгов может быть выделено как отдельная категория для последующего аннотирования.

Во всех типах проблемных ситуаций, приведенных выше, обработку сегмента/токена следует проводить путем комбинации формальных и нейросетевых подходов. Применение формальных методов обработки позволяет, во-первых, произвести первоначальную классификацию фрагментов для аннотирования и, во-вторых, сразу произвести аннотирование фрагментов, обработка которых не требует применения нейросетевых алгоритмов. К таким фрагментам относятся разного рода сокращения «г.» (*год* или *город*), «до н.э.», «и под.», «ЯНР» (*языки народов России*), «КМНР» (*коренные малочисленные народы России*), «P.S.», «etc.» и пр. В большинстве случаев подобную обработку можно проводить с использованием той или иной реализации недетерминированного конечного автомата²², потому что какая-либо более сложная грамматика потенциально может приводить к неоправданно большим временным затратам при работе. Список сокращений и текста для их замены составляется вручную и сохраняется в виде словаря. Если говорить об особенностях национальных языков, то в зависимости от степени их изученности подобный словарь может уже существовать. В случае же, если он не существует, возникает необходимость его составления. Так, например, отсутствует специализированный словарь сокращений бурятского языка, но есть отдельный раздел в Толковом словаре бурятского языка²³, где собраны как специальные сокращения: «н. п.» — *нэрын надеж* (рус. *именительный надеж*), так и общие: «г. м.» — *гэхэ мэтэ* (рус. *и так далее, и тому подобное*), «ж.» — *жэл* (рус. *год*), «з. ж.» — *зуун жэл* (рус. *век*).

Среди задач, которые требуют применения нейросетевых алгоритмов и более глубоких знаний о языке, можно выделить задачу распознавания семантиче-

²² Недетерминированный конечный автомат — это автомат, принцип работы которого состоит в том, что «любой его переход единственным образом определяется по текущему состоянию и входному символу; чтение входного символа требуется для каждого изменения состояния. Регулярным языком, или языком, распознаваемым автоматом, называется множество таких слов, после прочтения которых автомат из начального состояния q_0 попадает в конечное из множества F » (Лобарёв, Лобарёв, 2025: 10).

²³ Толковый словарь бурятского языка. URL: <https://edbl.ru/pometry-i-sokrashheniya/> (дата обращения: 15.08.2025).

ского и грамматического контекста. Решение этой задачи актуально и является ключевым вопросом при производстве аудиокниг на любом языке, поскольку во многом к нему сводится обработка слов-омографов и чтения чисел. Например, правильное определение семантического контекста позволяет производить расстановку ударения у части разноударных омографов, к числу которых в русском языке относятся *до́роги* и *до'роги*, *сто'ите* и *стои'те*, *о'дин* и *оди'н*, *мука'* и *му'ка*. В бурятском языке данная операция тоже необходима, но только при расстановке ударений в заимствованных из русского языка словах-омографах (*а'тлас* и *атла'с*), поскольку, как уже было сказано выше, в бурятском языке отсутствуют омографы, а тоническое ударение всегда падает на последний слог.

В бурятском языке долгота и краткость гласных выполняют смыслообразительную функцию: *үүдэн* (*дверь*) и *үдэн* (*перо*), *зуун* (*сто*) и *зун* (*лето*), *уула* (*гора*) и *ула* (*подошва*), *хирэ* (*размер*) — *хирээ* (*ворон*). Но необходимо учитывать, что произношение некоторых заимствованных слов может отличаться от норм, например, *англи хэлэн* (рус. *английский язык*) — в слове *англи* звук *а* произносится как двойная гласная и произносится как *аангли*.

Под определением грамматического контекста подразумевается задача определения ряда атрибутивных форм (например, падежных), числа и т.п. в тех случаях, когда это невозможно сделать по словарю, на котором осуществляется обучение синтезатора речи. Эта задача возникает также при обработке омографов (в русском языке *ма'стера* и *мастера'*), а также для корректного чтения чисел: в русском языке — «Он получил вознаграждение в размере 1 тыс. руб.» и «Его вознаграждение составляет 1 тыс. руб.»; в бурятском языке — *ср. 50* (*табин*) и *50-й* (*табидахи*) — «Мой брат носит 50 размер одежды» (Минии аха *табидахи хэмжээнэй хубсаһа үмдэдэг*).

Для русского языка до сих пор невозможно создать алгоритм, позволяющий точно обрабатывать подобные контексты. Можно говорить только о некотором наборе нейросетевых моделей, показывающих сравнительно неплохой результат. Они основаны как на простых конфигурациях нейронных сетей, состоящих только из линейных ячеек, так и на более сложных, предполагающих использование рекуррентных ячеек и долгой краткосрочной памяти. В случае решения аналогичных задач для ЯНР возможно применение этих же технологий, но потребуется их обучение, которое, вероятно, создаст необходимость поиска или сбора обучающей базы — текстов и соответствующих им аудиофайлов — на определенном языке.

Синтезирование аудиоданных и упаковка

Непосредственно синтезирование аудиосигнала, которое составит основное содержание аудиокниги, выполняет синтезатор речи. Программа принимает на вход аннотированный текст и преобразует его в аудиоматериал. При этом

в синтезированном аудиофайле должны быть соблюдены нормы просодии соответствующего языка и тембральные характеристики определенного голоса: ритм, ударение, интонация, длина и высота звука, тон и др. (Пунегова, 2025: 82; Tan et al., 2021: 6–7).

Несмотря на то, что синтезаторы речи, очевидно, разрабатываются для каждого языка отдельно, степень проникновения особенностей того или иного языка в ключевые технологии не такая глубокая, как это может показаться на первый взгляд.

Наивысшее качество позволяют получить технологии, построенные на основе интеллектуальных алгоритмов. В частности, активно используются рекуррентные нейронные сети и модели, основанные на Трансформере. Примером подобных моделей может служить модель Tacotron (Wang et al., 2017). Синтезаторы на основе интеллектуальных алгоритмов проходят процедуру обучения, в ходе которой производят обработку большого массива записей диктора с параллельной обработкой текста, прочитанного диктором (Mache, Baheti, Namrata Mahender, 2015: 56; Tan et al.: 8, 21). В ходе обработки интеллектуальный алгоритм производит глубокое сопоставление последовательности символов, составляющих алфавит языка, и сигнала, соответствующего этой последовательности (Tan et al., 2021: 7). После того, как необходимые закономерности установлены, алгоритм способен производить построение сигнала по любой предоставленной последовательности символов алфавита — синтезировать речь.

Из описания технологии следует, что для получения синтезированной речи требуется произвести обучение, для чего нужно подготовить достаточное количество качественных записей диктора. Обучение алгоритма полностью механизировано и не требует участия человека. Подобная работа требует сравнительно небольших усилий и легко выполняется для различных языков, но нередко предполагает доступность значительных вычислительных ресурсов.

Процесс сжатия аудиоданных и оформление комплекта аудиокниги не подразумевает учета каких-либо специфических знаний о языке. Единственный вопрос, который следует упомянуть, — кодировка символов в тегах файлов и текстовых документах, если аудиокнига создается в специализированном формате, требующим предоставления исходного текста.

Заключение

Таким образом, разработка аудиокниг с использованием синтезаторов речи на малоресурсных языках (в т.ч. некоторых ЯНР), требует значительной предварительной подготовки алгоритмов, с помощью которых планируется синтезирование аудиофайлов. Принципиальную значимость имеет

создание датасета, на котором будет происходить обучение алгоритма. Таким датасетом должен являться параллельный корпус письменных и озвученных текстов на соответствующем ЯНР. Поскольку целевая аудитория аудиокниг на ЯНР — их носители, в базу названного параллельного корпуса требуется включать образцы звучащей речи, созданные носителями определенного языка. Предположительно создание такой базы может происходить тремя путями: 1) создание аннотированного текста к существующим аудио- и видеозаписям на ЯНР (интервью, радиозаписи, голосовые сообщения в мессенджерах и др.); 2) создание аудиозаписей к существующим текстам посредством озвучивания дикторами-людьми; 3) формирование обучающей базы на специально подготовленных текстах для тренировки алгоритма создавать аудиокниги на определенную тематику, например, детские аудиокниги, которые должны быть озвучены разными голосами, некоторые из которых могут быть голосами животных и вымышленных существ.

Список литературы

- Алюнина Ю.М. «Геометрия по-русски»: организация учебного материала в электронном курсе по научному стилю речи // Современный русский язык: функционирование и проблемы преподавания: Вестник. XXVI Международная научно-практическая конференция, Будапешт, 14 мая 2021 года. Т. 35 / под ред. А.А. Уразбековой, Ю.М. Алюниной, А.С. Васильевой, В.В. Самсоновой, Е.С. Седовой, Т.А. Сиротиной. Будапешт : Российский центр науки и культуры в Будапеште, 2021. С. 7–17. EDN: UCTJWX
- Алюнина Ю.М. Цифровые технологии в переводе. СПб. : Лань, 2025. 144 с.
- Воркунова И.О., Кисиева А.А., Наумова А.А. Редактирование как один из основных этапов составления тифломаршрута // Теория и практика составления тифломаршрутов для навигации лиц с нарушением зрения на станциях метрополитена : монография / под ред. А.В. Козуляева. Казань : Бук, 2025. С. 112–116. EDN: EUWYYO
- Дрожжих Н.В., Ефимова Е.В. Лемматизация малоресурсных языков в диахронической лингвистике: проблемы и решения // Известия Российского государственного педагогического университета РГПУ им. А.И. Герцена. 2025. № 217. С. 302–311. <https://doi.org/10.33910/1992-6464-2025-217-302-311> EDN: AKDLRR
- Лобарёв Д.С., Лобарёв Н.Д. Синтез недетерминированных конечных автоматов по регулярным выражениям алгоритмом Глушкова в формате JFF // Вестник Полоцкого государственного университета. Серия С. Фундаментальные науки. 2025. № 1 (44). С. 9–13. <https://doi.org/10.52928/2070-1624-2025-44-1-9-13> EDN: TEVIHV
- Пунегова Г.В. Тембральные характеристики голоса персонажа (на примере прозаических произведений коми писателей) // Вестник урovedения. 2025. Т. 15. № 1 (60). С. 80–89. <https://doi.org/10.30624/2220-4156-2025-15-1-80-89> EDN: EMPANK
- Arulprakash A., Synthiya M., Vijila T., Rajabhusanam C. Tamil speech synthesizer app for android: text processing module enhancement // Indian Journal of Science and Technology. 2023. Vol. 16. № 7. P. 485–491. <https://doi.org/10.17485/IJST/v16i7.2165> EDN: ZDIRTC
- Li N., Liu S., Liu Y., Zhao S., Liu M. Neural speech synthesis with transformer network // Proceedings of the AAAI Conference on Artificial Intelligence. 2019. Vol. 33. № 01. P. 6706–6713. <https://doi.org/10.1609/aaai.v33i01.33016706>
- Mache S.R., Baheti M.R., Namrata Mahender C. Review on text-to-speech synthesizer // International Journal of Advanced Research in Computer and Communication Engineering. 2015. Vol. 4. № 8. P. 54–59. <https://doi.org/10.17148/IJARCC.2015.4812>
- Tan X., Qin T., Soong F., Liu T.-Y. A survey on neural speech synthesis // arXiv. 2021. <https://doi.org/10.48550/arXiv.2106.15561>

- Tosun M., Dincer K. Determination of sound transmission loss in lightweight concrete walls and modeling artificial neural network // Selçuk Üniversitesi Mühendislik Bilim Ve Teknoloji Dergisi. 2018. Vol. 6. № 3. P. 461–477. <https://doi.org/10.15317/Scitech.2018.145>
- Wang, Y., Skerry-Ryan, R.J., Stanton, D., Wu, Y., Weiss, R.J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Zh., Bengio, S., Le, Q., Ajiomyrgiannakis, Y., Clark, R., Saurous, R.A. Tacotron: Towards End-to-End Speech Synthesis // arXiv:1703.10135. 2017. <https://doi.org/10.48550/arXiv.1703.10135>
- Zheng Y., Li X., Xie F., Lu L. Improving end-to-end speech synthesis with local recurrent neural network enhanced transformer // ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona : ICASSP, 2020. P. 6734–6738. <https://doi.org/10.1109/ICASSP40776.2020.9054148>

References

- Alyunina, Yu.M. (2021). «Geometry in Russian»: online course on Russian language for specific purpose. In A.A. Urazbekova, Yu.M. Alyunina, A.S. Vasilieva, V.V. Samsonova, E.S. Sedova, T.A. Sirotina, *Modern Russian Language: Functioning and Teaching Problems: Bulletin. XXVI International Scientific and Practical Conference, Budapest, May 14, 2021. Volume 35. [Sovremennyyi russkii yazyk: funktsionirovanie i problemy prepodavaniya: Vestnik. XXVI Mezhdunarodnaya nauchno-prakticheskaya konferentsiya, Budapesht, 14 maya 2021 goda. Tom 35]*. Budapest: Russian Center for Science and Culture in Budapest Publ. P. 7–17. (In Russ.). EDN: UCTJWX
- Alyunina, Yu.M. (2025). *Tsifrovye tekhnologii v perevode [Digital technologies in translation]*. Lan' Publ. (In Russ.).
- Arulprakash, A., Synthiya, M., Vijila, T., & Rajabhusanam, C. (2023). Tamil speech synthesizer app for android: Text processing module enhancement. *Indian Journal of Science and Technology*, 16(7), 485–491. <https://doi.org/10.17485/IJST/v16i7.2165> EDN: ZDIRTC
- Drozashchikh, N.V., & Efimova, E.V. (2025). Lemmatization of low-resource languages in diachronic linguistics: problems and solutions. *Izvestia: Herzen University Journal of Humanities & Sciences*, (217), 302–311. (In Russ.). <https://www.doi.org/10.33910/1992–6464–2025–217–302–311> EDN: AKDLRR
- Li, N., Liu, S., Liu, Y., Zhao, S., & Liu, M. (2019). Neural speech synthesis with transformer network. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 6706–6713. <https://doi.org/10.1609/aaai.v33i01.33016706>
- Lobaryov, D.S., & Lobaryov, N.D. (2025). Synthesis of nondeterministic finite automaton from regular expressions by Glushkov's algorithm in JFF format. *Herald of Polotsk State University. Series C. Fundamental Sciences*, (1), 9–13. (In Russ.). <https://doi.org/10.52928/2070–1624–2025–44–1–9–13> EDN: TEVIHV
- Mache, S.R., Baheti, M.R., & Namrata Mahender, C. (2015). Review on text-to-speech synthesizer. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(8), 54–59. <https://doi.org/10.17148/IJARCC.2015.4812>
- Punegov, G.V. (2025). Timbral characteristics of a character's voice (on the example of prose works by Komi writers). *Bulletin of Ugric Studies*, 15(1), 80–89. (In Russ.). <https://doi.org/10.30624/2220–4156–2025–15–1–80–89> EDN: EMPAHK
- Tan, X., Qin, T., Soong, F., & Liu, T.-Y. (2021). *A survey on neural speech synthesis*. arXiv:2106.15561v3. <https://doi.org/10.48550/arXiv.2106.15561>
- Tosun, M., & Dincer, K. (2018). Determination of sound transmission loss in lightweight concrete walls and modeling artificial neural network. *Selçuk Üniversitesi Mühendislik Bilim Ve Teknoloji Dergisi*, 6(3), 461–477. <https://doi.org/10.15317/Scitech.2018.145>
- Wang, Y., Skerry-Ryan, R.J., Stanton, D., Wu, Y., Weiss, R.J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Zh., Bengio, S., Le, Q., Ajiomyrgiannakis, Y., Clark, R., & Saurous, R.A. (2017). Tacotron: Towards end-to-end speech synthesis. arXiv:1703.10135. <https://doi.org/10.48550/arXiv.1703.10135>
- Vorkunova, I.O., Kisieva, A.A., & Naumova, A.A. (2025). Editing as one of the main stages of compiling a typhlo route. In Kozulaev, A.V. *Teoriya i praktika sostavleniya tiflomarshrutov dlya navigatsii lits s narusheniem zreniya na stantsiyakh metropolitena [Theory and practice of creating tiflo-routes for navigation of visually impaired people at metro stations]*. Kazan': Buk Publ. P. 112–116. (In Russ.). EDN: EUWYYO
- Zheng, Y., Li, X., Xie, F., & Lu, L. (2020). Improving end-to-end speech synthesis with local recurrent neural network enhanced transformer. In *ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona: ICASSP. P. 6734–6738. <https://doi.org/10.1109/ICASSP40776.2020.9054148>

Информация об авторах:

ПОЖИДАЕВ Михаил Сергеевич, кандидат технических наук, доцент кафедры теоретических основ информатики института прикладной математики и компьютерных наук, Национальный исследовательский Томский государственный университет, Российская Федерация, 634050, Томск, пр. Ленина, д. 36. *Научные интересы:* компьютерная лингвистика, языковые модели, операционные системы семейства UNIX и программная инженерия, медиадоступность.

E-mail: msp@luwrain.org

ORCID: 0000-0002-5006-5975

SPIN-код: 5459-0852

ТЕПЛЫХ Елена Сергеевна, психолог, младший научный сотрудник лаборатории междисциплинарных исследований, Национальный исследовательский Томский государственный университет, Российская Федерация, 634050, Томск, пр. Ленина, д. 36. *Научные интересы:* психология личности, компьютерная лингвистика, языковые модели, медиадоступность.

E-mail: elena@luwrain.org

ORCID: 0000-0002-1825-7379

SPIN-код: 7424-5366

ДАНИЛОВ Сергей Ильич, аспирант кафедры общего и русского языкознания филологического факультета, Российский университет дружбы народов, Российская Федерация, 117198, Москва, ул. Миклухо-Маклая, д. 6. *Научные интересы:* социолингвистика, когнитивная лингвистика, миноритарные языки, репрезентации языка, бурятский язык.

E-mail: 1042250116@rudn.ru

ORCID: 0009-0005-6954-6874

SPIN-код: 6954-5385

Bionotes:

Mikhail S. POZHIDAEV, Ph.D. in computer science, Associate Professor at the Department of Theoretical Foundations of Computer Science at the Institute of Applied Mathematics and Computer Science of the National Research Tomsk State University, 36 Lenin ave., Tomsk, 634050, Russian Federation. *Research interests:* computational linguistics, language models, UNIX family operating systems, software engineering, media accessibility.

E-mail: msp@luwrain.org

ORCID: 0000-0002-5006-5975

SPIN-code: 5459-0852

Elena S. TEPLYKH, psychologist, a junior researcher at the Laboratory of Interdisciplinary Research of the National Research Tomsk State University, 36 Lenin ave., Tomsk, 634050, Russian Federation. *Research interests:* personality psychology, computational linguistics, language models, media accessibility.

E-mail: elena@luwrain.org

ORCID: 0000-0002-1825-7379

SPIN-code: 7424-5366

Sergey I. DANILOV, PhD student at the Department of General and Russian Linguistics, Faculty of Philology at RUDN University, 6 Miklukho-Maklaya st., Moscow, 117198, Russian Federation. *Research interests:* sociolinguistics, cognitive linguistics, minority languages, language representations, Buryat language.

E-mail: 1042250116@rudn.ru

ORCID: 0009-0005-6954-6874

SPIN-code: 6954-5385


ЯЗЫК И КУЛЬТУРА КАК СИМВОЛИЧЕСКИЙ КАПИТАЛ LANGUAGE AND CULTURE AS A FORM OF SYMBOLIC CAPITAL

DOI: 10.22363/2949-5997-2025-3-2-146-165

EDN HJMYAG

Research article / Научная статья

Social and cultural restrictions for movie titles translation: why AI will never do it

Yulia M. ALYUNINA^{1,2}  , Elvira R. GANEEVA³ ¹RUDN University, *Moscow, Russian Federation*²Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) of the Russian Academy of Sciences (MAE RAS), *Saint Petersburg, Russian Federation*³Financial University under the Government of the Russian Federation, *Moscow, Russian Federation* alyunina.y@mail.ru

Abstract. Numerous studies have been devoted to localising film titles. Most of them focus on adapting the film title to the target language environment from linguistic and cultural points of view. That said, the influence of extralinguistic factors on translating film titles seems to be under-investigated. Meanwhile, various translation forums and cinema websites publish reviews written both by professionals and laymen who discuss possible reasons behind the choice of a film title. Nowadays the issue is being further complicated by the widespread development of AI technologies that lead to questioning the very need for human translation in some fields of knowledge, especially when it comes to small format texts such as a film title. Nevertheless, while localisation of a film through translating its title pertains to the area of marketing expertise, not translation per se, marketing rules have to be observed as well. And current machine translation systems are not equipped to perform these functions now. In this light this study aims to investigate the strategies of adapting English film titles to the Russian market considering the social and cultural aspects of localisation. This study is based on reviews of Russian streaming platforms and translation companies, which share their observations on strategies for translating English-language titles in the context of localising British and American films to the Russian market. This study is analytical and based on reviews of Russian streaming platforms and translation companies, which share their observations on film translation

© Alyunina, Yu.M., Ganeeva, E.R., 2025

This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

strategies for English films in the context of localizing British and American films for the Russian market. The methodological core of the study consists of empirical methods. The study identified 7 key localisation strategies: expanding the title, using memes, developing the story, adding slang, lowering speech register, adding sexual subtext, minimising the risk of politicising the title, expanding a well-known franchise.

Key words: machine translation, audiovisual translation, marketing translation, language and culture, intercultural cooperation, localization, cultural potential of language

Authors' contribution: Alyunina Yu.M. — developing the vision, defining and establishing key goals and objectives, and conducting the study, approving the final version of the research results; Ganeeva E.R. — preparing and editing the text in English. Both authors approved the final version of the manuscript.

Conflict of interest. The authors declare no conflicts of interest.

Article history: received 25 August 2025; accepted 13 October 2025.

For citation: Alyunina, Yu.M., & Ganeeva, E.R. (2025). Social and cultural restrictions for movie titles translation: Why AI will never do it. *Macrosociolinguistics and Minority Languages*, 3(2), 146–165. <https://doi.org/110.22363/2949-5997-2025-3-2-146-165> EDN: HJMYAG


Социокультурные ограничения перевода кинозаголовков: почему ИИ их не преодолеть

Юлия Матвеевна АЛЮНИНА^{1,2}  ,
Эльвира Рустемовна ГАНЕЕВА³ 

¹Российский университет дружбы народов, Москва, Российская Федерация

²Музей антропологии и этнографии имени Петра Великого (Кунсткамера) Российской академии наук (МАЭ РАН), Санкт-Петербург, Российская Федерация

³Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация

 alyunina.y@mail.ru

Аннотация. Особенности локализации кинозаголовков посвящено много исследований. Большинство из них фокусируется на лингвистической и культурной специфике адаптации названия фильма к стране принимающего языка. Значительно реже в научной литературе поднимается вопрос влияния экстралингвистических факторов на перевод кинозаголовка: исторического, социального, политического контекстов, а также условий рыночной экономики. В то же время на различных переводческих форумах, сайтах стриминговых платформ и онлайн-кинотеатров встречается немало количество обывательских и профессиональных обзоров тех причин, которые могут влиять на перевод названия аудиовизуального произведения. В современных реалиях вопрос перевода осложняется развитием технологий искусственного интеллекта, массовость которых ставит под сомнение необходимость человеческого перевода в некоторых областях знания, особенно когда речь идет о небольшом фрагменте текста, как, например, название фильма. Однако в силу

того, что локализация фильма посредством его названия является областью деятельности маркетологов, а не переводчиков, перевод кинозаголовка подчиняется не только лингвистическим законам, но и маркетинговым, которые современные машинные переводчики пока не способны соблюдать. С учетом социокультурной сути локализации кинозаголовка, а также в свете развития машинных переводчиков на примере переводов кинозаголовков с английского языка на русский рассмотрены стратегии их адаптации к принимающему русскоязычному рынку. Исследование носит аналитический характер и основано на обзорах российских стриминговых платформ и переводческих компаний, которые делятся своими наблюдениями о стратегиях перевода англоязычных фильмов в контексте локализации британских и американских фильмов для российского рынка. Методологическое ядро исследования составляют эмпирические методы. В результате исследования удалось выделить 7 основных стратегий локализации: расширение названия, использование мемов, развитие истории, добавление сленга, добавление сексуального подтекста, снижение риска политизированности, расширение франшизы.

Ключевые слова: машинный перевод, аудиовизуальный перевод, маркетинговый перевод, язык и культура, межкультурное взаимодействие, локализация, культурный потенциал языка

Вклад авторов: Алюнина Ю.М. — разработка концепции, формулировка и развитие ключевых целей и задач, а также проведение исследования, утверждение окончательного варианта результатов исследования; Ганеева Э.Р. — подготовка и редактирование текста на английском языке. Оба автора одобрили окончательную версию рукописи.

Заявления о конфликте интересов. Авторы заявляют об отсутствии конфликта интересов.

История статьи: получена 25 августа 2025; принята в печать 13 октября 2025.

Для цитирования: Alyunina Yu.M., Ganeeva E.R. Social and cultural restrictions for movie titles translation: why AI will never do it // *Macrosociolinguistics and Minority Languages*. 2025. Т. 3. № 2. С. 146–165. <https://doi.org/110.22363/2949-5997-2025-3-2-146-165> EDN: HJMYAG

Introduction

Since the beginning of the 21st century, globalization processes (also known as homogenization and standardization (Pooch, 2016: 24)) have become particularly relevant, defining trends in global development in many fields, from economics itself to cultural studies and linguistics (Featherstone, 1992; Lang, 2006; Pooch, 2016). These processes of globalization have generated interest in the scientific study of such issues as the dominance of certain cultures and languages in the world. These issues fit into the problematic of symbolic and cultural capital, which P. Bourdieu addressed back in the late 1990s, introducing the concept of the “language market” (Bourdieu 1982, 1986). Today, English is undoubtedly the leader in this market. Despite the fact that in the global information space “there is always translation, adaptation, or ‘indigenization’ of the receiving culture” (Pooch, 2016: 25), the dominance of English leads to a unification of the global semiotic field. One manifestation of this

unification is the widespread use of machine translation technologies, which are being implemented in many fields.

Machine translation technologies mostly operate in language pairs with English, so in rare language pairs, even machine translators automatically use English as the intermediary language. Removing human involvement from the translation process in some areas and relying on English as the intermediary leads to the loss of significant cultural meanings that only a human can perceive, since in different languages, “cultures may well travel and move around the world, but ethnicity is still about the maintenance of social boundaries, something which remains a powerful force in the current phase of globalization” (Featherstone, 1992: 12).

In 2022 a Google engineer Blake Lemoine claimed that LaMDA (Language Model for Dialogue Applications) — Google’s system for building chatbots based on its most advanced large language models — was sentient, i.e. it was capable of feeling empathy¹. That was the conclusion that Blake Lemoine arrived at after having in-depth conversations with LaMDA. Consciousness and sentience, when ascribed to a language model, mean that it can understand the feelings of a person with whom it is having a conversation, as well as interpret human emotions and therefore identify them. Large-scale implementation of such a technology into machine translation has the potential to revolutionise translation industry by transforming the role of a human translator. Post-editing machine translation may become his primary function not only in the areas where texts are predominantly clichéd and have well-established equivalents in the source language and the target language (such as legal documents, technical texts, manuals, etc.), but also in the spheres where texts are required to express multiple meanings and trigger an emotional response (Toral, Wieling, Way, 2018). The latter encompasses literary fiction (Toral, Wieling, Way, 2018; Matusov, 2019; Constantine, 2019), audiovisual products as well as marketing texts, including film titles.

A film title is not only the name of the film that gives an idea about its plot. They “are created to draw attention to a film” (Haidegger, 2015: 425) and to urge the viewer to watch the film, i.e. to pay for it (Bae, Kim, 2019: 100). It could be argued that a film title contains a manipulative element that encourages the target audience to make the choice that the producers want them to make. In other words, “title should function as a means to market a film successfully” (Haidegger, 2015: 426). Thus, linguistically speaking, everything about the film title is of high importance — the choice of lexical units with all the meanings and connotations, their grammatical form (Haidegger, 2015: 425), the connection between the film title and the plot. If one looks at a film title in the original as a black and white text, then one might consider its lexical and grammatical aspects in conjunction with the content of the film and the parameters of the target audience (age, gender, social or professional status, creed, preferences,

¹ Tikku, N. The Google engineer who thinks the company’s AI has come to life. Retrieved 15 August 2025, from <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>

etc.) to be the key means of attracting attention of the viewer. However, when it comes to the translated version of the film title, these factors may be insufficient to adapt it successfully to a foreign film market.

Since a film title's main function is to sell the film in a foreign market, it would not suffice to resort to interlingual translation without adapting it to the foreign culture. One should consider cultural features of the host culture as well as social and political context that will accommodate the film along with its title. Moreover, the title should be engaging and marketable (Krasina, Moctar, 2020: 293–294). Indisputably, these factors may vary depending on the language pairs involved in translation or, what is more important, on the two cultures and the differences that have to be reconciled in the process of film title adaptation. Thus, studies show that, for instance, when translating film titles from English into Thai, the dominating strategy is to create a new title in the Thai language that conveys the main idea of the film instead of translating the original one (Kettongma, 2024: 210), whereas, when translating from French into Russian, the priority is given to the adequacy of the title and its correspondence to the genre, plot and target audience of the film (Anissimov, Borissova, Konson, 2019: 453).

The findings of academic research lead us to the conclusion that it is difficult to compile an exhaustive list of all translation strategies used when adapting a film title in a particular language pair. If such a list could be compiled, it would be possible to identify patterns of translating film titles and use them as a foundation for training artificial intelligence to translate the titles of audiovisual works. However, scientific research into translation of marketing texts, including film title adaptation, and the practice of translation do not align. The translation process presupposes careful analysis of the factors that may influence translation decisions but escape the attention of a researcher who views a foreign-language film title as a result of translation and marketing efforts.

Thus, **this research focuses** on the study of translation strategies from a practical viewpoint, taking into account cultural, social and political contexts.

Methodology and Material

This article is of analytical nature, and it is primarily based on the observations derived from the practice of translating film titles, which can be found on the Internet in the public domain in large quantities. This is accounted for by the fact that scientific understanding of film title localisation strategies remains an open question for research in linguistics, and while seeking an answer one should start with marketing.

According to audiovisual translation practitioners, there are no available technologies that can translate a film title without referring to the content of the audiovisual work in question. The current machine translation systems render film titles as stand-alone sentences that are not contextualised. The context for a film title comprises not only the film itself, but also the target audience as a set of cultural and

social parameters. The cultural, historical or political context in which the film is released is also of great importance.

This article is based on reviews of Russian streaming platforms and translation companies², which share their observations on strategies for translating English-language titles in the context of localising British and American films to the Russian market.

But before discussing the process of translation and localisation that inherently requires human participation, it is necessary to review the strengths and weaknesses of machine translation diachronically.

AI Translation Algorithm and Where it Comes from

Artificial Intelligence technologies are far from being the first generation of machine and automated translation. The advent of machine translation dates back to the 50s of the 20th century, and the idea for this method of translation is rooted in the practical necessity to automate human labour. By that time the amount of scientific and technical information in the world that needed to be translated had been growing rapidly, and human translation resources were insufficient. Thus, people came up with the idea of automating translation process (Sin-wai, 2014; DeCamp, Zetsche, 2014).

Historically speaking, there have been five main approaches to machine translation.

Rule-based Machine Translation

This can be represented as follows: *dictionary + grammar = translation*. This type of machine translation functions according to the same principle as human translation. It relies heavily on detailed dictionaries and grammars of both languages involved in translation (Jones et al., 2012: 1364). The first machine translation systems were rule-based and were generally not highly successful. This was due to, firstly, low flexibility (rules and grammars were fixed, making it difficult to adapt to new language constructions and features); secondly, high labour intensity (creating and maintaining extensive rule sets required significant resources and time); thirdly, problems with lexical polysemy (polysemy of lexical units was a difficult task, as the choice of the correct meaning depended on the context, which was sometimes difficult to take into account in predefined rules) (Shiwen, Xiaojing, 2014).

² Zayacz, A. (2016, May 6). Kak lokalizuyut nazvaniya fil'mov v Rossii [How film titles are localised in Russia]. Film.ru. Retrieved 28 July 2025, from <https://www.film.ru/articles/kinoslovar-trudnosti-perevoda>

Trudnosti perevoda: pochemu russkii variant nazvaniya fil'ma chasto ne sootvetstvetuet originalu? [How film titles are localised in Russia. Difficulties of translation: why does the Russian version of a film title often not correspond to the original?]. Lexxis. Retrieved 28 July 2025, from <https://lexxis.ru/articles/trudnosti-perevoda-pochemu-russkiy-variant-nazvaniya-filma-chasto-ne-sootvetstvetuet-originalu/#close>

Darbinyan, A. (2020, December 2). Kak pravil'no lokalizovyvat' fil'my v Rossii? Kratkaya instruktsiya [How to localise films in Russia? Brief instructions]. Kinopoisk. <https://www.kinopoisk.ru/media/article/4003560/>

Blicz perevod. Pochemu perevody nazvanii fil'mov byvayut daleki ot originalov? [Why do translations of film titles tend to be far from the originals?]. Retrieved 28 July 2025, from <https://blitz-perevod.ru/articles/pochemu-perevody-nazvanii-filmov-byvaiut-daleki-ot-originalov>

20 fil'mov s nepravil'no perevedennymi nazvaniyami smotret' online [20 films with mistranslated titles]. @kino. Retrieved 28 July 2025, from https://kino.mail.ru/cinema/selection/2733_trudnosti_perevoda_filmi_s_nepravilnimi_nazvaniyami/

Example-based machine translation

This machine translation method presupposes finding similar examples of sentence pairs in the source language and the target language. Example-based machine translation relies on data from parallel language corpora, and examples are extracted from large sets of bilingual texts. The mechanism of this type of translation is as follows: the computer algorithm analyses the sentence to be translated → compares this sentence with the parallel corpus → identifies segments in the parallel corpus that are similar to the sentence to be translated → extracts the necessary components from the corpus and constructs a sentence in the target language (Tak-ming, Webster, 2014; Maučec, Donaj, 2020).

Statistical machine translation

It boils down to the following scheme: *parallel texts + statistics = translation*. Thus, this type of translation is based on statistics and machine learning. The computer is given a database of parallel texts that consists of sentences paired with their translation into a foreign language. Trained on parallel texts, the computer records how a certain lexical unit was translated in this or that language environment, and how frequently that occurred. This way it analyses the statistical data of translation. The larger the database of parallel texts is, the higher is the translation quality of the algorithm that was trained on the basis of these texts (Yang, Min, 2014: 202).

Hybrid machine translation

This combines the first and the third approaches mentioned above, so it could be represented as follows: *rules + statistics = translation*. The combination of the two approaches can be implemented in various ways, but the two most common variants are: (1) integration of a statistical module into translation, and (2) integration of rules into a statistical model (Qun, Xiaojun, 2014).

Neural machine translation

This type of machine translation uses artificial neural networks that imitate human neural networks. The main advantage of neural machine translation is its ability to account for contextual relationships and grammar. In contrast to statistical machine translation, neural machine translation ensures a higher degree of adequacy and equivalence, and the translation outcome appears more authentic. This result is attributed to the fact that neural networks analyse individual language units and whole sentences within a given context, rather than selecting a translation equivalent in isolation, based on the frequency of its occurrence in a similar linguistic environment (Koehn, 2020; Maučec, Donaj, 2020).

Back in the mid-2000s, many people made fun of how a computer was handling translation of texts and could scarcely believe that we would soon be questioning the

relevance of translation as a profession. However, the situation has changed dramatically since neural machine translation was introduced in 2016. Neural models are capable of taking into account the context and semantics of sentences, allowing them to produce more accurate and natural sounding translations as compared to earlier versions of machine translation. Presently, neural machine translation is considered to be the most successful software solution in translation industry in a variety of different subject areas. The quality of present-day machine translation raises the issue of preserving translation as a profession and drives various experiments that are, to some extent, competitions between a machine translation system and a human translator. Let us consider one of them.

In 2018, two years after the declared launch of neural machine translation, the technology started being actively adapted to translation tasks in different subject areas. *Intento*³, a company which specialises in implementing AI technologies into machine translation to run various international platforms, has evaluated the effectiveness of neural machine translation.

The *Intento* experiment involved translations from a neural network, a human translator and several machine translation systems. All participants of the experiment translated the same biomedical texts in the English-German language pair. Text fragments that were translated with significant differences were submitted for manual review to *Logrus IT*⁴ — a large international high-tech company that provides services of translation, localisation, testing and language consulting to the business sector. *Logrus IT* linguists received the source texts in English, their anonymised translations into German from 13 machine translation systems, as well as a human translation that was considered exemplary under the terms of the experiment. None of the experts knew which translation they were checking — a machine translation or a human translation⁵.

Logrus IT linguists used their own quality assessment methodology to analyse the translations. It included criteria such as the degree of translation adequacy, readability, terminological consistency with the subject matter of the text, style, as well as a scale of translation errors, which range in severity from critical to minor. A comparison of translations performed by a neural network, a human translator and several machine translation systems revealed that the human translation was not rated as exemplary by the experts.

In a sense, this experiment demonstrated that neural machine translation, i.e. a human-designed translation system, was able to outperform a human translator. One could say that the student has surpassed the master.

³ Enterprise Language Hub. Retrieved 28 July 2025, from <https://inten.to/>

⁴ *Logrus IT*. Retrieved 28 July 2025, from <https://www.logrusit.com/en>

⁵ Samsonova, V. (2018, November 26). *Kompaniya Logrus IT pouchastvovala v sravnenii sistem neironnogo mashinnogo perevoda [Logrus IT participated in a comparison of neural machine translation systems]*. *LogrusIT*. Retrieved 28 July 2025, from <https://www.logrusit.com/ru/blog/testing-nmt-engines/>

Although neural machine translation has made significant advances in recent years, it is still far from ideal. AI continues to stumble over the complexity of a natural language. The most challenging type of translation for the system is so far the translation of literary fiction. Only a human translator is capable of perceiving the meaning of prose and poetry that is concealed in a wide range of stylistic elements of a text. Machine algorithms often translate word for word, following the rules of language, i.e. translating words — the external linguistic form of a literary work, rather than sense — the inner semantic content of a literary text.

Automatically translating marketing texts, including film titles, presents another serious challenge to translation systems and it is closely linked to the machine learning technology. Contemporary machine translation algorithms are trained on the database of written texts. This means that they require a training database that the algorithm uses to record certain morphological, lexical, grammatical, syntactic patterns in multiple languages or in a certain language pair. Machine translation systems rely on these patterns to generate the most adequate version of translation, taking into account the formal aspects of the context that is analysed by AI. When it comes to the translation of a film title, such an analysis proves to be quite difficult, since its context encompasses not only the film or TV series, but also the culture of the country of localisation (Anissimov, Borissova, Konson, 2019; Kiran, 2023; Jiang, Yin, 2024), as well as the social and political context (Crane, 2014; Zhang, 2023). It entails ideological and censorship requirements that accompany the release of an audiovisual work in a foreign market.

In addition, an audiovisual work may be a film adaptation of a literary text of the same title (the play by Anton Chekhov *The Seagull* 1895 → TV drama *The Seagull* 2018) or of a different title (Anton Chekhov's short story *The Duel* 1891 → TV drama *The Duel* 2010), or it may be an adaptation of a work of fiction (the 2013 tragicomedy *The Double* based on Dostoyevsky's novella of the same title *The Double* (1846)). In both cases, the title of the film should serve as a clear reference to the title of the literary text.

To understand the film's semantic content, the marketing expert, who is translating the film title, should watch the film or at least read the script or the screenplay. Furthermore, the marketing expert/translator has to consider the distinctive features of the target audience for whom the film is made and, therefore, who the film's title should be addressed to.

Thus, it is only through an analysis of these factors — the social and political context, the culture of the country of localisation, the intertextual links of the audiovisual work, the target audience — that the most adequate decision can be made about the choice of one or another variant of the translation of the film title.

Naturally, one can assume that if modern AI technologies are capable of translating and dubbing audiovisual works without human participation (e.g. browser-based Yandex video translation⁶, machine translation platforms for dubbing and neural

⁶ Yandex browser translator. Retrieved 28 July 2025, from https://yandex.ru/project/browser/video_translation/m/03

network systems for voice-over such as Rask AI⁷, Veed⁸, HeyGen⁹, Kapwing¹⁰, AI Video Translator¹¹, Clideo¹² etc.), then neural machine translation will be able to easily ‘watch’ a film, analyse it, and offer several variants of translation of the film title taking into account the content of the film.

Nevertheless, neural machine translation does not go further than the audiovisual work itself when analysing the film. The neural machine translation process includes recognising diegetic speech → transcribing it → generating machine translation of the transcript → voicing of the translation → superimposing the voiced text on a video sequence (Martín-Mor, Sánchez-Gijón, 2016; Sulubacak, et al. 2020). When given a transcript of the speech in the original or in translation, a machine algorithm can analyse it and provide several variants of the title in the source language or the target language.

Why does this scenario of creating an original and/or translated film title seem questionable? Firstly, because when translating a film, the machine translates a written text with no reference to the film’s video sequence and its sound or music tracks, which also carry a semantic and emotional message. Hence the visual component (video sequence), audio component (music) and the plot line (conceptual content) of the film will also be disregarded in the process of creating the film title. Secondly, at the present stage of development of machine translation technology, neural machine translation does not provide an integrated option of analysing the social and political agenda and current affairs as well as the cultural features of the localisation market. AI technologies are trained on the databases that have already been created. Therefore, they do not reflect the reality in its dynamic aspect. Only a human translator can take into consideration the nuances of the ever-changing reality, selecting the most adequate strategy for translating a film title, while also considering all the necessary factors.

Movie Titles Translation Strategies

Having analysed strategies for translating film titles that streaming companies and translation agencies resort to in their practice (including social, political, and cultural localisation factors), we identified the following groups of marketing strategies for localising film titles.

Expanding a well-known franchise

Marketing experts claim that sometimes it is quite difficult to sell a film to the audience without a brand affiliation. Therefore, localising a film title may involve

⁷ Rask. Retrieved 28 July 2025, from <https://www.rask.ai/>

⁸ Veed. Retrieved 28 July 2025, from <https://www.veed.io/>

⁹ HeyGen. Retrieved 30 July 2025, from <https://www.heygen.com/translate>

¹⁰ Kapwing. Retrieved 30 July 2025, from <https://www.kapwing.com/tools/translate>

¹¹ Maestra. Retrieved 30 July 2025, from <https://maestra.ai/tools/video-translator>

¹² Clideo. Retrieved 30 July 2025, from <https://clideo.com/ru/video-translator>

a well-established franchise or film. Thus, for example, the 2019 horror film *Polaroid* was titled *Пункт назначения: Смайл* [Final Destination: Smile] in Russian. Another example of the same strategy is the 2019 feature film *Dark Phoenix* that was localised as *Люди Икс: Тёмный Феникс* [X-men: Dark Phoenix]. A literal translation of the title — *Dark Phoenix* — might seem recognisable only to the fans of the franchise, who are well-versed in the X-Men universe, but not to those who might have once watched one of the previous films.

It turned out that the Russian word *братва* [gang] was a popular choice for rendering titles of cartoons and films for children from English into Russian. After the 2004 cartoon *Shark Tale* hit the Russian big screen under the title *Подводная братва* [Underwater gang], other localised film titles started to feature the word *братва* [gang]: *Over the Hedge* (2006) → *Лесная братва* [Forest gang]; *Delhi Safari* (2011) → *Братва из джунглей* [Gang from the Jungle]; *Alpha and Omega* (2010) → *Альфа и Омега: клыкастая братва* [Alpha and Omega: fang gang]; *Joey and Ella* (2021) → *Прыгучая братва* [Bouncy gang]. As we can see, the original film titles were in no way related to each other and represented independent stories, not united by the same franchise, however, in the Russian language they seem to be artificially linked together during localisation process. This marketing move brought together films for children about animal adventures under the umbrella of the word *братва* [gang].

Thus, this strategy involves analysing already existing translated film titles that have similar genre and plot features and have been successfully localised in the host market in terms of reaching their target audience.

Minimising the risk of politicising the title

This strategy presupposes avoiding situations where viewers may see political subtext in the film title. This is the case when the word *America* ordinarily disappears from the localised title when the film is released in Russia. For instance, the film, well-known to the Russian audience as *Первый мститель* [The first avenger], is originally called *Captain America. The First Avenger* (2011); while *American Sniper*, released in 2014, was distributed in Russian cinemas as *Снайпер* [Sniper], losing its American affiliation.

Another case to support this strategy is the film title *Siberia* (2018), which originally was accompanied by a slogan *A dangerous land breeds a dangerous man* (Fig. 1). The film tells the story of an American gem dealer who travels to Siberia to make a deal. It was released in Russia under the title *Профессионал* [Professional] (Fig. 2), while the slogan was adapted to fit the context *Один против русской мафии* [Alone against Russian mafia] (Fig. 2). It should be mentioned that the localised slogan is not available on popular Russian streaming services (Okko, Kinopoisk) and could be found only by searching images.



Fig. 1. Original movie poster for Siberia in English

Source: Kinopoisk.

Retrieved 21 July 2025,
from <https://www.kinopoisk.ru/film/1038843/posters/page/1/>



Fig. 2. Localised movie poster for Siberia in Russian

Source: Kinopoisk.

Retrieved 21 July 2025,
from <https://www.kinopoisk.ru/film/1038843/posters/page/1/>

Adding sexual subtext

From a marketing point of view, in order to attract the attention of a certain target audience, erotic or sexual innuendo should be imparted into the title of the film. This localisation strategy could be seen in the case of the 2011 film *Friends with Benefits*. The word combination in the original film title is a set expression that could be translated into Russian as *Друзья с привилегиями* [Friends with privileges] or *Друзья с бонусами* [Friends with bonuses], referring to friends who have an intimate relationship but are not boyfriend and girlfriend or husband and wife. Nevertheless, the film was released in Russia under the title *Секс по дружбе* [Sex for friendship's sake]. The original title implies the same, albeit in a less direct fashion, so the film title was translated adequately. At the same time another possible variant of translation — *Друзья с привилегиями* [Friends with privileges] — would not give any hint to the Russian-speaking viewers as to the implicit meaning of this phrase. Thus, the strategy that was used in the process of localisation is to make the sexual subtext of the original title explicit in translation.

Another case in point is the film *No Strings Attached* (2010) that could be translated as *Без всяких обязательств* [without any commitment]. But the Russian audience was offered a more explicit film title — *Больше чем секс* [More than sex].

Although this localisation strategy is used quite often, it requires some caution. In November 2020 *The Boss Baby: Family Business* was promoted in Russia as *Босс-молокосос 2* [Boss-milk sucker 2]. The poster in English read *Playtime is over*, which could be rendered into Russian as *Игры закончились* [Games came to an end]. Instead, the slogan was localised as *Новая соска — просто бомба* [New baby pacifier is a bomb/sensation/smash]. The Russian viewers have spotted in this phrase not just a sexual or erotic subtext, but propaganda of paedophilia, and criticised the film distributor¹³.

Lowering speech register

This strategy is characterised by the use of colloquial vocabulary that could be considered offhand or even rude, express a negative attitude and include evaluative characteristics. For instance, the 2013 film *Identity Thief* that may be translated as *Кража личности* [Stealing identity] (Fig. 3), has nothing in its title that may suggest emotional, evaluative or rude connotations. Meanwhile the Russian localised film title read *Поймай толстуху, если сможешь* [Catch the fatso if you can] (Fig. 4), which included an offensive colloquial word *толстуха*, referring to a fat woman.

Adding slang

This localisation strategy involves adding slang vocabulary to the titles of films and TV series. It can be illustrated by the Russian localised film titles that ended up with the word *перцы*, which literally means “pepper”, but is also used figuratively by younger generation to denote someone cool¹⁴: *Superbad* (2007) → *SuperПерцы* [Super dudes]; *Your Highness* (2010) → *Храбрые перцем* [Brave dudes]; *Last Vegas* (2013) → *Старперцы* [Star dudes].

In the case of *Superbad*, the inclusion of the slang word *перец* (the plural form — *перцы*) in the Russian film title directly indicates that the main characters of the comedy are young people: *SuperПерцы* [Super dudes]. The word *перец* in the localised title functions as a substitute for *bad*, which could refer to sexually provocative behaviour¹⁵ portrayed in the film.

¹³ Mineeva, A. (2020, November 27). “Propaganda pedofili”: «Bossa-molokososa 2» razgromili iz-za slogana [“Paedophilia propaganda”: *Boss-milk sucker 2* was trashed because of the tagline]. *Gazeta.ru*. Retrieved 28 July 2025, from https://www.gazeta.ru/culture/2020/11/27/a_13377493.shtml

¹⁴ Clovar’ sovremennoi leksiki, zhargona i slenga. Stat’i na bukvu “P” [Dictionary of modern vocabulary, jargon and slang. The letter “П” [P]]. Retrieved 31 July 2025, from <http://rus-yaz.niv.ru/doc/jargon-dictionary/fc/slovar-207-1.htm#zag-3792>

¹⁵ Bad. Retrieved 31 July 2025, from <https://www.urbandictionary.com/define.php?term=bad>



Fig. 3. Original movie poster for Identity Thief in English

Source: Kinopoisk.

Retrieved 21 July 2025,

from <https://www.kinopoisk.ru/film/606698/posters/page/1/>



Fig. 4. Localised movie poster for Identity Thief in Russian

Source: Kinopoisk.

Retrieved 21 July 2025,

from <https://www.kinopoisk.ru/film/606698/posters/page/1/>

As for the film *Your Highness*, its Russian counterpart, *Храбрые перцем*, also includes the word *перец*, however, its use seems to be intentionally ambiguous. It may mean “a soldier who is serving while violating the regulations”¹⁶, which fits the plot line as one of the main characters is a knight who tries to avoid problems by any means necessary. It may invoke associations with the phrase *храбрые сердцем* [brave-hearted] as it sounds very similar in Russian. Other interpretations are also possible.

And in the final example illustrating the strategy of adding slang to the translated film title, *Last Vegas*, a comical effect is achieved through a play on words as the Russian title *Старперцы* sounds very similar to the slang word *старпёр*, designating an old man with a derogatory connotation: *старик, пенсионер, старикашка*¹⁷. Thus, the original title, *Last Vegas*, is referring to the city in the USA Las Vegas and at the same time hinting at the age of the protagonists, while the localised title of the film, *Старперцы*, mirrors the main storyline — an attempt by several retired people

¹⁶ Clovar' sovremennoi leksiki, zhargona i slenga. Stat'i na bukvu "P" [Dictionary of modern vocabulary, jargon and slang. The letter "П" [P]]. Retrieved 31 July 2025, from <http://rus-yaz.niv.ru/doc/jargon-dictionary/fc/slovar-207-1.htm#zag-3792>

¹⁷ Slovar' russkikh sinonimov (svodnyi iz interneta, 2010). Slova na bukvu "S" [Dictionary of Russian Synonyms (compiled from the Internet, 2010). Words with the letter "С" [C]]. Retrieved 31 July 2025, from <http://rus-yaz.niv.ru/doc/synonym/fc/slovar-209-12.htm#zag-70631>

to have fun in the city, which is famous for its vibrant nightlife. Using the English word *star* in the Russian film title may also serve the function of attracting attention to the star-studded cast.

Film titles that are localised using such strategies as ***Adding slang*** and ***Lowering speech register*** tend to appeal to a younger audience and therefore act as a filter to discourage those viewers for whom the film is not intended.

Developing the story

The film distributor may use the localisation strategy of developing the story by hinting at the scale of the film. As a result, words like *history*, *chronicles* and *saga* are added to the title. For example, this strategy was implemented while translating the title of a vampire film: *Let Me In* (2010) was localised for the Russian market as *Впусти меня. Сага* [Let me in. Saga]. The localised title serves as a reference to another famous vampire saga — *Twilight*.

Another case in point — the 2024 film *Furiosa: A Mad Max Saga*, released in Russia under the title *Фуриоса: хроники безумного Макса* [Furiosa: chronicles of mad Max], which was a sequel to the 2015 film *Mad Max: Fury Road*, translated literally into Russian as *Безумный Макс: дорога ярости* [Mad Max: fury road].

The same strategy can be identified in the localisation of the following films and TV series: *Masters of Science Fiction* (2007) → *Хроники будущего* [Chronicles of the future]; *Dylan Dog: Dead of Night* (2010) → *Хроники вампиров* [Vampire chronicles]; *Mortal Engines* (2018) → *Хроники хищных городов* [Chronicles of predator cities]; *The Secret Kingdom* (2022) → *Хроники Панголинов* [Chronicles of the Pangolin]; *The Creep Tapes* (2024) → *Ублюдские хроники* [Chronicles of a creep].

Unlike the ***Franchise expansion strategy***, the ***Strategy of developing the story*** presupposes incorporating into the film titles only those words that increase the artistic value of an audiovisual work, hinting that it may have a sequel, prequel or is a TV series.

Using memes

This method of localisation is arguably less common as it targets even more precisely the audience of the source language and culture. It could be observed in the translation of the 2011 film *Salmon Fishing in Yemen* (Fig. 5), which in Russian literally means *Ловля лосося в Йемене*, however, it was released under the title *Рыба моей мечты* [The fish of my dreams] (Fig. 6).

Another example is the 2012 comedy *Lola Versus*, i.e. *Лола против* in Russian, which was localised as a colloquial expression *Давай, до свидания!* [Good riddance!].

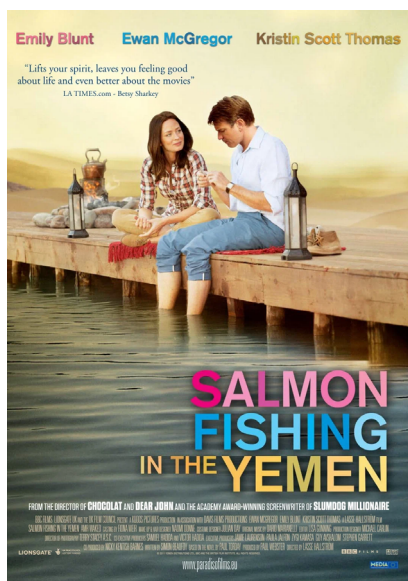


Fig. 5. Original movie poster for Salmon Fishing in the Yemen in English

Source: Kinopoisk. Retrieved 21 July 2025, from <https://www.kinopoisk.ru/film/484559/posters/page/1/>



Fig. 6. Localised movie poster for Salmon Fishing in the Yemen in Russian

Source: Kinopoisk. Retrieved 21 July 2025, from <https://www.kinopoisk.ru/film/484559/posters/page/1/>

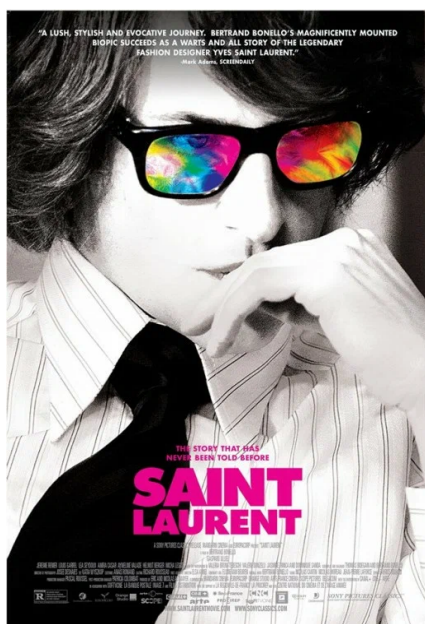


Fig. 7. Original movie poster for Saint Laurent in English

Source: Kinopoisk. Retrieved 21 July 2025, from <https://www.kinopoisk.ru/film/735178/posters/page/1/>



Fig. 8. Localised movie poster for Saint Laurent in Russian

Source: Kinopoisk. Retrieved 21 July 2025, from <https://www.kinopoisk.ru/film/735178/posters/page/1/>

Expanding the title

This strategy is employed when it makes sense to use the original title, but it is not always sufficient for the target audience to grasp the message of the film. In this case the title is expanded to include an explanation. For instance, the 2014 film *Saint Laurent* (Fig. 7), with the name of a famous designer in the original title, in translation gained an extension in the form of a sentence: *Сен-Лоран. Стиль — это я* [Saint Laurent. I am style] (Fig. 8). It seems that the Russian film distributors deemed it necessary to clarify for the target audience that the film is about fashion.

One more example of this localisation strategy is the film *Elysium* (2013), which in Russian had an explanatory note: *Элизиум. Рай на земле* [Elysium. Heaven on Earth].

Discussion

Evidently, while localising a film, it is not always necessary to resort to certain strategies for translating the title. In some cases, the title is self-evident and does not require changes in translation. For example, when a film is based on a book with the same title: *Harry Potter and the Philosopher's Stone* (2001) → *Гарри Поттер и философский камень*; *Memoirs of a Geisha* (2005) → *Мемуары Гейши*; *The Great Gatsby* (2013) → *Великий Гэтсби*.

Another instance is when the title could easily be translated into the target language word by word. To name a few, *Titanic* (1997) → *Титаник*; *Avatar* (2009) → *Аватар*; *Once Upon a Time in Hollywood* (2019) → *Однажды в Голливуде*. These are examples of literal translations.

Another case where translation as such is not performed is when the film title is transliterated. This may be due to the use of proper names in the film title or the inability to find a euphonious equivalent in the target language. This case can be illustrated by the example of the series *Virgin River* (2019 — present), which is broadcast in Russia under the title *Виржин Ривер* [Virgin River].

Sometimes the film's title and its meaning originate from some well-known phrase, which should be preserved in translation. One such example is the 2014 film *Ex Machina* which was rendered into Russian as *Из машины* [From machine]. The film tells a story about a scientist who makes robots and, by implanting them with artificial intelligence, tries to ensure that his creatures are as close to humans as possible in their behaviour, thinking, sensations and feelings, level of social adaptation and empathy. The title *Ex Machina* refers to the ancient Greek *Deus ex machina*¹⁸, that was used in Greek and Roman theatre setting. To play a deity that is capable of flying, the actor was suspended by a rope and brought on stage with a device similar to a modern crane (from Greek *mēchanē*). That gave rise to the expression *Deus ex machina*, which in Russian was traditionally rendered as *Бог из машины*¹⁹ [God

¹⁸ *Deus ex machina*. Britannica. Retrieved 31 July 2025, from <https://www.britannica.com/art/deus-ex-machina>

¹⁹ *Bog iz mashiny* [God from machine]. Slovar' antichnosti [Dictionary of Antiquity]. Retrieved 31 July 2025, from <https://ancientrome.ru/dictio/article.htm?a=168889981>

from machine]. Thus, the Russian localised title *Из машины* makes a reference to the already existing phrase in the target language.

This article does not aspire to provide an exhaustive list of film title localisation strategies, however, it offers an overview of general tendencies concerning film title translation taking into consideration social and cultural aspects. Moreover, it is not feasible to give a complete list of marketing strategies for the localisation of film titles, since translation strategy is always connected to a multitude of external factors that are constantly changing and reflecting the time period of the film release in a foreign market. Although some of the film titles seem quite easy to translate, they are always localised by a human marketing expert, who has seen the film itself, or at least its trailer. Machine translation systems are unable to take into consideration all the linguistic connotations, the variety of characteristics of the target audience as well as localisation trends. Even a human may err, as we have seen in case of the localised cartoon title *Босс-молокосос 2* [Boss-milk sucker 2].

Conclusion

This study has shown that, despite the rapid progress of machine translation technologies, it is impossible to use them to translate such marketing texts as film titles, at least at the current stage of development of neural machine translation systems.

A neural network can be trained to script the entire film and analyse its content, provide a summary, and suggest variants of the film title translation on the basis of diegetic speech scripts, summary and the key ideas of the film. Nevertheless, the neural network as of now is unable to analyse the video sequence and match it with diegetic speech, as well as the emotions conveyed by the film's cast and musical background. Therefore, a film title suggested by an AI on the basis of diegetic speech scripts cannot perform all the functions that a film title has to perform to ensure a box office hit. Even a marketing expert, when translating a film title, relies on the film or its trailer if the film itself is not available due to some reasons.

In addition, one has to consider the time period in which a film is translated and released: certain social, cultural, economic, political factors that affect translation may become relevant. AI is currently unable to take into account the dynamic nature of the context because it is trained on already existing databases. AI may be unaware of the events and their implications if they are very recent. It is likely to miss connotations and associations that are in the process of being established in the target language and culture. So far, the AI is lagging behind the present timeline, so it can not translate a film title in view of all the marketing and localisation requirements that humans bear in mind. The conclusions of this study can be regarded as a prospect for the development of AI technologies in the translation industry.

References

- Anissimov, V.E., Borissova, A.S., & Konson, G.R. (2019). Linguocultural localization of movie titles. *Russian Journal of Linguistics*, 23(2), 435–459. (In Russ.). <https://doi.org/10.22363/2312-9182-2019-23-2-435-459> EDN: QJHIQQ
- Bae, G.W., & Kim, H.-J. (2019). The impact of movie titles on box office success. *Journal of Business Research*, 103, 100–109. <https://doi.org/10.1016/j.jbusres.2019.06.023>
- Bourdieu, P. (1982). *Ce que parler veut dire: L'économie des échanges linguistiques*. Fayard.
- Bourdieu, P. (1986). The forms of capita. In J.G. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood Press.
- Constantine, P. (2019). Google translate gets voltaire: Literary translation and the age of artificial intelligence. *Contemporary French and Francophone Studies*, 23(4), 471–479. <https://doi.org/10.1080/17409292.2019.1694798>
- Crane, D. (2014). Cultural globalization and the dominance of the American film industry: Cultural policies, national film industries, and transnational film. *International Journal of Cultural Policy*, 20(4), 365–382. <https://doi.org/10.1080/10286632.2013.832233>
- DeCamp, J., & Zetzsche, J. (2014). A history of translation technology in the United States. In Sin-wai, Ch. (Ed.), *Routledge Encyclopaedia of Translation Technology* (pp. 375–392). Routledge. <https://doi.org/10.4324/9781315749129>
- Featherstone, M. (Ed.). (1992). *Cultural Theory and Cultural Change*. Sage.
- Haidegger, I. (2015). What's in a name? The art of movie titling. *Word & Image*, 31(4), 425–441. <https://doi.org/10.1080/02666286.2015.1053037>
- Jiang, F., & Yin, H. (2024). Exploring cultural identity and localized aesthetics in Chinese science fiction cinema: A study of community building. *Educational Administration: Theory and Practice*, 30(3), 702–706. <https://doi.org/10.53555/kuey.v30i3.1336>
- Jones, B., Andreas, J., Bauer, D., Hermann, K.M. Knight, K. (2012). Semantics-based machine translation with hyperedge replacement grammars. In Kay, M. & Boitet, Ch. (Eds.), *Proceedings of COLING 2012: Technical Papers, Mumbai, The COLING 2012 Organizing Committee* (pp. 1359–1376). Mumbai. <https://aclanthology.org/volumes/C12-1/>
- Kettongma, N. (2024). Movie titles' translation strategies from English into Thai in Monomax Application. *Community and Social Development Journal*, 25(3), 195–213. <https://doi.org/10.57260/csdj.2024.269277>
- Kiran, A. (2023). Between global and local: Translation and localization in Netflix Turkey's media paratexts. *Translation Studies*, 16(3), 361–378. <https://doi.org/10.1080/14781700.2023.2201282> EDN: RVANXJ
- Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press. <https://doi.org/10.1017/9781108608480>
- Krasina, E.A., & Moctar, A. (2020). On film titles: Translation or retitling? *Bulletin of Moscow Region State University*, (2), 283–297. <https://doi.org/10.18384/2224-0209-2020-2-1014> EDN: NEINVL
- Lang, M. (2006). Globalization and its history. *The Journal of Modern History*, 78(4), 899–931. <https://www.jstor.org/stable/10.1086/511251>
- Martín-Mor, A., & Sánchez-Gijón, P. (2016). Machine translation and audiovisual products: a case study. *The Journal of Specialised Translation*, 26, 172–186. https://jostrans.soap2.ch/issue26/art_martin.php
- Matusov, E. (2019). The challenges of using neural machine translation for literature. In: Hadley, J., Popović, M., Afli, H., & Way, A. (Eds.), *Proceedings of the Qualities of Literary Machine Translation, Dublin, European Association for Machine Translation* (pp. 10–19). Dublin. <https://aclanthology.org/W19-7302/>
- Maučec, S.M., & Donaj, G. (2020). Machine translation and the evaluation of its quality. *Recent trends in Computational Intelligence*. <https://doi.org/10.5772/intechopen.89063>
- Pooch, M.O. (2016). Globalization and its effects. In M.U. Pooch (Ed.), *DiverCity — Global Cities as a Literary Phenomenon: Toronto, New York, and Los Angeles in a Globalizing Age* (pp.15–26). Lettre. <https://www.jstor.org/stable/j.ctv1wxt87>
- Qun, L., & Xiaojun, Z. (2014). Machine translation: general. In Sin-wai, Ch. (Ed.), *Routledge Encyclopaedia of Translation Technology* (pp. 105–119). Routledge. <https://doi.org/10.4324/9781315749129>
- Shiwen, Y., & Xiaojing, B. (2014). Rule-based machine translation. In Sin-wai, Ch. (Ed.), *Routledge Encyclopaedia of Translation Technology* (pp. 186–200). Routledge. <https://doi.org/10.4324/9781315749129>
- Sin-wai, Ch. (2014). The development of translation technology: 1967–2013. In Sin-wai Ch. (Ed.), *Routledge Encyclopaedia of Translation Technology* (pp. 3–31). Routledge. <https://doi.org/10.4324/9781315749129>
- Sulubacak, U., Caglayan, O., Grönroos, S.-A., Rouhe, A., Elliott, D., Specia, L. & Tiedemann, J. (2020). Multimodal machine translation through visuals and speech. *Machine Translation*, 34(2–3), 97–147. <https://doi.org/10.1007/s10590-020-09250-0> EDN: GZTTGD

- Tak-ming, B. W., Webster, J. J. (2014). Example-based machine translation. In Sin-wai, Ch. (Ed.), *Routledge Encyclopaedia of Translation Technology* (pp. 137–151). Routledge. <https://doi.org/10.4324/9781315749129>
- Toral, A., Wieling, M., & Way, A. (2018). Post-editing effort of a novel with statistical and neural machine translation. *Frontiers in Digital Humanities*, 5, 1–11. <https://doi.org/10.3389/fdigh.2018.00009>
- Yang, L., & Min, Z. (2014). Statistical machine translation. In Sin-wai, Ch. (Ed.), *Routledge Encyclopaedia of Translation Technology* (pp. 201–212). Routledge. <https://doi.org/10.4324/9781315749129>
- Zhang, H. (2023). Strategic localization and screen adaptations of journey to the West from the 1920s to the 2010s. In Schultz, C. & Mello, C. (Eds.), *Chinese Film in the Twenty-First Century. Movements, Genres, Intermedia* (pp. 158–173). Routledge. <https://doi.org/10.4324/9781003371694>

Bionotes:

Yulia M. ALYUNINA, candidate of Philological Sciences and Ph.D. in Linguistics, senior lecturer and research fellow, Department of Foreign Languages, Faculty of Philology, RUDN University, 6 Miklukho-Maklaya st., Moscow, 117198, Russian Federation; research assistant, Laboratory of Museum Technologies, Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) of the Russian Academy of Sciences, 7–9 Universitetskaya Embankment, Saint Petersburg, 199034, Russian Federation. *Research interests*: corpus linguistics, audiovisual translation, digital technologies in translation, localization, language and culture, intercultural cooperation.

E-mail: alyunina.y@mail.ru

ORCID: 0000-0002-4970-5584

SPIN-code: 1972-7723

Elvira R. GANEVA, candidate of Philological Sciences, associate professor, Department of Foreign Languages and Intercultural Communication, Faculty of International Economic Relations, Financial University under the Government of the Russian Federation, 49 Leningradsky Avenue, bldg. 2, Moscow, 125167, Russian Federation. *Research interests*: linguistic aspects of translation and interpretation, simultaneous interpretation through the prism of cognitive linguistics and the influence of modern technologies on the industry of translation in general.

E-mail: erganeeva@fa.ru

ORCID: 0000-0001-6473-6576

SPIN-code: 9831-6986

Информация об авторах:

АЛЮНИНА Юлия Матвеевна, кандидат филологических наук и Ph.D. in Linguistics, старший преподаватель и научный сотрудник кафедры иностранных языков филологического факультета, Российский университет дружбы народов, Российская Федерация, 117198, Москва, ул. Миклухо-Маклая, д. 6; стажер-исследователь лаборатории музейных технологий, Музей антропологии и этнографии имени Петра Великого (Кунсткамера) Российской академии наук, Российская Федерация, 199034, Санкт-Петербург, Университетская набережная, д. 7–9. *Научные интересы*: корпусная лингвистика, аудиовизуальный перевод, цифровые технологии в переводе, локализация, язык и культура, межкультурное взаимодействие.

E-mail: alyunina.y@mail.ru

ORCID: 0000-0002-4970-5584

SPIN-код: 1972-7723

ГАНЕЕВА Эльвира Рустемовна, кандидат филологических наук, доцент кафедры иностранных языков и межкультурной коммуникации факультета международных экономических отношений, Финансовый университет при Правительстве Российской Федерации, Российская Федерация, 125167, Москва, ул. пр-т Ленинградский, д. 49/2. *Научные интересы*: лингвистический аспект письменного и устного перевода, синхронный перевод в аспекте когнитивной лингвистики и влияния современных технологий на индустрию перевода в целом.

E-mail: erganeeva@fa.ru

ORCID: 0000-0001-6473-6576

SPIN-код: 9831-6986

DOI: 10.22363/2949-5997-2025-3-2-166-191

EDN HMVEBY

Научная статья / Research article

Социолингвистические параметры образа СССР в романе У. Тевиса «The Queen's gambit» и его экранизации: легитимация доминирования

Анна Сергеевна КУЗНЕЦОВА 

Российский университет дружбы народов, Москва, Российская Федерация

✉ kuznetsova-ansr@rudn.ru

Аннотация. Рассмотрены языковые средства для создания образа СССР в межсемиотической плоскости на материале речи героев романа У. Тевиса «The Queen's Gambit» (1984), его русскоязычного перевода «Ход королевы» (2020), экранизации романа и его аудиовизуальной адаптации на русский язык. Исследовательский корпус составили реплики главных персонажей литературного и кинематографического произведений. Лексический материал реплик был распределен по группам: имена и фамилии героев, топонимы, шахматная терминология и лексика, описывающая реалии Советского Союза. Отдельное внимание обращено на роль русского языка в англоязычном оригинале романа, стратегии передачи функционала русской речи в переводе рассмотренных произведений на русский язык. Методология исследования представлена количественным и качественным подходами. Количественный подход включает инструменты дальнего чтения (работа с корпусным менеджером *Sketch Engine* и его инструментами *Word List* и *Keyword*), качественный — кластеризацию и лексико-семантический анализ фокусных лексических единиц. Выявлены ключевые стратегии репрезентации образа «другого». В оригинале романа лингвистическая маркировка советского пространства достигается за счет транслитерации русских лексем, создающей для англоязычной аудитории эффект культурной дистанции. В русскоязычном переводе данный эффект нивелируется, а языковой барьер передается имплицитно через авторские ремарки. Экранизация формирует образ СССР преимущественно через комплекс аудиовизуальных кодов (кадры архитектуры, предметов быта, невербальная коммуникация и др.), создавая детализированный визуальный контекст. При этом в локализованном сериале утрачивается исходная языковая полифония,

© Кузнецова А.С., 2025



This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

приводящая к снижению аутентичности и смысловым упрощениям. Исследование показало, что анализируемый образ предстает как динамический конструкт, формирующийся в результате активизации вербальных, переводческих и визуальных стратегий.

Ключевые слова: имагология, репрезентация, корпусная лингвистика, социолингвистика, киноязык

Заявление о конфликте интересов. Автор заявляет об отсутствии конфликта интересов.

История статьи: получена 30 октября 2025; принята к публикации 10 ноября 2025.

Для цитирования: Кузнецова А.С. Социолингвистические параметры образа СССР в романе У. Тевиса «The Queen’s Gambit» и его экранизации: легитимация доминирования // *Macrosociolinguistics and Minority Languages*. 2025. Т. 3. № 2. С. 166–191. <https://doi.org/110.22363/2949-5997-2025-3-2-166-191> EDN: HMVEBY

Sociolinguistic parameters of the image of the USSR in W. Tevis’s Novel “The Queen’s gambit” and its film adaptations: legitimization of dominance

Anna S. KUZNETSOVA 

RUDN University, *Moscow, Russian Federation*

✉ kuznetsova-ansr@rudn.ru

Abstract. The article describes the linguistic means used to create the image of the USSR on an intersemiotic plane, based on the analysis of characters’ speech in W. Tevis’s novel “The Queen’s Gambit” (1984), its Russian translation «Ход королевы» (2020), the novel’s screen adaptation, and its Russian audio-visual adaptation. The research corpus comprises the lines of key characters from the literary work and the film. The lexical material from these lines was categorized into groups: characters’ first and last names, toponyms, chess terminology, and vocabulary describing Soviet realities. Special attention is paid to the role of the Russian language in the English-language original of the novel and to the strategies for conveying the function of Russian speech in the translation of the analyzed works into Russian. The research methodology combines quantitative and qualitative approaches. The quantitative approach utilizes distant reading tools (working with the Sketch Engine corpus manager and its Word List and Keyword tools), while the qualitative approach involves clustering and lexical-semantic analysis of focal lexical units. The analysis revealed key strategies for representing the image of the ‘Other’. In the original novel, the linguistic marking of the Soviet space is achieved through the transliteration of Russian lexemes, creating an effect of cultural distance for the English-speaking audience. In the Russian translation, this effect is neutralized, and the language barrier is conveyed implicitly through authorial remarks. The screen adaptation constructs the image of the USSR primarily through a complex of audio-visual codes (shots of architecture, everyday objects, non-verbal communication, etc.), creating a detailed visual context. However, in the localized series, the original linguistic polyphony is lost, leading to reduced authenticity and semantic simplifications. The

study showed that the analyzed image emerges as a dynamic construct formed through the activation of verbal, translational, and visual strategies.

Key words: imagology, representation, corpus linguistics, sociolinguistics, film language

Conflict of interest. The author declares no conflict of interest.

Article history: received 30 October 2025; accepted 10 November 2025.

For citation: Kuznetsova, A.S. (2025). Sociolinguistic parameters of the image of the USSR in W. Tevis's Novel "The Queen's gambit" and its film adaptations: Legitimization of dominance. *Macrosociolinguistics and Minority Languages*, 3(2), 166–191. (In Russ.). <https://doi.org/10.22363/2949-5997-2025-3-2-166-191> EDN: HMVEBY

Введение

Изучение механизмов культурного доминирования в эпоху глобализации требует обращения к концепциям, объясняющим несиловые формы влияния. С одной стороны, это теория символического капитала Пьера Бурдьё, согласно которой язык функционирует как капитал на «рынке языков» (Bourdieu, 1982), а владение доминирующим языком (сегодня — английским) открывает доступ к социальным и культурным ресурсам (Bourdieu, 1986). С другой стороны, это концепция «мягкой силы» Дж. Ная, описывающая способность государства добиваться влияния благодаря привлекательности своей культуры и ценностей (Nye, 1990; Nay, 1991). Предлагается рассмотреть синтез этих подходов для анализа унификации глобального семиотического пространства. Английский язык и американская медиаиндустрия, выступая носителями «мягкой силы», легитимируют англосаксонские смыслы не только на вербальном, но и на визуальном уровне. Репрезентация образа страны в литературе и кинематографе становится значимым элементом культурной коммуникации, формируя восприятие определенного исторического периода и его социальных реалий. Один из маркеров актуальности того или иного образа — реакция общества. Так со второго десятилетия XXI в. (рис. 1 и 2) в мире наблюдается рост интереса к «русскому», прослеживаемый в укреплении и популяризации таких концептов, как *Slavic bimbo* или *Russian bimbo* (*русская красавица*)¹, *Slavic stare* (*русский взгляд*)²; *Slavic core* и *Slavic style* (*славянский стиль*, преподносимый как

¹ Что такое стиль Slavic Bimbo и почему весь мир играет в русскую красавицу // Dzen. 2024. URL: <https://dzen.ru/a/ZbEATc7TaHro8w3R> (дата обращения: 20.12.2025); Моисеева М. Почему все подражают русским девушкам из 90-х: разбираем стиль slavic bimbo // Spletnik.ru. 2023. URL: <https://spletnik.ru/pochemu-vse-podrazhayut-russkim-devushkam-iz-90-h-razbiraem-stil-slavic-bimbo-316849> (дата обращения: 20.12.2025); Skiba K.M. What's Lurking Behind the 'Slavic Bimbo' Aesthetic? // Polyesterzine.com. URL: <https://www.polyesterzine.com/features/whats-lurking-behind-the-slavic-bimbo-aesthetic> (дата обращения: 20.12.2025).

² Тега Н. Как тренд на славянскую эстетику захватил тикток — меха, песни, мамыны советы и картошка с селедочкой // Мел. URL: <https://mel.fm/zhizn/razbor/2370519-na-30-travma-na-40-osuzhdeniye-i-na-30-prinyatiye-togo-chto-ty-prozhil-eshche-odin-den-kak-trend-na-> (дата обращения: 20.12.2025).

«традиционная русская эстетика», которая «захватила соцсети»³) (Karamalak, Verlib, 2025), что демонстрируют результаты запроса по лексеме *Slavic* в *Google Trends*. В голливудском кинематографе вне зависимости от внешнеполитических обстоятельств устойчиво появляются фильмы и сериалы с сюжетом, обращенным к русской и советской эстетике (Алюнина, 2024; Алюнина, Роговец, 2025; Алюнина, Роговец, 2024; Рябов, 2024; Riabov, 2025). В таких кинокартинах одним из инструментов формирования образа русского человека является речь, а для исполнения своих ролей англоязычные актеры говорят на русском языке. Примерами тому служат новые сериалы «Heated Rivalry»⁴ («Жаркое соперничество» 2025 г.) про русского хоккеиста и «Ponies»⁵ (анонсированный на 2026 г. телесериал «Пони») про американских женщин, чьи мужья погибают на территории СССР при загадочных обстоятельствах.

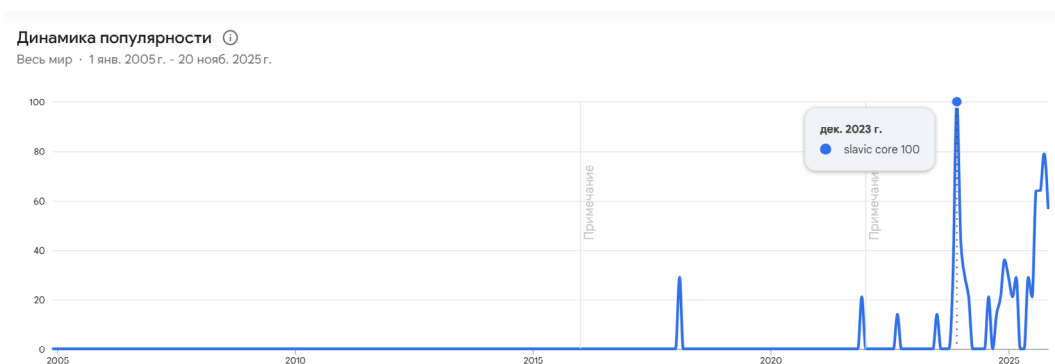


Рис. 1. Динамика популярности концепта Slavic core по интернет-запросам

Источник: Google Trends. URL: <https://trends.google.com/explore?q=slavic%20core&date=2005-01-01%202025-11-20&geo=Worldwide> (дата обращения: 20.11.2025).

Fig. 1. Dynamics of the popularity of the Slavic core concept based on internet requests

Source: Google Trends. Retrieved 20 November 2025, from

<https://trends.google.com/explore?q=slavic%20core&date=2005-01-01%202025-11-20&geo=Worldwide>

³ Slavic core или культурный код. Погружаемся в новый тренд // Looky.media. URL: <https://www.looky.media/blog/slavic-core-ili-kulturnyy-kod-pogruzhaemysya-v-novyy-trend> (дата обращения: 20.12.2025).

⁴ Sengupta S. Is Heated Rivalry Episode 2 only in the Russian language? Here's what we know // Soapcentral. 2025. URL: <https://www.soapcentral.com/shows/is-heated-rivalry-episode-2-russian-language-here-s-know-english-subtitles> (дата обращения: 20.12.2025); Swift A. Heated Rivalry's Connor Storrie Re-veals The 'Most Scary Thing Ever' About Playing Ilya Rozanov // Tvline. 2025. URL: <https://www.tvline.com/2064501/heated-rivalry-connor-storrie-russian-accent-explained/> (дата обращения: 20.12.2025); LeGardye Q. 'Heated Rivalry' Star Connor Storrie Studied Russian and Channeled His Own Love Language to Play Ilya // Yahoo. 2025. URL: <https://www.yahoo.com/entertainment/tv/articles/heated-rivalry-star-connor-storrie-211546856.html> (дата обращения: 20.12.2025).

⁵ Billington A. Emilia Clarke & Haley Lu Richardson in 'PONIES' 70s Spy Series Trailer // Firstshowing. 2025. URL: <https://www.firstshowing.net/2025/emilia-clarke-haley-lu-richardson-in-ponies-70s-spy-series-trailer/> (дата обращения: 20.12.2025); Онопдэра Ю. Реасок показал трейлер сериала «Пони» с Эмилией Кларк и Хейли Лу Ричардсон // Daily.afisha. 2025. URL: <https://daily.afisha.ru/news/104903-peасок-pokazal-treyer-seriala-poni-s-emiliey-klark-i-heyli-lu-richardson/> (дата обращения: 20.12.2025).

Пользователи, которые вводили запрос **slavic core**, также выполняли поиск по указанным ниже запросам

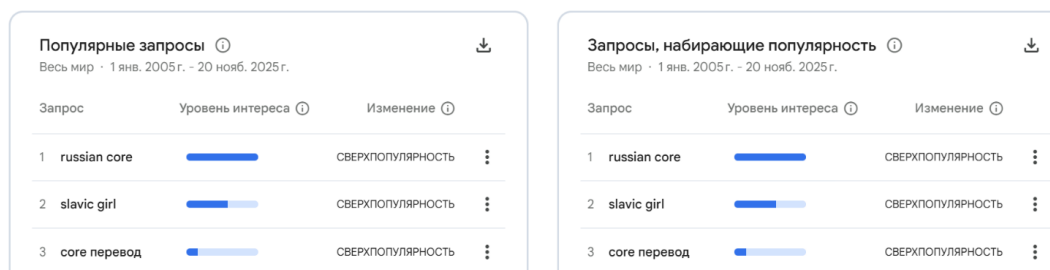


Рис. 2. Поисковые запросы интернет-пользователей, которые интересуются концептом Slavic core

Источник: Google Trends. URL: <https://trends.google.com/explore?q=slavic%20core&date=2005-01-01%202025-11-20&geo=Worldwide> (дата обращения: 20.11.2025).

Fig. 2. Search queries from internet users who are interested in the Slavic core concept

Source: Google Trends. Retrieved 20 December 2025, from <https://trends.google.com/explore?q=slavic%20core&date=2005-01-01%202025-11-20&geo=Worldwide>

В связи со сложившимся в 2000-е гг. трендом на «русскость» особую научную актуальность приобретает исследование образа Советского Союза, который продолжает оставаться объектом рефлексии и множественных интерпретаций в современном культурном дискурсе как в отечественном пространстве, так и в зарубежном (Алюнина, Роговец, 2024: 237; Белов, 2024: 51; Karamalak, Vertlib, 2025: 4; Riabov, 2025: 24; Hill, 1972: 669–671). Роман американского писателя Уолтера Тевиса «The Queen's Gambit»⁶ (1984) и его последующая экранизация⁷ представляют исследовательский интерес не только в силу своей художественной значимости и популярности, но и в силу социолингвистических и социокультурных инструментов — механик, использованных автором книги и режиссером экранизации для конструирования образа СССР.

В романе У. Тевиса «Ход королевы»⁸ показан путь становления личности и карьеры американской шахматистки Элизабет Хармон, главным противником которой является советский шахматист Василий Боргов. Уникальность героини заключается в парадоксальном сочетании беспрецедентной одаренности и крайней личностной уязвимости. Главной целью шахматных турниров для нее становится не столько спортивное превосходство, сколько попытка обретения контроля над собственным сознанием.

Сериал «Ход королевы» вызвал широкий общественный резонанс в России и позволяет говорить о нем как о ярком примере так называемого «эффекта Netflix»⁹. Как отмечает журнал Forbes, после выхода сериала продажи шахматных

⁶ Tevis W. The Queen's Gambit. New York: RosettaBooks, 2014. 600 p.

⁷ Netflix. URL: <https://www.netflix.com/title/80234304> (дата обращения: 20.12.2025).

⁸ Тевис У. Ход королевы. М.: АСТ, 2021. 382 с.

⁹ Бакиева М. Эффект Netflix: как сериалы делают модными шахматы и помогают продавать кроссовки // Forbes Life. URL: <https://www.forbes.ru/forbeslife/420381-effekt-netflix-kak-serialy-delayut-modnymi-shahmaty-i-pomogayut-prodavat-krossovki> (дата обращения: 20.12.2025).

наборов на таких площадках, как 1stDibs и eBay, выросли на 87...130 %, а спрос на учебники по шахматам — на 603 %⁹. Это свидетельствует не просто о временном интересе, а о глубоком культурном сдвиге: благодаря сериалу шахматы были романтизированы, они стали ассоциироваться с кинематографическими образами гениальности, стиля и интеллектуальной победы. Социологи объясняют этот феномен тем, что эпизоды из сериала связываются в сознании аудитории с определенным стилем жизни и ценностями⁹. Компания Netflix прошла путь от сервиса почтовой доставки DVD до одного из лидеров мировой индустрии развлечений, что демонстрирует, как технологическая инновация и бизнес-модель, основанная на персонализированной подписке и data-driven производстве контента, создали беспрецедентный по охвату канал распространения культурных продуктов¹⁰. Этот канал можно назвать одним из ярчайших примеров концепции «мягкой силы», предложенной Дж. Наем. Согласно интерпретации О.Г. Леоновой, современная «мягкая сила» все менее является монополией государства и все более — «полем деятельности транснациональных структур, которые предлагают притягательные нарративы и модели идентичности» (Леонова, 2018: 102).

В контексте глобализации данный феномен обретает амбивалентное значение. С одной стороны, универсальность шахмат как интеллектуального кода, понятного любой культуре, и единый эстетический язык сериала способствуют глобальной унификации культурного потребления. С другой — именно эта унификация создает скрытую угрозу для локальных идентичностей. Адаптируя западный нарратив о гениальности и успехе, российский зритель неизбежно вступает в диалог с ценностями, которые могут вытеснять или маргинализировать отечественные культурные коды. Особенно остро этот процесс ощущается в сфере социального капитала: язык и культура перестают быть исключительно инструментами национальной консолидации и превращаются в товар. Восхищаясь игрой Бет Хармон, аудитория потребляет не просто шахматы, а глобальный бренд «интеллектуальной исключительности». Таким образом, рост продаж шахматных наборов парадоксальным образом сигнализирует не столько о возрождении интереса к игре как элементу национального наследия, сколько о включении российского зрителя в глобальный процесс унификации культурных предпочтений, где идентичность формируется под воздействием транснациональных медиагигантов.

В рамках настоящего исследования анализируются специфика вербальной репрезентации образа СССР в оригинальном тексте романа на английском языке и его русскоязычном переводе, а также особенности аудиовизуальной репрезентации образа СССР: «видео (визуальные элементы образа <...>), речь (вербальные отсылки к образу <...>) и сюжет» (Алюнина, Роговец, 2025: 169). Ключевыми проблемными вопросами избраны межсемиотический перевод и культурная рецепция

¹⁰ Мазина М. Эффект Netflix: как компания устроила революцию в индустрии развлечений // РБК Инвестиции. 2020. URL: <https://www.rbc.ru/quote/news/article/5ec383c59a7947e3453b32c7> (дата обращения: 20.12.2025).

советского прошлого. Исследование осуществлено в парадигме межкультурной коммуникации, имагологии, теории локализации и социолингвистики.

Теоретическим фундаментом исследования послужили работы в области социолингвистики, семиотики и теории культурного обмена. На макроуровне анализ опирается на концепцию «глобального лингвистического рынка» П. Бурдьё, где язык и культура функционируют как символический капитал, а различные формы культурной продукции вступают в конкурентную борьбу за легитимность и признание (Bourdieu, 1982). В условиях глобализации эта конкуренция приобретает характер соперничества между национальными культурами и транснациональными нарративами, которые, как отмечает С.Л. Маккей в своих исследованиях лингвистического империализма и глобального распространения языков, способны переструктурировать локальные иерархии ценностей и идентичности (McKay, 2002). Именно в этом контексте платформы вроде Netflix становятся агентами «мягкой силы», продвигающими определенные модели поведения и мышления в качестве универсальных и желаемых.

Переходя от макросоциологического анализа к механизмам конструирования смыслов (микроуровень), необходимо обратиться к семиотике. Вслед за Р. Якобсоном используем понятие межсемиотического перевода (трансмутации), определяемого как интерпретация вербальных знаков через невербальные знаковые системы, например, экранизацию литературного произведения (Якобсон, 1985: 363). Понятие расширяет поле переводоведения, включая в него семиотические трансформации между разными видами искусства. У. Эко развил эту идею, используя термин «интерсемиотический перевод» (Эко, 2006: 11). Он подчеркнул, что это не механический перенос, а сложная интерпретация, при которой содержание неизбежно трансформируется при переходе из одной знаковой системы (романа) в другую (фильм, картину). Названные концепции позволяют деконструировать то, каким именно образом глобальный культурный продукт адаптирует, перекодирует и транслирует ценности, превращая их в элемент идентичности массового зрителя на разных локальных рынках.

С учетом выше сказанного, целью исследования стало выявление и описание комплекса языковых и экранных средств, участвующих в конструировании образа Советского Союза в романе У. Тевиса «The Queen's Gambit» (1984) и его русскоязычном переводе «Ход королевы» (2020), а также в его одноименной экранизации на английском языке¹¹ и ее русскоязычной локализации¹². Объектом исследования послужил образ Советского Союза, репрезентированный в упомянутом материале. Предметом исследования выступили ключевые

¹¹ The Queen's Gambit // IMDb. URL: https://www.imdb.com/title/tt10048342/?ref=fn_t_1 (дата обращения: 07.10.2025).

¹² «Ход Королевы» (мини-сериал 2020) // Кинопоиск. URL: https://www.kinopoisk.ru/series/1253633/?ysclid=mkz75a06v611276174&utm_referrer=ya.ru (дата обращения: 07.10.2025).

и частотные лингвистические элементы (в оригинале и переводе романа) и аудиовизуальные коды (в экранизации), формирующие названный образ.

Материалы и методы исследования

Как было сказано выше, эмпирическую базу исследования составили следующие материалы:

- роман Уолтера Тевиса «The Queen's Gambit» (1984) на английском языке;
- официальный перевод романа на русский язык, изданный под названием «Ход королевы» (2020) и выполненный О. Павловской;
- англоязычный сериал — экранизация романа 2020 г. производства студии Netflix;
- официальный перевод сериала, выполненный под дубляж на русский язык студией «Невафильм»;
- англоязычные и русскоязычные субтитры¹³ к сериалу на сайте *OpenSubtitles*.

Для проведения количественного анализа текстовые версии романа преобразовали в цифровые корпуса с помощью корпусного менеджера *Sketch Engine*¹⁴. Объем англоязычного корпуса составил 101 798 слов (20 815 токенов), русскоязычного — 91 356 слов (113 946 токенов). В качестве референтных корпусов для сравнительного анализа использовались встроенные в *Sketch Engine* коллекции текстов: корпус *enTenTen21*¹⁵ — для текста в оригинале и корпус *ruTenTen17*¹⁶ — для перевода.

Аудиовизуальный материал объемом 385 минут в оригинале и 385 минут в переводе был подвергнут контент-анализу с фиксацией вербальных и визуальных репрезентаций образа СССР.

Методологическую основу работы составил комплексный подход, включающий количественные и качественные методы. Количественный анализ реализован с помощью инструментария корпусного менеджера *Sketch Engine*. Использовались такие функции корпусного менеджера как *Keyword* и *Word List*. Первая функция позволяет автоматически выявлять статистически значимые ключевые слова в целевом корпусе относительно референтного (Алюнина, 2022: 25), вторая — формирует ранжированные списки лексических единиц по критерию частотности. Качественный анализ включал лексико-семантический анализ отобранных единиц для определения их концептуальной нагрузки, а также метод кластеризации (тематического группирования), который позволил

¹³ «The Queen's Gambit» // Open subtitles. URL: <https://www.opensubtitles.org/en/search/sublanguageid-all/idmovie-773292> (дата обращения: 07.10.2025).

¹⁴ Sketch Engine. URL: <https://www.sketchengine.eu> (дата обращения: 07.10.2025).

¹⁵ enTenTen: Corpus of the English Web. URL: <https://www.sketchengine.eu/ententen-english-corpus/> (дата обращения: 07.10.2025).

¹⁶ ruTenTen: Corpus of the Russian Web. URL: <https://www.sketchengine.eu/rutenten-russian-corpus/> (дата обращения: 07.10.2025).

систематизировать выявленную лексику по смысловым категориям: антропонимы (имена и фамилии персонажей), топонимы, шахматная терминология, лексика, репрезентирующая советский социокультурный и бытовой контекст.

Такое сочетание методов обеспечивает объективность выборки ключевых слов и последующую глубину их концептуальной интерпретации.

Репрезентация образа СССР в оригинале и переводе романа У. Тевиса «The Queen's Gambit»

Противостояние СССР и США в период холодной войны выступает в романе У. Тевиса «The Queen's Gambit» (1984) в качестве одного из центральных сюжетнообразующих и идеологических контекстов.

Результаты корпусного анализа демонстрируют значительную концентрацию антропонимов, соотносимых с образом Советского Союза, в тексте произведения. Для наглядного сравнения распределения и количественного соотношения антропонимов в англоязычном оригинале и русскоязычном переводе ниже приведены соответствующие графики (рис. 3, а и б).

Как видно из графиков (см. рис. 3, а и б), данная категория лексики формирует ядро высокочастотных имен собственных как в оригинальном тексте, так и в его переводе. Эти лексические единицы можно распределить по четырем кластерам: антропонимы (антропонимы советского — *Borgov*, *Girev* и несоветского происхождения — *Harmon*), топонимы (*Moscow*, *USSR*), шахматная терминология (*shakmatni*), лексика, репрезентирующая советский социокультурный и бытовой контекст (*blinchiki* и *Pravda*).

Как показывает анализ частотности слов (см. рис. 3, а и б), количество употреблений проанализированных имен собственных в оригинале и в переводе не совпадает. Например, фамилия *Petrosian* в оригинале имеет частотность 82,77 ipm (от англ. *instance per million* — частотность на миллион), в переводе Петросян — 87,76 ipm. Как следствие, мы можем видеть несоответствие на графиках.

Среди имен шахматистов лидирующую позицию по показателю частотности занимает фамилия Боргов (*Borgov* — 1216,74 ipm / *Боргов* — 1895,63 ipm). Данный персонаж, сильнейший советский шахматист, выполняет функцию главного антагониста главной героини Бэт Хармон. Он символизирует вершину шахматного мастерства, которую Бэт предстоит достичь.

Второй по значимости частотный кластер образуют фамилии других советских гроссмейстеров: *Luchenko* (347,64 ipm) / *Лученко* (614,33 ipm), *Girev* (115,88 ipm) / *Гирев* (96,54 ipm), *Petrosian* (82,77 ipm) / *Петросян* (87,76 ipm). Замыкают список менее частотные, но маркированные в тексте антропонимы: *Flento* (74,49 ipm) / *Фленто* (87,76 ipm) и *Shapkin* (74,49 ipm) / *Шапкин* (61,43 ipm).

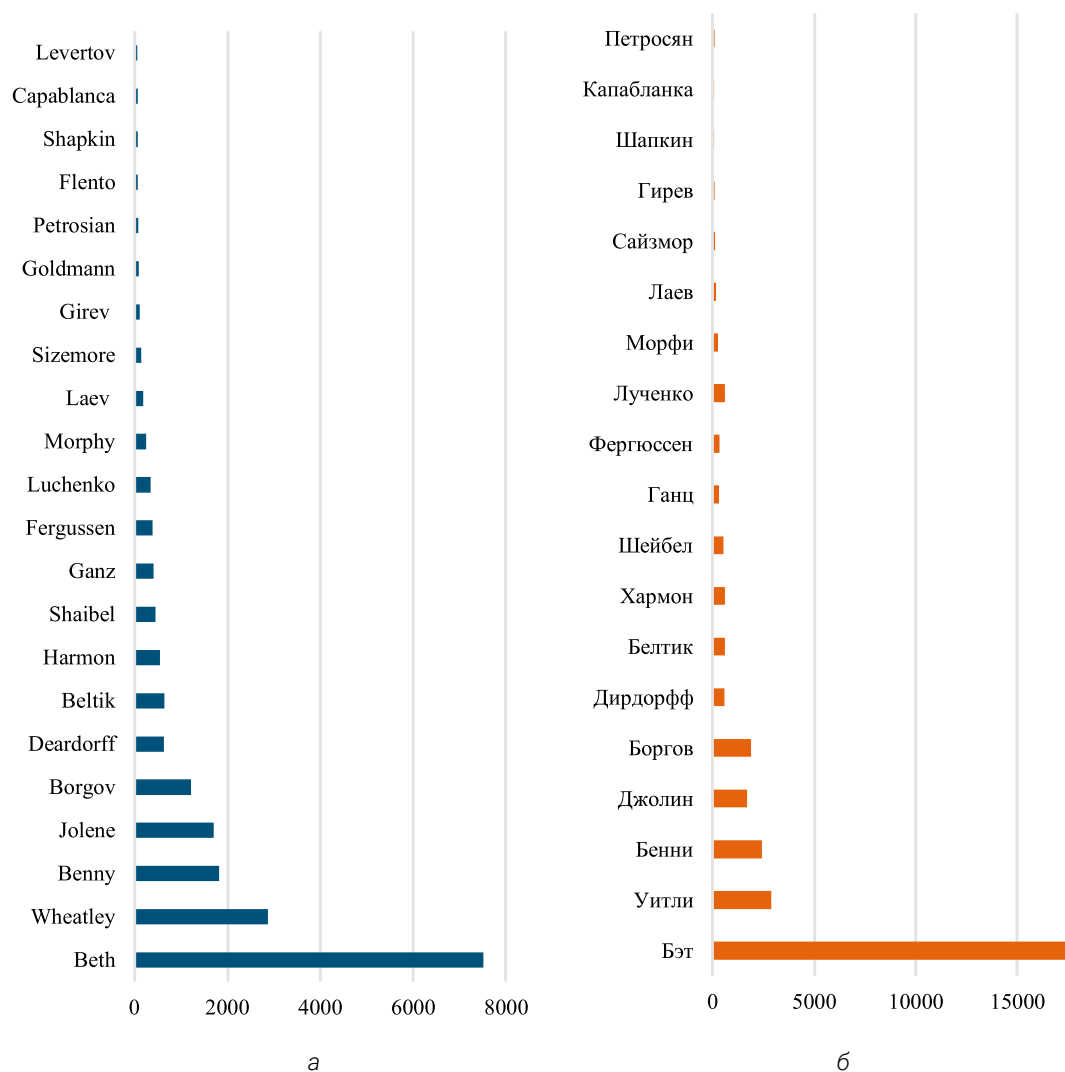


Рис. 3. Имена и фамилии героев: а – в оригинале романа, б – в переводе

Источник: составлено А.С. Кузнецовой.

Fig. 3. Names and surnames of the characters: a – in the original novel, б – in the novel's translation

Source: compiled by A.S. Kuznetsova.

При сопоставительном анализе выявлены случаи значительной диспропорции в частотности употребления некоторых лексем. Наиболее показателен пример лексемы *шахматный* и ее эквивалента *shakmatni* в оригинальном тексте. Статистические данные демонстрируют принципиальное различие: показатель частотности для англоязычной лексемы составляет 41,39 ipm, в то время как ее эквивалент в русскоязычном переводе фиксируется с частотностью 2246,68 ipm.

С точки зрения лингвокультурной адаптации такая асимметрия объясняется различием в прагматических функциях лексем в исходном и переводном текстах. В оригинальном романе лексема *shakmatni* функционирует как лингвистический маркер, конструирующий образ Советского Союза, который посещает главная героиня для участия в турнире. В таком случае лексема выполняет роль специфического коннотативного элемента, усиливающего эффект культурной инаковости для англоязычного читателя. В русскоязычном переводе стилистически нейтральная лексема *шахматный* теряет эту маркирующую функцию, поскольку она не воспринимается носителем русского языка как экзотизм или культурно специфический знак. Она естественным образом интегрируется в повествование.

Демонстрация изучения Бэт Хармон русского языка служит важным сюжетным мотивом, подчеркивающим стратегический характер ее подготовки к противостоянию. Осознание того, что родным языком ее главного противника, Боргова, является русский, становится для героини стимулом к освоению этого языка, что позволяет ей к моменту поездки в СССР достичь значительного уровня понимания устной речи (пример 1).

(1) *She was enrolled as a special student at the university for a class in Russian*¹⁷.

Важную роль в конструировании образа советской действительности в оригинале играет вкрапление лексем, транслитерированных латиницей, таких как *blinchiki* и *Pravda* (примеры 2, 3). Данные единицы выполняют функцию лингвистических маркеров культурной специфики, создавая для англоязычного читателя эффект погружения в иную среду. Однако в русскоязычном переводе эти слова теряют свою экзотизирующую функцию, поскольку воспринимаются как нейтральные элементы родного языка.

(2) *Beth found one of the hotel restaurants — another carpeted gymnasium of a room — and managed enough Russian to order **blinchiki** and tea with blackberry jam*¹⁸.

(3) *It was **Pravda**. She couldn't penetrate the boldface Cyrillic of the headlines, but when she flipped it over, the bottom of page one had her picture on it, playing Flento*¹⁹.

Частотный анализ также подтверждает доминирование советского контекста в профессиональной сфере героини. Прилагательное *Russian* (347,64 ipm) регулярно встречается в сочетаниях, указывающих на источник ее шахматного образования и основной вызов. Как показывают примеры 4 и 5, героиня целенаправленно изучает труды советских мастеров (*Boleslavski, Luchenko*), что обосновывает ее рост.

(4) *She was able to use some of the things she had learned over the past weeks from studying her new book by the Russian Master Boleslavski*²⁰.

(5) *She spent an hour in her room with *The Middle Game in Chess, translated from the Russian of D. Luchenko**²¹.

¹⁷ Tevis W. *The Queen's Gambit*. New York: RosettaBooks, 2014. P. 210.

¹⁸ Там же. P. 112.

¹⁹ Там же. P. 61.

²⁰ Там же. P. 138.

²¹ Там же. P. 147.

Характерная лексическая сочетаемость этой лексики *Russian player, Russian grandmaster* устойчиво акцентирует тему профессионального превосходства советских шахматистов, представляя их как основную конкурентную силу и эталон мастерства в повествовании (примеры 6 и 7).

(6) *She had won her first game with a Russian grandmaster, and that was more important to her than anything she might see outside in the huge city*²².

(7) *The Russian players were at the other end with their wives*²³.

Таким образом, анализ показывает, что советские шахматисты, и в частности, Боргов, позиционируются как ключевые антагонисты, определяющие траекторию развития главной героини. Ответом на этот вызов становится ее решение изучить русский язык. Оригинальный текст активно использует стратегию транслитерации русских лексем, вкрапляя в англоязычное повествование русизмы, для создания эффекта культурной дистанции, которая очевидным образом нивелируется в переводе. При этом сам русский язык представлен в романе имплицитно: факт его использования персонажами реконструируется читателем исключительно через использование русского языка.

Репрезентация образа СССР в экранизации романа

У. Тевиса «The Queen's Gambit»

В ходе адаптации романа к аудиовизуальному формату произошла значительная тематическая переакцентуация. Если в литературном оригинале тема Советского Союза является важным, но не доминирующим контекстом, то в сериале она выдвигается на первый план, становясь центральным идеологическим и сюжетным элементом. Противостояние систем, воплощенное в фигурах Бэт Хармон и Василия Боргова, превращается в ключевую движущую силу всего повествования. Сюжет экранизации выстраивается как целенаправленная подготовка героини к кульминационному событию — матчу за звание чемпиона мира против советского гроссмейстера, что визуально подчеркивается в монтажной логике сериала (рис. 4).

Структурные изменения в экранизации носят принципиальный характер и служат драматургическим целям. В отличие от линейной хронологии романа, начинающегося с детства героини, первая серия использует прием предвосхищающей композиции. Зритель сразу погружается в напряженную атмосферу первой игры Бэт и Боргова. Такой прием позволяет сделать акцент на глубоком психологическом противостоянии: паника и смятение главной героини контрастируют с абсолютной невозмутимостью и хладнокровием советского чемпиона.

Подобная перестановка, нарушающая первоначальную хронологию, фокусирует внимание аудитории на судьбоносном значении данной встречи, задавая

²² Tevis W. *The Queen's Gambit*. New York: RosettaBooks, 2014. P. 100.

²³ Там же. P. 72.

высокий уровень драматического напряжения с самого начала и трансформируя сюжет из истории становления гения в напряженное идеологическое и психологическое противоборство.



Рис. 4. Кадр из сериала «Ход Королевы», показывающий встречу Бэт Хармон и Василия Боргова

Источник: Netflix. URL: <https://www.netflix.com/title/80234304> (дата обращения: 20.11.2025).

Fig. 4. A frame from the movie “The Queen’s Gambit”, showing the meeting between Bat Harmon and Vasily Borgov

Source: Netflix. Retrieved 20 November 2025, from <https://www.netflix.com/title/80234304>

В ходе анализа переводческих решений в локализации сериала была выявлена последовательная стратегия смягчения агрессивных или излишне прямых формулировок, относящихся к советским шахматистам. Яркой иллюстрацией служит эпизод, в котором мать Бэт интересуется стилем игры в СССР, а ее друзья дают характеристику потенциальным соперникам.

Таблица 1 / Table 1

Разговор о советских шахматистах / A conversation about Soviet chess players

Текст в оригинале / The original text of the movie	Текст в переводе на русский язык / Text in the movie’s translation into Russian
<p>— <i>What about Russia?</i> — <i>The Soviets are murder. They eat Americans for breakfast there. I don’t think there’s been an American with a prayer against them in 20 years. It’s like ballet. They pay people to play chess.</i></p>	<p>— <i>А в Союзе?</i> — <i>В Союзе все серьезно. Щелкают американцев как орешки. Ни один американец за последние 20 лет рядом с ними и близко не стоял. Это как балет. Там людям платят за шахматы.</i></p>

Источник: составлено А.С. Кузнецовой.

Source: compiled by A.S. Kuznetsova.

Переводчики отходят от буквального воспроизведения оригинальных высказываний, выбирая менее резкие эквиваленты. Вместо прямого перевода оборота, содержащего лексему *murder* (убийца), используется идиоматичное выражение *в Союзе все серьезно*. Далее образная гипербола *they eat Americans for breakfast there* (они там едят американцев на завтрак), намеренно грубая и провокационная в оригинале, заменяется на более привычную для русского языка метафору *щелкают американцев как орешки* (табл. 1).

Несмотря на то, что общий смысл о превосходстве советских игроков сохраняется, подобная замена приводит к нивелированию степени угрозы. В оригинале используется намеренно резкая риторика, которая в переводе трансформируется в более спортивную, соревновательную метафору. Это свидетельствует о сознательном смягчении тональности и снижении градуса идеологического противостояния в русскоязычной версии диалога (рис. 5).

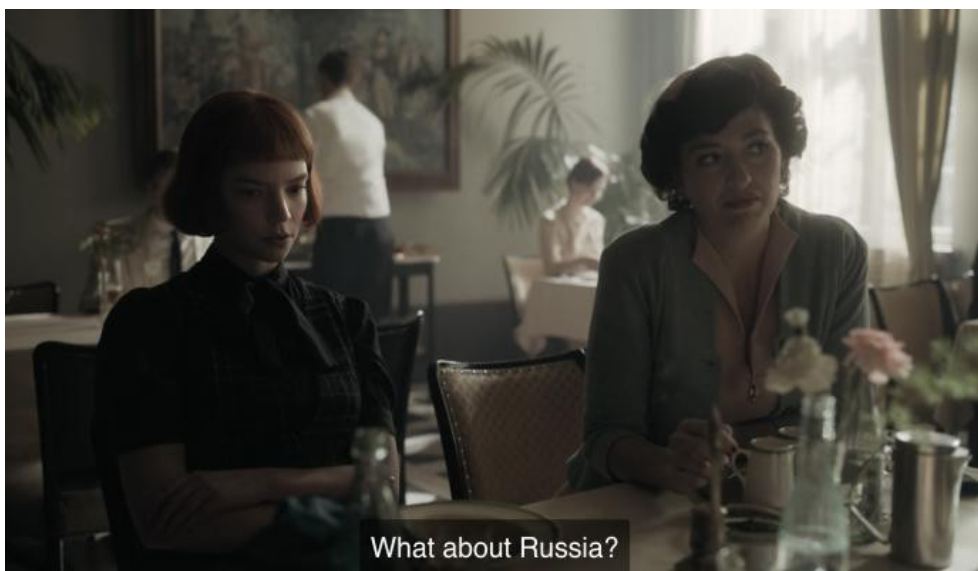


Рис. 5. Кадр из сериала «Ход Королевы», показывающий разговор о советских шахматистах
 Источник: Netflix. URL: <https://www.netflix.com/title/80234304> (дата обращения: 20.11.2025).

Fig. 5. A frame from the movie “The Queen’s Gambit”, showing a conversation about Soviet chess players
 Source: Netflix. Retrieved 20 November 2025, from <https://www.netflix.com/title/80234304>

В одном из эпизодов, в котором Бэт общается со знакомым с курсов русского языка, продемонстрированы характерные для западного сознания стереотипные представления о Советском Союзе. Диалог становится показателем противостояния социалистической и капиталистической систем, выраженного через ироничное самоопределение героя как *раба капитализма* (*a slave to capitalism*) (рис. 6).



Рис. 6. Кадр из сериала «Ход Королевы», в котором говорится о том, почему нужно учить русский язык

Источник: Netflix. URL: <https://www.netflix.com/title/80234304> (дата обращения: 20.11.2025).

Fig. 6. A frame from the movie "The Queen's Gambit" that explains the reasons for learning Russian

Source: Netflix. Retrieved 20 November 2025, from <https://www.netflix.com/title/80234304>

Параллельно в этой же коммуникативной ситуации проявляется иной, более сложный культурный стереотип, связанный с восприятием русской литературы. Упоминание желания героя прочитать Достоевского в оригинале служит индикатором всемирного культурного признания и автономной ценности русского классического наследия. Данный мотив вводит в повествование идею о существовании сферы интеллектуального и эстетического интереса, находящейся вне рамок непосредственного политического противостояния государств (табл. 2).

Таблица 2 / Table 2

Причины учить русский язык / Reasons to learn Russian

Текст в оригинале / The original text of the movie	Текст в переводе на русский язык / Text in the movie's translation into Russian
<p>— <i>So, why are you taking Russian classes if you're a slave to capitalism?</i></p> <p>— <i>I wanna read Dostoevsky in the original.</i></p>	<p>— <i>А тогда зачем ты учишь русский язык, если ты раб капитализма?</i></p> <p>— <i>Хочу читать Достоевского в оригинале.</i></p>

Источник: составлено А.С. Кузнецовой.

Source: compiled by A.S. Kuznetsova.

Георгий Гирев, первый советский шахматист, с которым сталкивается Бэт Хармон, представляет интерес для анализа с точки зрения лингвистической и культурной репрезентации. В оригинальной версии сериала данный персонаж характеризуется идиолектом, включающим грамматически правильную английскую речь с выраженным русским акцентом. Эта особенность может быть интерпретирована двояко: с одной стороны, как использование стереотипного клише о характерном «русском» произношении в англоязычной среде; с другой — как вполне реалистичная для исторического контекста деталь, обусловленная ограниченным доступом к аутентичным языковым материалам и практике живого общения (табл. 3). При этом общее высокое качество его речи косвенно указывает на хорошее образование, а также на уважительное отношение к языку и культуре соперника.

Таблица 3 / Table 3

Встреча с Григорием Гиревым / Meeting with Grigory Girev

Текст в оригинале / The original text of the movie	Текст в переводе на русский язык / Text in the movie's translation into Russian
— Good afternoon. I am Georgi Girev. (С русским акцентом)	— Добрый день. Я Георгий Гирев.

Источник: составлено А.С. Кузнецовой.

Source: compiled by A.S. Kuznetsova.

В процессе локализации сериала на русский язык произошла значительная трансформация данного лингвистического портрета. Вместо передачи акцента, значимого для оригинала, переводчики избрали стратегию перевода диалогов на русский язык, придав речи персонажа намеренную упрощенность и грамматическую небрежность (например, реплика *may I ask you something* преобразована в *можно задавать вопрос*). Подобное решение представляется неоднозначным с точки зрения переводческой адекватности, так как приводит не только к потере исходной лингвистической характеристики, но и к формированию искаженного образа, где носитель языка демонстрирует низкую грамотность в родной речи на фоне безупречной речи американской героини. Более релевантным решением, сохраняющим авторский замысел, могло бы стать сохранение оригинальной английской звуковой дорожки с субтитрованием.

Содержательно ключевой репликой персонажа становится вопрос о возможности смотреть кино в автомобиле в США. Этот бытовой, внешне шахматный интерес, проявленный после напряженной отложенной партии, выполняет важную функцию. Он гуманизирует образ советского соперника, смещая

акцент с образа «идеологического противника» на образ обычного ребенка, разделяющего детские интересы и любопытство к иной культуре (рис. 7).

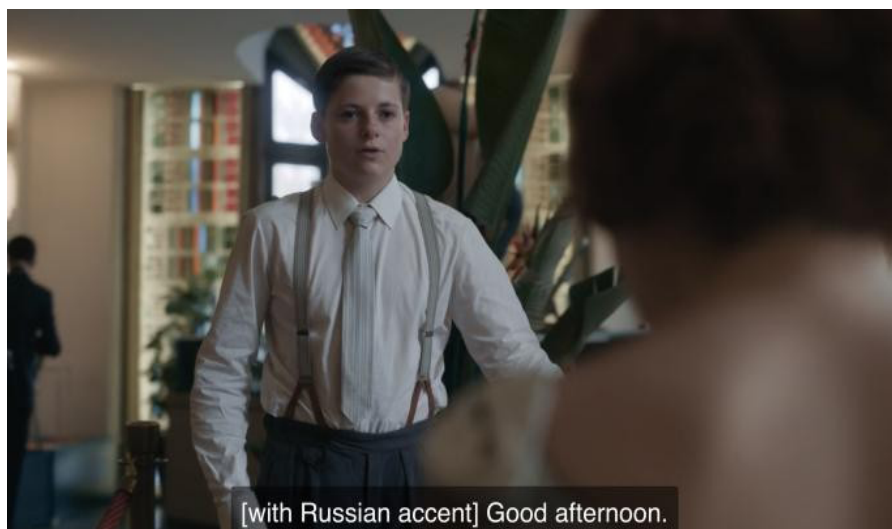


Рис. 7. Кадр из сериала «Ход Королевы», в котором показана встреча с Георгием Гиревым
 Источник: Netflix. URL: <https://www.netflix.com/title/80234304> (дата обращения: 20.11.2025).

Fig. 7. A frame from the movie “The Queen’s Gambit” showing a meeting with Georgy Girev
 Source: Netflix. Retrieved 20 November 2025, from <https://www.netflix.com/title/80234304>

Этот момент знаменует перелом в восприятии Гирева главной героиней. Бэт перестает видеть в нем исключительно абстрактную угрозу или продукт враждебной системы, начиная распознавать в нем индивидуальность. Таким образом, через призму лингвистического выбора и бытового диалога конструируется более сложный и неоднозначный образ, способствующий преодолению стереотипов.

В рассматриваемом эпизоде, где Бэт и Бенни совместно отрабатывают фразы повседневного общения на русском языке, был допущен принципиальный методологический просчет в процессе локализации. Исходная сцена построена как интерактивный учебный диалог: один персонаж произносит фразу на английском, другой осуществляет ее устный перевод на русский, и наоборот. Данный прием визуализирует процесс лингвистической подготовки героев.

Однако переводчики избрали стратегию полного перевода всех реплик на русский язык, включая те, что в оригинале звучат на английском как исходный материал для перевода. Это приводит к абсурду в восприятии русскоязычного зрителя: диалог, задуманный как демонстрация межъязыкового перевода, трансформируется в обмен семантически идентичными репликами на одном языке (табл. 4).

Таблица 4 / Table 4

Бэт и Бенни учат русский язык / Beth and Benny are learning Russian

Текст в оригинале / The original text of the movie	Текст в переводе на русский язык / Text in the movie's translation into Russian
<ul style="list-style-type: none"> — <i>The weather is hot today.</i> — <i>Сегодня жарко. (С английским акцентом)</i> — <i>I feel seek.</i> — <i>Мне плохо. (С английским акцентом)</i> 	<ul style="list-style-type: none"> — <i>Погода сегодня жаркая.</i> — <i>Сегодня жарко.</i> — <i>Мне нездоровится.</i> — <i>Мне плохо.</i>

Источник: составлено А.С. Кузнецовой.

Source: compiled by A.S. Kuznetsova.

В результате сцена утрачивает исходный смысл. Вместо отображения усилий по преодолению языкового барьера зритель наблюдает бессмысленную с точки зрения логики коммуникации имитацию диалога. Подобное решение не только нивелирует авторский замысел, но и создает когнитивный диссонанс, нарушая правдоподобие ситуации для целевой аудитории. Этот случай наглядно демонстрирует, как технически корректный, но концептуально неадекватный перевод может привести к полному искажению смысла целой сцены (рис. 8).



Рис. 8. Кадр из сериала «Ход Королевы», показывающий, как Бэт и Бенни учат русский язык
 Источник: Netflix. URL: <https://www.netflix.com/title/80234304> (дата обращения: 20.11.2025).

Fig. 8. A frame from the movie “The Queen’s Gambit” showing Bat and Benny learning Russian
 Source: Netflix. Retrieved 20 November 2025, from <https://www.netflix.com/title/80234304>

Данный фрагмент служит для демонстрации стереотипного представления о коллективистской модели поведения, приписываемой советскому социуму. Бенни подчеркивает необходимость понять принципы мышления русских и адаптировать поведение к их нормам для того, чтобы играть с ними на равных. В оригинальной версии эта мысль получает развитие через контраст, что можно увидеть в реплике: *Мы, американцы, работаем в одиночку, потому что мы все такие индивидуалисты (Us Americans, we work alone because we're all such individualists).*

Данную реплику можно трактовать как форму иронии или скрытой самокритики, направленной на ключевую ценность американской идентичности — индивидуализм. В высказывании Бенни проскальзывает насмешка над национальной особенностью, которая, в контексте противостояния с советской командой, предстает не как сила, а как потенциальная слабость (табл. 5, рис. 9).



Рис. 9. Кадр из сериала «Ход Королевы», в котором Бэт и Бенни обсуждают секрет успеха советских шахматистов

Источник: Netflix. URL: <https://www.netflix.com/title/80234304> (дата обращения: 20.11.2025).

Fig. 9. A frame from the movie "The Queen's Gambit" where Bat and Benny discuss the secret of Soviet chess players' success

Source: Netflix. Retrieved 20 November 2025, from <https://www.netflix.com/title/80234304>

Таблица 5 / Table 5

**Секрет успеха советских шахматистов /
The secret of Soviet chess players' success**

Текст в оригинале / The original text of the movie	Текст в переводе на русский язык / Text in the movie's translation into Russian
<p>— <i>The kind of chess played by the best players in the world, the Soviets.</i> — <i>And you know why they're the best players in the world?</i> — <i>They have the best suits?</i> — <i>It's because they play together as a team, especially during adjournments. They help each other out. Us Americans, we work alone because we're all such individualists. We don't like to let anyone help us.</i></p>	<p>— <i>Так играют лучшие шахматисты на планете. Советские.</i> — <i>А знаешь, почему у них это получается лучше всех?</i> — <i>У них костюмы лучше?</i> — <i>Они играют в команде, особенно на отложенной партии. Выручают друг друга. А в Америке мы работаем поодиночке. Мы же все индивидуалисты. И не позволяем никому нам помогать.</i></p>

Источник: составлено А.С. Кузнецовой.

Source: compiled by A.S. Kuznetsova.

В киноадаптации показаны довольно типичные пейзажи, которые ассоциируются с Россией у практически всех иностранцев. Это Кремль, Красная Площадь и гостиница Дворец Москва (рис. 10 и 11).

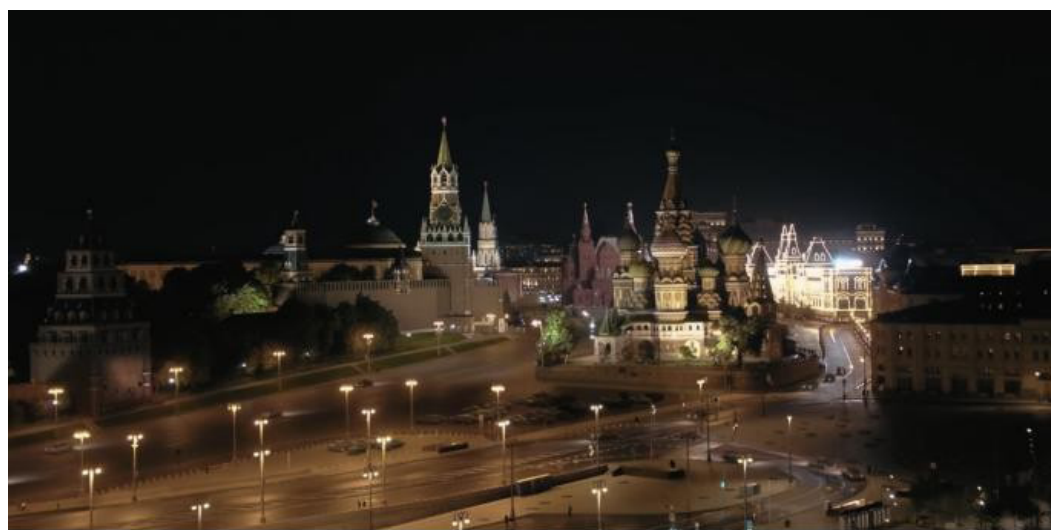


Рис. 10. Кадр из сериала «Ход Королевы», показывающий советские пейзажи
Источник: Netflix. URL: <https://www.netflix.com/title/80234304> (дата обращения: 20.11.2025).

Fig. 10. A frame from the movie “The Queen’s Gambit”, showing Soviet landscapes
Source: Netflix. Retrieved 20 November 2025, from <https://www.netflix.com/title/80234304>



Рис. 11. Кадр из сериала «Ход Королевы», показывающий советские пейзажи
 Источник: Netflix. URL: <https://www.netflix.com/title/80234304> (дата обращения: 20.11.2025).

Fig. 11. A frame from the movie “The Queen’s Gambit”, showing Soviet landscapes
 Source: Netflix. Retrieved 20 November 2025, from <https://www.netflix.com/title/80234304>

Кульминационная сцена, в которой Бэт Хармон одерживает победу над Василием Борговым, характеризуется сознательным отказом от вербальной экспрессии в пользу исключительно аудиовизуальных средств выразительности. Напряжение, эмоциональная разрядка и триумф передаются не через диалог, а через синтез операторской работы и музыкального сопровождения. Динамичная смена ракурсов, крупные планы, контрастное освещение и музыка берут на себя всю смысловую нагрузку, создавая эффект погружения зрителя в субъективное переживание момента.

Особого внимания заслуживает трансформация невербального поведения персонажа Боргова. На протяжении всего повествования он воплощает хладнокровную, отстраненную рациональность. Однако в финале его символический жест — объятие победительницы (рис. 12, а и б) — имеет большое значение. Этот невербальный акт выходит за рамки спортивного этикета и может быть интерпретирован как жест признания и преодоления идеологических границ. Таким образом, кульминация конфликта разрешается не на уровне идеологического противостояния систем, а на уровне личностного и профессионального уважения, что визуально подчеркивает центральную тему преодоления стереотипов.



а



б

Рис. 12. Кадры из сериала «Ход Королевы», показывающие решающую партию Бэт Хармон с Василием Борговым

Источник: Netflix. URL: <https://www.netflix.com/title/80234304> (дата обращения: 20.11.2025).

Fig. 12. A frame from the movie "The Queen's Gambit", showing the decisive match between Beth Harmon and Vasily Borgov

Source: Netflix. Retrieved 20 November 2025, from <https://www.netflix.com/title/80234304>

Заключение

Выполнен сравнительный анализ репрезентации образа Советского Союза в оригинальном романе У. Тевиса «The Queen's Gambit», его русскоязычном переводе и последующей экранизации. Основное внимание уделялось лингвистическим и аудиовизуальным средствам конструирования данного образа, а также проблемам его межъязыковой и межкультурной трансляции. В результате анализа сделаны следующие выводы.

Во-первых, сюжетно-идеологическим ядром произведения стало противостояние в эпоху холодной войны, реализуемое через образы шахматистов. Советские гроссмейстеры, и в особенности фигура Бориса Боргова, сюжетно позиционируются как ключевые антагонисты и главный вызов для протагонистки Бэт Хармон. Ее ответом на этот вызов становится изучение русского языка.

Во-вторых, в изученном материале активно эксплуатируются лингвистические стратегии маркировки «другого». В оригинальном художественном тексте образ СССР конструируется, в частности, через стратегию транслитерации русизмов (*blinchiki, Pravda*), что создает для англоязычного читателя эффект культурной дистанции и инаковости. В переводе романа эта специфическая маркировка неизбежно утрачивается, поскольку указанные лексемы воспринимаются русскоязычным читателем как обыденные, привычные, не маркированные культурно элементы родного языка. А языковой барьер и факт использования русской речи в оригинале передаются имплицитно — через авторские ремарки, а не прямое цитирование.

В-третьих, в адаптации Netflix визуальная репрезентация советского пространства становится значительно более детализированной и насыщенной. Образ создается через комплекс аудиовизуальных кодов: архитектуру, предметы быта, музыкальное сопровождение, невербальное поведение персонажей. Несмотря на минимальное использование русской речи, эти средства формируют плотный и узнаваемый визуальный контекст.

В-четвертых, были выделены некоторые проблемы перевода в аудиовизуальном формате. Локализация сериала на русский язык столкнулась с принципиальной сложностью — утратой лингвистической полифонии. В оригинале намеренное включение русской речи служит мощным средством аутентификации и маркировки «чужого» пространства. Полный перевод всего диалогового ряда на русский язык нивелирует этот эффект, лишая зрителя значимого слоя восприятия и приводя к неизбежным смысловым потерям или упрощениям в передаче культурных нюансов.

В-пятых, анализ языка персонажей и используемой ими лексики, в т.ч. с помощью такого инструмента как *ipm* (например, высокая частотность лексемы «шахматный») косвенно свидетельствует о значимости данного явления в об-

щественной жизни СССР и доминировании Советского Союза в этой области. В пользу этого также говорит эпизод, связанный с изучением героиней русского языка для чтения шахматной литературы. Однако победа главной героини над действующим чемпионом мира задает и легитимизирует нарративы американского превосходства. Факты же таковы, что в описываемую эпоху советские спортсмены были безусловно сильнейшими и удерживали шахматную корону с конца 40-х до начала 90-х гг. XX в. с коротким перерывом с 1972 по 1975 г., когда чемпионом мира был американец Р. Фишер.

Также необходимо отметить эволюцию восприятия и преодоление стереотипов, которые произошли в сознании главной героини. Если первоначальный образ СССР в сериале строится на ряде клишированных представлений, характерных для западного дискурса, то по мере развития сюжета зритель видит постепенное преодоление этих стереотипов. Через профессиональное взаимодействие и личный контакт происходит движение от восприятия советской стороны как угрозы к ее узнаванию как многогранной культурной реальности.

Таким образом, образ Советского Союза в исследуемом произведении и его киноадаптации представляет собой динамичный конструкт, формируемый взаимодействием вербальных, переводческих и визуальных стратегий. Исследование демонстрирует, как лингвистические особенности оригинала, проблемы межсемиотического перевода и медийная специфика киноязыка совместно влияют на восприятие историко-культурного контекста в современном международном культурном пространстве.

Список литературы

- Алюнина Ю.М. Где живут чудовища? Корпусный метод обнаружения англицизмов и их производных в русскоязычном Интернете // Вестник Томского государственного университета. Филология. 2022. № 80. С. 5–29. <https://doi.org/10.17223/19986645/80/1> EDN: LHFJGN
- Алюнина Ю.М. Структурные компоненты образа России в американском кино и методика их выявления // Метаморфозы истории. 2024. № 34. С. 1–16. DOI <https://doi.org/10.37490/S241436770032774-1> EDN: VPMATQ
- Алюнина Ю.М., Роговец А.А. Вербальная репрезентация образа России в американском кино: возможности и ограничения анализа // Славистика. 2024. Т. 28. № 2. С. 236–254 <https://doi.org/10.18485/slavistika.2024.28.2.15> EDN: FULUEW
- Алюнина Ю.М., Роговец А.А. Интермедиальное воплощение образа русского персонажа в американском фильме: портреты мужчины и женщины // Научный диалог. 2025. Т. 14. № 9. С. 162–188. <https://doi.org/10.24224/2227-1295-2025-14-9-162-188> EDN: IOGDPV
- Белов С.И. «Враг интерактивный»: американские видеоигры 1980-х гг. как ресурс позиционирования «советского чужого» // Электронный научно-образовательный журнал «История». 2024. Т. 15. № 2 (136). <https://doi.org/10.18254/S207987840030742-5> EDN: CSMNYJ
- Леонова О.Г. Джозеф Най и «мягкая сила»: попытка нового прочтения // Социально-гуманитарные знания. 2018. № 1. С. 101–114. EDN: YPLPTX
- Рябов О.В. Легитимация Холодной войны и массовая культура СССР и США // Имагология Холодной войны: «Свои» и «чужие» в массовой культуре СССР и США : сб. научных статей. СПб. : Изд-во РГПУ им. А.И. Герцена, 2024. С. 6–14. EDN: ZCIYIS
- Эко У. Сказать почти то же. СПб. : Симпозиум, 2006. 577 с.
- Якобсон Р.О. О лингвистических аспектах перевода // Вопросы теории перевода в зарубежной лингвистике. М. : Прогресс, 1985. С. 361–368.

- Bourdieu P. Ce que parler veut dire: L'économie des échanges linguistiques. Paris : Fayard, 1982. 239 p.
- Bourdieu P. The forms of capital // Handbook of Theory and Research for the Sociology of Education / ed. by J.G. Richardson. New York : Greenwood Press, 1986. P. 241–258.
- Hill S.P. A quantitative view of Soviet cinema // Cinema Journal. 1972. Vol. 11. № 2. Pp. 18–25. <https://doi.org/10.2307/1225047>
- Karamalak O., Vertlib K. Following the trace in fashion: From Zhivago look to Slavic core // Fashion communication in the digital age : Proceedings of the FACTUM 25 Conference, Stockholm, Sweden, 2025 (pp. 3–14) / ed. by von P. Wachenfeldt, L. Cantoni, N. Sabatini, T. Sádaba. Springer, 2025. https://doi.org/10.1007/978-3-031-99481-4_1
- McKay S.L. Teaching English as an international language: Rethinking goals and approaches. Oxford : Oxford University Press, 2002.
- Nye J.S. Jr. Bound to Lead: The Changing Nature of American Power. New York : Basic Books, 1991. 336 p.
- Nye J.S. Jr. Soft power // Foreign Policy. 1990. № 80. P. 153–171. <https://doi.org/10.2307/1148580> EDN: GWIBN
- Riabov O.V., Kulikova M.N. 'The voice of the enemy': using linguistic means for representing 'Amerika' in the Cold war Soviet films // Telekinet. 2025. № 2 (31). P. 23–29. <https://doi.org/10.24412/2618-9313-2025-231-23-29> EDN: PZHRQV

References

- Alyunina, Yu.M. (2022). Where do the wild things live? Corpus method to detect anglicisms and their derivatives on Russian Internet. *Tomsk State University Journal of Philology*, (80), 5–29. (In Russ.). <https://doi.org/10.17223/19986645/80/1> EDN: LHFGJN
- Alyunina, Yu.M. (2024). Structural components of Russian image in American cinema and methods of their identification. *Metamorphoses of History*, (34), 1–16. (In Russ.). <https://doi.org/10.37490/S241436770032774-1> EDN: VPMATQ
- Alyunina, Yu.M., & Rogovets, A.A. (2024). Verbal representation of the image of Russia in American cinema: Possibilities and limitations of analysis. *Slavistics*, 28(2), 236–254. (In Russ.). <https://doi.org/10.18485/slavistika.2024.28.2.15> EDN: FULUEW
- Alyunina, Yu.M., & Rogovets, A.A. (2025). Intermedial representation of a Russian character in an American film: Portraits of a man and a woman). *Nauchnyi dialog*, 14(9), 162–188. (In Russ.). <https://doi.org/10.24224/2227-1295-2025-14-9-162-188> EDN: IOGDPV
- Belov, S.I. (2024). “The interactive enemy”: American video games of the 1980s as a resource for positioning the “Soviet Other”). *Elektronnyy nauchno-obrazovatel'nyy zhurnal «Istoriya»*, 15(2). (In Russ.). <https://doi.org/10.18254/S207987840030742-5> EDN: CSMNYJ
- Bourdieu, P. (1982). *Ce que parler veut dire: L'économie des échanges linguistiques*. Fayard.
- Bourdieu, P. (1986). The Forms of Capita. In J.G. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood Press.
- Eco, U. (2006). *Skazat' pochti to zhe samoe* [To say almost the same thing]. Saint Petersburg: Symposium. (In Russ.).
- Hill, S.P. (1972). A quantitative view of Soviet cinema. *Cinema Journal*, 11(2), 18–25. <https://doi.org/10.2307/1225047>
- Jakobson, R. (1985). On linguistic aspects of translation. In *Voprosy teorii perevoda v zarubezhnoi lingvistike*. Moscow: Progress. (In Russ.).
- Karamalak, O., & Vertlib, K. (2025). Following the trace in fashion: From Zhivago look to Slavic core. In P. von Wachenfeldt, L. Cantoni, N. Sabatini, & T. Sádaba (Eds.), *Fashion communication in the digital age: Proceedings of the FACTUM 25 Conference, Stockholm, Sweden, 2025* (pp. 3–14). Springer. https://doi.org/10.1007/978-3-031-99481-4_1
- Leonova, O.G. (2018). Joseph Nye and “Soft power”: An attempt to read it again. *Social and Humanitarian Knowledge*, (1), 101–114. (In Russ.). EDN: YPLPTX
- McKay, S.L. (2002). *Teaching English as an international language: Rethinking goals and approaches*. Oxford: Oxford University Press.
- Nye, J.S., Jr. (1990). Soft power. *Foreign Policy*, (80), 153–171. <https://doi.org/10.2307/1148580> EDN: GWIBN
- Nye, J.S., Jr. (1991). *Bound to Lead: The Changing Nature of American Power*. Basic Books.
- Riabov, O.V., & Kulikova, M.N. (2025). “The voice of the enemy”: Using linguistic means for representing “Amerika” in the Cold War Soviet films. *Telekinet*, (2), 23–29. <https://doi.org/10.24412/2618-9313-2025-231-23-29> EDN: PZHRQV

Ryabov, O.V. (2024). Legitimation of the Cold War and Soviet and American mass culture. *Imagology of the Cold War: “Us” and “Them” in the mass culture of the USSR and the USA.* (pp. 6–14). Herzen State Pedagogical University of Russia. EDN: ZCIYIS

Информация об авторе:

КУЗНЕЦОВА Анна Сергеевна, лаборант-исследователь кафедры иностранных языков филологического факультета, Российский университет дружбы народов, Российская Федерация, 117198, Москва, ул. Миклухо-Маклая, д. 6. *Научные интересы:* корпусная лингвистика, когнитивная лингвистика, компьютерная лингвистика.

E-mail: kuznetsova-ansr@rudn.ru

ORCID: 0009-0003-7373-4056

SPIN-код: 4878-3885

Bionote:

Anna S. KUZNETSOVA, research assistant at the Department of Foreign Languages, Faculty of Philology, RUDN University, 6 Miklukho-Maklaya st., Moscow, 117198, Russian Federation. *Research interests:* corpus linguistics, cognitive linguistics, computer linguistics.

E-mail: kuznetsova-ansr@rudn.ru

ORCID: 0009-0003-7373-4056

SPIN-code: 4878-3885

DOI: 10.22363/2949-5997-2025-3-2-192-213

EDN HOJKYE

Научная статья / Research article

Символический капитал и конструирование «другого»: цвет как маркер образа России в американском кинематографе

Арина Анатольевна РОГОВЕЦ 

Российский университет дружбы народов, Москва, Российская Федерация

✉ RinaRogovets@yandex.ru

Аннотация. Процессы глобализации среди прочего направлены на унификацию семиотического пространства, что приводит к трансформации структур языка и культуры как символического капитала. Один из наиболее действенных инструментов мягкой силы — кинематограф, полимодальная семантика которого обладает эффективным набором различных инструментов воздействия на зрителя. Помимо развлекательной функции кинокартины могут использоваться в качестве инструмента «мягкой силы». Существующие исследования этого феномена фокусируются на сюжете и комплексном представлении событий и персонажей, репрезентация которых является центральным элементом картины. Изучение таких элементов репрезентации, как звучащий и написанный текст, костюмы, музыкальные произведения, постановка кадра и цвет, позволит сформировать наиболее эффективный набор инструментов воздействия. Цель исследования — проследить инструментализацию цвета при формировании образа России в американском кино. Материалом выступили американские фильмы, отобранные по подготовленным критериям. Проведение исследования в два этапа позволило рассмотреть 50 кинокартин с дальнейшим более детальным изучением 30 из них. Корпусный анализ текстов субтитров к кинофильмам выявил вхождения с упоминанием России, русских людей или элементов русской культуры. Впоследствии было решено рассмотреть визуальную составляющую в момент звучания реплик, а затем и применение цвета. В результате выявлена наибольшая частотность использования в кадре красного цвета с отсылкой к элементам государственной символики. Отмечено: с функциональной точки зрения цвет в большинстве случаев лишь дополняет и усиливает значение, переданное другими инструментами, а не формирует его самостоятельно. Полученные результаты показали, что выбор цвета не всегда обусловлен эстетическими требованиями, но может быть следствием макропроцессов социального и языкового характера, за которыми стоит языковое и культурное доминирование.

© Роговец А.А., 2025

This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

Ключевые слова: глобализация, рынок языков, визуальная семиотика, корпусная лингвистика, имагология, мягкая сила

Заявление о конфликте интересов. Автор заявляет об отсутствии конфликта интересов.

История статьи: получена 20 августа 2025; принята к публикации 20 сентября 2025.

Для цитирования: *Роговец А.А. Символический капитал и конструирование «другого»: цвет как маркер образа России в американском кинематографе // Macrosociolinguistics and Minority Languages. 2025. Т. 3. № 2. С. 192–213. <https://doi.org/10.22363/2949-5997-2025-3-2-192-213> EDN: HOJKYE*

Symbolic capital and the construction of the “Other”: color as a marker of the image of Russia in American cinema

Arina A. ROGOVETS 

RUDN University, Moscow, Russian Federation

✉ RinaRogovets@yandex.ru

Abstract. Globalization processes, among other things, aim to unify the semiotic space, leading to the transformation of the structures of language and culture as symbolic capital. One of the most effective instruments of soft power is cinema, whose multimodal semantics offers an effective set of diverse tools for influencing the viewer. Beyond entertainment, films can be used as a tool of «soft power». Existing research on this phenomenon focuses on plot and the complex presentation of events and characters, whose representation is the central element of the film. Studying such representational elements as spoken and written text, costumes, musical scores, framing, and color will help develop the most effective toolkit. The aim of this study is to trace the instrumentalization of color in the formation of the image of Russia in American cinema. American films, selected according to established criteria, served as the material. Conducting the study in two stages, 50 films were examined, with a subsequent, more detailed examination of 30 of them. A corpus analysis of film subtitles revealed instances of references to Russia, Russian people, or elements of Russian culture. Subsequently, it was decided that the visual component needed to be considered when the lines were being spoken, which led to a color study. This revealed that the most frequent use of red in the frame, referencing elements of national symbols, was the color. Functionally, color in most cases merely complements and enhances the meaning conveyed by other instruments, rather than shaping it independently. The results of this study show that color choice is not always determined by aesthetic requirements, but can be a consequence of social and linguistic macroprocesses, behind which lies linguistic and cultural dominance.

Key words: corpus linguistics, imagology, color in cinema, soft power

Conflict of interest. The author declares no conflict of interest.

Article history: received 20 August 2025; accepted 20 September 2025.

For citation: Rogovets, A.A. (2025). Symbolic capital and the construction of the “Other”: Color as a marker of the image of Russia in American cinema. *Macrosociolinguistics and Minority Languages*, 3(2), 192–213. (In Russ.). <https://doi.org/10.22363/2949-5997-2025-3-2-192-213> EDN: HOJKYE

Основная задача цвета — служить выразительности¹.
Анри Матисс

Введение

Процессы глобализации, магистральные для 1980–2000-х гг. и определяющие векторы мирового развития, обусловили научный интерес и необходимость исследования таких концептов, как культурное и языковое доминирование, лингвистический империализм, символический капитал. В кругу подобных концептов особое место занимает понятие «рынка языков» (Bourdieu, 1982), где доминирующее положение занимает английский. Английский язык, а также связанная с ним медиаиндустрия США создают условия для легитимации идеологического доминирования англосаксонской культуры и внедрение ее смыслов как на собственно языковом, так и, шире, на семиотическом уровне в иные языковые и культурные пространства. Результатом этих процессов становится унификация глобального семиотического пространства. Инструментом унификации может быть, среди прочего, создание образа «чужого» не только через вербальный ряд, но и через специфику выбора цвета для визуального ряда. Таким образом, цвет становится инструментом формирования национального образа в массовой культуре, поскольку, закрепляясь за определенными реалиями или ментальными конструктами, он превращается в символ и становится знаком и частью коммуникативной системы.

Исследованием смыслов цветообозначения занимается раздел семиотики, получивший название визуальной семиотики. Визуальные коммуникативные системы, как и любые системы естественных или искусственных языков, можно рассматривать как на микро-, так и на макроуровнях. Наделение цвета функциональной нагрузкой на микроуровне можно проиллюстрировать примером из литературы. В творчестве один и тот же цвет может использоваться для передачи разных смыслов. Так у Ф.М. Достоевского желтый известен как маркер бледности, болезненности, нищеты и вместе с тем как репрезентант государственных реформ и либерализма (Ерохина, 2021: 147, 149, 153; Мусаева, 2024: 170), а в современной русской поэзии желтый встречается в значениях «свет, счастье, любовь, красота» (Мусаева, 2024: 172). Если говорить о макроуровне, то здесь цвет, его восприятие, связанные с ним репрезентации, а также верба-

¹ Матисс А. Заметки живописца. 1908.

лизация данной визуальной системы могут стать значимыми в конструировании образа страны, ее культуры и языка, а значит соприкасаются с областью интересов макросоциолингвистики в перспективе конструирования мягкой силы и ее влияния на языковую и культурную ситуации.

Концепция мягкой силы (англ. *soft power*), введенная в 1990-е гг. Джозефом Наем и наиболее подробно изложенная в труде «Призвание к лидерству: меняющаяся природа американской мощи» (Nye, 1991) достаточно быстро приобрела популярность, поскольку обратила внимание на разнообразие инструментов, с помощью которых может быть оказано влияние на массовую аудиторию. Мягкая сила, по мнению Дж. Нае, как правило, черпается из таких ресурсов, как «культурная и идеологическая привлекательность» («*cultural and ideological attraction*») (Nye, 1990: 168). В качестве примеров автором приводятся коммунистическая идеология раннего поствоенного Советского Союза² (Nye, 1990: 167) и американская поп-культура: телевидение США за рубежом, англоязычные песни и одежда с надписями на английском языке по всему миру³ (Nye, 1990: 168–169).

Предложенный Дж. Наем спектр категорий мягкой силы перекликается с тремя основными видами капитала у П. Бурдьё — экономическим, культурным и социальным (Бурдьё, 2002: 60; Bourdieu, 1982; 1986: 243). Экономический капитал понимается П. Бурдьё в традиционном смысле — как финансовое благосостояние (Bourdieu, 1986: 242). Культурный капитал имеет наиболее сложную природу, которая выражается в трех ипостасях: *инкорпорированной* — в виде культурного капитала индивида; *институционализированной* — в виде формального выражения признания «культурного капитала индивида со стороны определенной организации» (диплом, звание, премия «Оскар» и др.); и *объективированной* — в виде совокупности «материальных объектов культуры и науки (книги, картины, фильмы и т.д.)» (Савченко, 2020: 64), каждый из которых ассоциируется с определенным автором или группой авторов и, соответственно, со страной происхождения. Социальный капитал предлагается рассматривать как некий конгломерат «реальных или потенциальных ресурсов» (Бурдьё, 2002: 66), обладание которыми обеспечивается членством в группе (Bourdieu, 1986: 248–249). Вхождение в группу предоставляет «опору в виде коллективного капитала» (Bourdieu, 1986: 248–249), именуемого репутацией. Дополнительно П. Бурдьё вводит понятие символического капитала, который, не являясь «самостоятельной формой капитала» (Савченко, 2020:

² «In the early postwar period, the Soviet Union profited greatly from such soft resources as communist ideology, the myth of inevitability, and transnational communist institutions» [В ранний послевоенный период Советский Союз извлек огромную выгоду из таких «мягких» ресурсов, как коммунистическая идеология, миф о неизбежности и транснациональные коммунистические институты. — здесь и ниже пер. наш. — А.А. Роговец] (Nye, 1990: 167).

³ «jackets with the names of American colleges» [куртки с названиями американских колледжей]; «Nicaraguan television broadcast American shows» [телевидение в Никарагуа транслирует американские шоу]; «Soviet teenagers wear blue jeans and seek American recordings» [советские подростки носят голубые джинсы и гоняются за американскими песнями] (Nye, 1990: 169).

64), используется в качестве своего рода разменной монеты для трех основных форм капитала. При этом последние могут брать на себя эту символическую функцию. Так «культурный капитал предрасположен функционировать в качестве символического капитала, т.е. оставаться непризнанным в качестве капитала и признаваться в качестве легитимной компетенции, в виде силы, влияющей на узнавание (или неузнавание)» (Бурдые, 2002: 62) определенного объекта, явления, образа.

Возвращаясь к концепции мягкой силы Дж. Ная, можно сказать, что ее ресурсом служит культурный капитал, элементы которого способны выполнять символическую функцию в репрезентации того или иного объекта. Репрезентация в свою очередь определяется как «представление познаваемого явления с помощью посредников»⁴, как модели, символы, знаки, рисунки, схемы, цвета и др. Форма такого «посредника» зависит от сферы ее реализации или использования, а назначение всегда состоит в «замещении чего-то, находящегося за ней»⁴. Язык и культура также могут быть рассмотрены в качестве объекта символической капитализации, а кино, как наиболее массовый и доступный вид искусства, может быть удобным инструментом такой капитализации.

Кино относят к аудиовизуальным произведениям, которые представляют собой «мультимодальный текст, объединяющий в себе изображение, звук и вербальные средства, то есть разные модусы или средства передачи смысла» (Александрова, 2023: 156). Кинокартины способны оказывать влияние на зрителя посредством одновременного использования целого набора инструментов. Выделяют следующие каналы восприятия информации и соответствующие им инструменты передачи информации: аудиальный, визуальный, кинестетический (Барило, Христенко, Рахимов, 2023: 263). Киноискусство способно задействовать одновременно аудиальный и визуальный каналы, а в отдельных случаях и кинестетический, если мы говорим об иммерсивном кинотеатре⁵. Так, одна идея может быть представлена на нескольких уровнях, каждый из которых способен раскрыть ее со своей стороны. Вместе с тем кино способно выполнять не только развлекательную функцию, но и выступать инструментом «мягкой силы», под которой понимается «способность влиять на то, чего хотят другие страны», связанная «с нематериальными ресурсами, такими как культура, идеология и институты» (Леонова, 2018: 105).

Несмотря на то, что голливудские компании-кинопроизводители, такие как Disney, Sony, Warner Brothers, Universal и др., после февраля 2022 г. приостановили сотрудничество с кинотеатрами в России, спрос на американское кино не упал. Опрос, проведенный в 2022 г., показал, что среди 1700 россиян старше 18 лет

⁴ Касавин И.Т. Энциклопедия эпистемологии и философии науки. М.: Канон, 2009. С. 827.

⁵ В Москве открылся первый иммерсивный кинотеатр для профориентации старшеклассников // Мой социальный помощник. URL: <https://msp.dszn.ru/news/kago89rxn1-v-moskve-otkrilsya-pervii-v-rossii-immer> (дата обращения: 14.08.2025).

59 % респондентов ожидает возвращения зарубежных картин в кинотеатры, а 67 % — предпочли кинокартины производства США⁶. Такой масштаб охвата аудитории говорит о том, что репрезентации, транслируемые американским кинематографом, могут закрепляться в мировосприятии зрительской аудитории любого возраста. Потому важно изучать инструменты, используемые для репрезентации России, и ее непосредственный образ в продуктах зарубежной массовой культуры.

Целью стало исследование цвета как элемента, дополняющего вербальный образ России в американском кинематографе, в дальнейшей перспективе анализа инструментов мягкой силы на символическом рынке языкового и культурного капиталов. Исследование является вкладом в проект по изучению репрезентации образа России в американском кино⁷. Результаты этого проекта позволили выделить ключевые тенденции в представлении образа России через портреты русского мужчины и русской женщины (Алюнина, Роговец, 2025), а также общие тенденции в репрезентации России в кинематографических произведениях США (Алюнина, Роговец, 2024: 248; Алюнина, 2024: 9). В ходе работы над проектом было замечено, что цвет, среди прочего, играет важную роль в репрезентации образа России в кинематографе, являясь визуальным средством передачи настроения, маркером определенных событий, а также инструментом непрямого, эксплицитного ассоциирования того или иного объекта, явления, персонажа с Россией.

Материал и методы исследования

Материалом исследования стал собственный корпус, включающий 50 кинокартин, отобранных по критериям, утвержденным в рамках указанного выше проекта (Алюнина, Роговец, 2024: 241–242; Алюнина, 2024: 8). Критерии выборки приведены ниже.

1. Страна-производитель. При выборе кинокартин для анализа приоритет отдавался фильмам и сериалам, созданным американскими кинокомпаниями, однако допустимым было совместное кинопроизводство с Великобританией и Канадой, в случае если первым производителем были указаны США. Подобный формат создания кинокартин в XXI в. набирает популярность, что является еще одним показателем ориентированности кинематографического искусства на массовую аудиторию. В связи с этим было решено не ограничивать выборку фильмами, созданными исключительно в США, чтобы рассмотреть репрезентацию России более подробно.

⁶ «Аватар», «Брат», «Любовь и голуби»: россияне рассказали, какие фильмы хотят видеть в кинотеатрах // Анкетолог. URL: <https://iom.anketolog.ru/2022/05/17/kinematograf> (дата обращения: 18.12.2025).

⁷ Языковая репрезентация образа России в американском кинематографе: корпусный подход // Российский научный фонд. URL: https://www.rscf.ru/project/E_5AoRIVKOTbCqRKJmw_pioxqoaQKETpWsgXkjRzjz_NHJEGISnYVQ03hr1owA6Q1sE1hLC3/ (дата обращения: 19.08.2025).

2. Сюжет. Помимо кинокартин, чей сюжет полностью связан с Россией или отдельными событиями из ее истории, в выборку были включены фильмы, действующим лицом которых является русский персонаж в главной либо значимой второстепенной роли. Также включались картины с Россией в качестве места событий.

3. Режиссерское кино. Под ним понимаются «фильмы, где режиссер выступает как главный автор, полностью реализуя свое видение через управление всеми аспектами создания, от сценария до монтажа, создавая целостное авторское произведение»⁸. Важным условием в рамках этого критерия было исключение фильмов — экранизаций произведений русской литературы во избежание влияния на режиссера взглядов автора художественного текста как носителя русской культуры.

4. Англоязычные субтитры. Все кинематографические произведения, которые попали в исследовательскую выборку, должны были иметь англоязычные субтитры, необходимые для составления корпуса субтитров к фильмам в целях дальнейшего их анализа при помощи корпусного менеджера *Sketch Engine*⁹. Суть этого анализа описана далее.

Анализ корпуса фильмов производился в два этапа.

На первом этапе использовался корпусный менеджер *Sketch Engine*, а именно его функция *Word Sketch*, которая позволяет обнаруживать сочетания с употреблением конкретной лексической единицы в правом или левом контексте. В рамках описываемого исследования в *Word Sketch* выполнялся поиск лексем с корнем *-Russ-* в роли существительного и прилагательного. В результате в корпусе субтитров к фильмам были обнаружены сочетания с лексемами с корнем *Russ-* в роли подлежащего, дополнения и обстоятельства (*a Russian, the Russians, Russia, Russia's, in Russia, in Russian* и др.) и с лексемами с корнем *Russ-* в роли определения (*Russian bear, Russian woman, Russian patriot, Russian space, Russian desk* и др.).

Первичное рассмотрение полученных сочетаний выявило необходимость расширения контекста до целостных смысловых отрывков и их дальнейшего анализа в сопоставлении с видеорядом и сюжетом конкретного аудиовизуального произведения для получения более точных результатов о репрезентации образа России в американском кинематографе. Выбранный подход от контекста или от материала в корпусной лингвистике принято называть *corpus-driven* (дословно — ведомый корпусными данными), который, в отличие от *corpus-based* (дословно — основанный на корпусных данных), заключается в анализе

⁸ Алимтаева А. Зачем нужно смотреть фильмы в режиссерской версии // Steppe. URL: [https://the-steppe.com/razvlecheniya/zachem-nuzhno-smotret-filmy-v-rezhisserskoj-versii-ili-zachem-kromsayut-kino#:~:text=Режиссёрская%20версия%20—%20это%20фильм%2C%20который,только%20один%20человек%20—%20сам%20режиссер\(дата%20обращения:18.08.2025\).](https://the-steppe.com/razvlecheniya/zachem-nuzhno-smotret-filmy-v-rezhisserskoj-versii-ili-zachem-kromsayut-kino#:~:text=Режиссёрская%20версия%20—%20это%20фильм%2C%20который,только%20один%20человек%20—%20сам%20режиссер(дата%20обращения:18.08.2025).)

⁹ Sketch Engine. URL: <https://www.sketchengine.eu/> (дата обращения: 18.08.2025).

языковых фактов «с целью обоснования и обобщения полученных результатов и выхода на новые гипотезы» (Дмитриев, Коган, Вдовина, 2020: 212).

В ходе изучения аудиовизуального ряда в контексте формирования образа России на экране выявлено специфическое использование цвета в кадре, что стало фокусом второго этапа исследования. Так от анализа вербальных упоминаний России был совершен переход к изучению невербальных элементов, в частности цвета, который выступает одним из структурных элементов в построении рассматриваемой репрезентации.

При первичном анализе фильмов выявили 415 текстовых фрагментов и соответствующих им отрывков видеоряда в 30 кинокартинах, в которых показан образ России (табл. 1).

Таблица 1

Перечень источников исследовательского материала

№	Название фильма	Год	Режиссер	Страны
1	«Красная жара»	1988	Уолтер Хилл	США
2	«Мстители»	2012	Джосс Уидон	США
3	«Москва на Гудзоне»	1984	Пол Мазурски	США
4	«Белые ночи»	1985	Тейлор Хэкфорд	США
5	«Рокки 4»	1985	Сильвестр Сталлоне	США
6	«Русский отдел»	1990	Фред Скеписи	США
7	«Армагеддон»	1998	Майкл Бэй	США
8	«Ложное искушение»	2006	Роберт Де Ниро	США
9	«Индиана Джонс и Королевство хрустального черепа»	2008	Стивен Спилберг	США
10	«Железный человек 2»	2010	Джон Фавро	США
11	«Солт»	2010	Филлип Нойс	США
12	«Порок на экспорт»	2007	Дэвид Кроненберг	Великобритания, Канада, США
13	«Красный воробей»	2018	Фрэнсис Лоуренс	США
14	«Миссия невыполнима: Протокол Фантом»	2011	Брэд Берд	США
15	«Жертвуй пешкой»	2014	Эдвард Цвик	США
16	«Агенты А.Н.К.Л.»	2015	Гай Ричи	Великобритания, США

Окончание табл. 1

№	Название фильма	Год	Режиссер	Страны
17	«Профессионал»	2018	Мэттью М. Росс	Канада, США
18	«Клаустрофобы: Квест в Москве»	2020	Уилл Верник	США
19	«Касабланка»	1942	Майкл Кертиц	США
20	«Довод»	2020	Кристофер Нолан	Великобритания, США
21	«Черная Вдова»	2021	Кейт Шортланд	США
22	«Полицейская академия: Миссия в Москве»	1994	Алан Меттер	США
23	«Охота за «Красным Октябрем»»	1990	Джон Мактирнан	США
24	«Красный рассвет»	1984	Джон Милиус	США
25	«Святой»	1997	Филлип Нойс	США
26	«Великий уравниатель»	2014	Антуан Фукуа	США
27	«Золотой глаз»	1995	Мартин Кэмпбелл	Великобритания, США
28	«Эвакуация»	2013	Тони Джильо	США
29	«С любовью, Антоша»	2019	Гаррет Прайс	США
30	«До скорой встречи»	2019	Дэвид Махмуди	США

Источник: составлено А.А. Роговец.

Table 1

List of sources of research material

№	Movie title in English	Year	Director	Countries
1.	"Red Heat"	1988	Walter Hill	USA
2.	"The Avengers"	2012	Joss Whedon	USA
3.	"Moscow on the Hudson"	1984	Paul Mazursky	USA
4.	"White Nights"	1985	Taylor Hackford	USA
5.	"Rocky IV"	1985	Sylvester Stallone	USA
6.	"The Russia House"	1990	Frederic Schepisi	USA
7.	"Armageddon"	1998	Michael Bay	USA

Continuing table 1

№	Movie title in English	Year	Director	Countries
8.	"The Good Shepherd"	2006	Robert De Niro	USA
9.	"Indiana Jones and the Kingdom of the Crystal Skull"	2008	Steven Spielberg	USA
10.	"Iron Man 2"	2010	Jonathan Favreau	USA
11.	"Salt"	2010	Phillip Noyce	USA
12.	"Eastern Promises"	2007	David Cronenberg	United Kingdom, Canada, USA
13.	"Red Sparrow"	2018	Francis Lawrence	USA
14.	"Mission: Impossible — Ghost Protocol"	2011	Bradley Bird	USA
15.	"Pawn Sacrifice"	2014	Édward Zwick	USA
16.	"The Man from U.N.C.L.E."	2015	Guy Ritchie	United Kingdom, USA
17.	"Siberia"	2018	Matthew Ross	Canada, USA
18.	"Follow Me"	2020	Will Wernick	USA
19.	"Casablanca"	1942	Kertész Mihály	USA
20.	"Tenet"	2020	Christopher Nolan	United Kingdom, USA
21.	"Black Widow"	2021	Cate Shortland	USA
22.	"Police Academy: Mission to Moscow"	1994	Alan Metter	USA
23.	"The Hunt for Red October"	1990	John McTiernan	USA
24.	"Red Dawn"	1984	John Milius	USA
25.	"The Saint"	1997	Phillip Noyce	USA
26.	"The Equalizer"	2014	Antoine Fuqua	USA
27.	"Golden Eye"	1995	Martin Campbell	United Kingdom, USA
28.	"Extraction"	2013	Tony Giglio	USA

Ending table 1

№	Movie title in English	Year	Director	Countries
29.	"Love, Antosha"	2019	Garret Price	USA
30.	"See You Soon"	2019	David Mahmoudieh	USA

Source: compiled by A.A. Rogovets.

На втором этапе исследования появилась необходимость в разработке критериев, позволяющих считать использование цветowych элементов целенаправленным, т.е. применять их в качестве инструмента для репрезентации образа России.

Для анализа цвета в качестве инструмента конструирования образа России сформулировали следующие критерии.

1. Пространство в кадре и контраст. Оттенок должен занимать в кадре от 40 % площади и выше. В спорных ситуациях, где площадь покрытия оттенка граничила с выбранным значением, на кадр накладывалась виртуальная сетка, позволяющая определить, удовлетворяется ли критерий отбора (рис. 1). Также отбирались кадры, где цвет выделялся за счет контраста с соседствующими ему, определяемого с использованием цветового колеса Иттена (рис. 2).

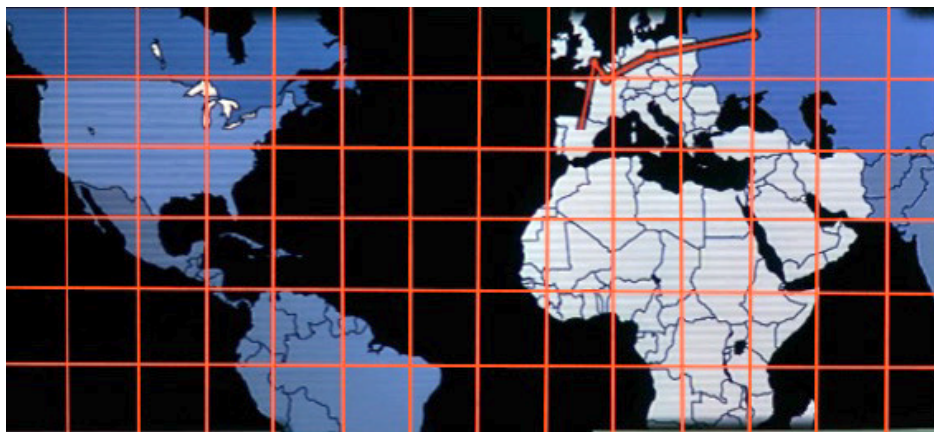


Рис. 1. Кадр из кинокартины «Golden Eye», 1995

Источник: Кинопоиск. URL: <https://www.kinopoisk.ru/film/3556/> (дата обращения: 12.08.2025).

Fig. 1. A frame from the movie "Golden Eye", 1995

Source: Kinopoisk. Retrieved 12 August 2025, from <https://www.kinopoisk.ru/film/3556/>

2. Специфика. Цвет не должен принадлежать уже имеющим конвенциональную смысловую нагрузку в контексте сцены объектам, таким как, например, красная дорожка, красная форма сотрудников отеля и т.д.



Рис. 2. Изображение цветового колеса Иттена
 Источник: Skillbox Media. URL: https://skillbox.ru/media/design/chto_takoe_tsvetovoy_krug_ittena/
 (дата обращения: 15.08.2025).

Fig. 2. Image of the Itten color wheel
 Source: Skillbox Media. Retrieved 15 August 2025, from https://skillbox.ru/media/design/chto_takoe_tsvetovoy_krug_ittena/

С учетом названных выше критериев среди 415 кадров, в которых присутствует вербальная отсылка к «русскому», выделили кадры, где цвет выступает элементом, конструирующим образ России в американском кинематографе (табл. 2).

2. Специфика. Цвет не должен принадлежать уже имеющим конвенциональную смысловую нагрузку в контексте сцены объектам, таким как, например, красная дорожка, красная форма сотрудников отеля и т. д.

С учетом названных выше критериев среди 415 кадров, в которых присутствует вербальная отсылка к «русскому», выделили кадры, где цвет выступает элементом, конструирующим образ России в американском кинематографе (табл. 2).

Таблица 2

Соотношение кадров с цветом, подлежащих анализу, с общим числом кадров

Общее число кадров	Кадры, где цвет выступает инструментом передачи смысла	
	Количество кадров	Доля от общего числа, %
415	18	4

Источник: составлено А.А. Роговец.

Table 2

The proportion of color frames to be analyzed to the overall number

Overall number of frames	Frames where color acts as a tool for conveying meaning	
	Number of frames	Percentage of the overall number, %
415	18	4

Source: compiled by A.A. Rogovets.

Таким образом, выявлено и проанализировано 18 кадров, в которых цвет выступает инструментом конструирования образа России, что составляет лишь 4 % от общего числа кадров с вербальной отсылкой к «русскому». Подобное соотношение дает основания сформулировать гипотезу о том, что цвет является комплементарным инструментом формирования рассматриваемой репрезентации.

Основные цвета, задействованные в построении кинематографического образа России, приведены ниже (табл. 3).

Таблица 3

Распределение кадров по цветам

Цвет в кадре	Количество кадров	Доля от общего числа, %
Красный	10	55,6
Синий	6	33,3
Зеленый	2	11,1

Источник: составлено А.А. Роговец.

Table 3

The distribution of frames by color

Color in the frame	Number of frames	Percentage of the overall number, %
Red	10	55.6
Blue	6	33.3
Green	2	11.1

Source: compiled by A.A. Rogovets.

Красный цвет как смысловой инструмент

Кадры, в которых присутствуют указанные цвета, разделили на три основные группы: красный, зеленый и синий, в соответствии с цветовой моделью RGB (red, green, blue), которая представляет собой «аддитивную модель», что значит, что «цвета создаются за счет добавления одного к другому: красного — red, зеленого — green, и синего — blue»¹⁰. Оттенки цветов не учитывались.

Рассмотрим наиболее репрезентативные примеры включения в кадр красного цвета (рис. 3 и 4, табл. 4 и 5).

¹⁰ Гринмирис А., Соловьева Ж., Павлова А. Что такое цветовое пространство: виды и настройка // Яндекс Практикум. URL: <https://practicum.yandex.ru/blog/chto-takoe-cvetovoe-prostranstvo-podrobnyj-razbor/> (дата обращения: 19.08.2025).

В кинокартине «Golden Eye» на экране компьютера показан профиль генерала Урумова (см. рис. 3). По сюжету на него падают подозрения в организации крушения летательных аппаратов. Помимо красной полосы на фуражке для сопровождения образа русского персонажа выбран красный фон. Зеленый цвет, противоположный красному по цветовому колесу Иттена (см. рис. 2), повышает контраст и усиливает акцент на персонаже. Цветной элемент головного убора является частью формы государственного служащего и, соответственно, связан с государственной символикой. Красный — основной цвет флага СССР и часть флага Российской Федерации. Нет официально принятой и единственной трактовки значения цвета в контексте символики, однако существует мнение о том, что красный ассоциируется с «энергией, силой и мужеством»¹¹, а также «с кровью и огнем, а следовательно с опасностью» (Пастуро, 2019: 34). В психологии принято считать, что красный цвет ассоциируется с агрессией и озлобленностью. Таким образом, Урумов представлен властным и сильным человеком как с точки зрения цветовой репрезентации, так и с точки зрения вербальной — *считает себя новым правителем России; не вписывается в образ предателя* (см. табл. 4).



Рис. 3. Кадр из кинокартины «Golden Eye», 1995

Источник: Кинопоиск. URL: <https://www.kinopoisk.ru/film/3556/> (дата обращения: 12.08.2025).

Fig. 3. A frame from the movie "Golden Eye", 1995

Source: Kinopoisk. Retrieved 12 August 2025, from <https://www.kinopoisk.ru/film/3556/>

¹¹ Эволюция триколора: ко Дню государственного флага // Культура Петербурга. URL: <https://spbcult.ru/articles/istoriya-peterburga/evolyutsiya-trikolora-ko-dnyu-gosudarstvennogo-flaga/> (дата обращения: 28.07.2025).

Таблица 4 / Table 4

**Реплика в момент демонстрации кадра в фильме «Golden Eye» /
A line from the film “Golden Eye” as it was being shown**

Оригинал / Original	Звучащий перевод / A sounding translation
<i>Ourumov. They made him a general. He sees himself as the next “iron man of Russia”. Our political analysts say he doesn’t fit the profile of a traitor.</i>	<i>Урумов. Уже генерал. Считает себя новым правителем России. Наши аналитики исключают его причастность, не вписывается в образ предателя.</i>

Источник: составлено А.А. Роговец.

Source: compiled by A.A. Rogovets.

В следующем примере задействован цвет в репрезентации русского персонажа на экране американского фильма (рис. 4, табл. 5).



Рис. 4. Кадр из кинокартины «The Saint», 1997

Источник: Кинопоиск. URL: <https://www.kinopoisk.ru/film/4133/> (дата обращения: 12.08.2025).

Fig. 4. A frame from the movie “The Saint”, 1997

Source: Kinopoisk. Retrieved 12 August 2025, from <https://www.kinopoisk.ru/film/4133/>

В кинокартине «The Saint» промышленный магнат Иван Третьяк выступает с речью перед зрителями. Фон за ним красного цвета, на полотне надпись: «Демократические реформы — сегодня, сильная Россия — завтра». Кроме того, на сцене можно заметить представителя церкви и мужчину в военной форме. События фильма разворачиваются в постсоветской Москве. В этой связи красный цвет интерпретируется как инструмент, усиливающий силу и власть в образе главного героя, подкрепляемые вербальной составляющей его репрезентации (табл. 5).

Таблица 5 / Table 5

**Реплика в момент демонстрации кадра в фильме «The Saint» /
A line from the film “The Saint” as it was being shown**

Оригинал / Original	Звучащий перевод / A sounding translation
<i>There were three great empires that dominated the world, both culturally and militarily: Rome, Constantinople and Russia. All three have fallen. Only one can be restored.</i>	<i>Три империи доминировали в мире в культурном и военном плане: Рим, Константинополь и Россия. Все три пали и только одну можно возродить.</i>

Источник: составлено А.А. Роговец.

Source: compiled by A.A. Rogovets.

Таким образом, в 7 случаях из 10 красный цвет в кадре был частью государственной символики СССР, что предположительно направлено на усиление образа сильного и влиятельного государства. В иных случаях контекст ситуаций позволяет сделать вывод о использовании цвета с целью указать на опасность или агрессию.

Синий цвет как инструмент передачи смысла

Перейдем к рассмотрению роли синего цвета в конструировании кинематографического образа России.

В кинокартине «The Saint» звучит сводка новостей о России и в этот момент показаны кадры с обстановкой в стране (рис. 5, табл. 6). В кадре преобладает синий цвет, который ассоциируется со спокойствием и меланхолией, что пришло из эпохи романтизма (Пастуро, 2017: 41). Позднее синий начал считаться символом уверенности и диалога (Пастуро, 2017: 41). В контексте ситуации цвет создает ощущение грусти и безвыходности, подчеркивает, что ситуация в стране напряженная.



Рис. 5. Кадр из кинокартины «The Saint», 1997

Источник: Кинопоиск. URL: <https://www.kinopoisk.ru/film/4133/> (дата обращения: 12.08.2025).

Fig. 5. A frame from the movie “The Saint”, 1997

Source: Kinopoisk. Retrieved 12 August 2025, from <https://www.kinopoisk.ru/film/4133/>

Таблица 6 / Table 6

**Реплика в момент демонстрации кадра в фильме «The Saint» /
A line from the film “The Saint” as it was being shown**

Оригинал / Original	Звучащий перевод / A sounding translation
<i>The troops will depose the president and install Ivan tretiak as leader of Russia.</i>	<i>Военные скинут действующего президента и сделают главой государства Ивана Третьяка.</i>

Источник: составлено А.А. Роговец.

Source: compiled by A.A. Rogovets.

В кинокартине «Armageddon» у шаттла, летящего к астероиду, возникают проблемы с двигателем. Российскому космонавту Льву Андропову удастся решить эту проблему, благодаря чему всей команде, включая американцев, удастся выжить. Кадр, показывающий данную ситуацию, окрашен в синий цвет, который в данном контексте может быть интерпретирован как символ диалога и взаимопонимания между русским и американскими космонавтами (рис. 6, табл. 7).



Рис. 6. Кадр из кинокартины «Armageddon», 1998

Источник: Кинопоиск. URL: <https://www.kinopoisk.ru/film/2941/> (дата обращения: 12.08.2025).

Fig. 6. A frame from the movie “Armageddon”, 1998

Source: Kinopoisk. Retrieved 12 August 2025, from <https://www.kinopoisk.ru/film/2941/>

Таблица 7 / Table 7

**Реплика в момент демонстрации кадра в фильме «Armageddon» /
A line from the film “Armageddon” as it was being shown**

Оригинал / Original	Звучащий перевод / A sounding translation
<i>Thank you. Thank you, thank you, thank you, thank you. Lev! Now I'm really a Russian hero!</i>	<i>Слава тебе господи! Слава! Слава! Слава! Слава! Лев! Теперь я настоящий русский герой!</i>

Источник: составлено А.А. Роговец.

Source: compiled by A.A. Rogovets.

Так, в равном числе кадров с учетом контекста синий символизирует диалог и взаимопонимание, также, как и спокойствие, меланхолию и грусть.

Зеленый цвет как инструмент передачи смысла

В кинокартине «Armageddon» во время миссии по присоединению к российской космической станции теряется связь с командой американских космонавтов, сотрудники с земли пытаются ее восстановить. Зеленый цвет как правило ассоциируется со спокойствием и безопасностью (Пастуро, 2024: 50). В данном контексте использование цвета направлено на формирования образа надежности американских государственных структур, что контрастирует с сотрудниками русской космической службы, а именно с космонавтом Львом Андроповым, который в момент миссии находится в состоянии алкогольного опьянения (рис. 7, 8, табл. 8, 9).

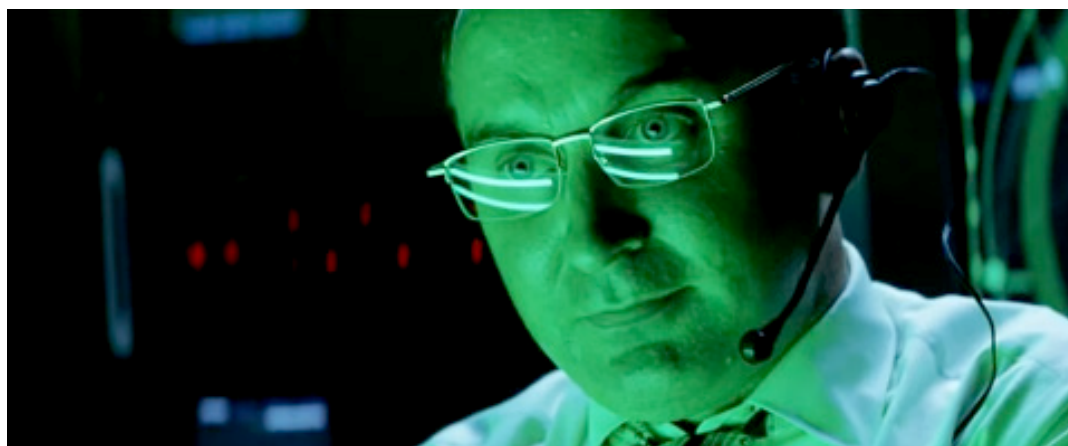


Рис. 7. Кадр из кинокартины «Armageddon», 1998

Источник: Кинопоиск. URL: <https://www.kinopoisk.ru/film/2941/> (дата обращения: 12.08.2025).

Fig. 7. A frame from the movie “Armageddon”, 1998

Source: Kinopoisk. Retrieved 12 August 2025, from <https://www.kinopoisk.ru/film/2941/>

Таблица 8 / Table 8

**Реплика в момент демонстрации кадра в фильме «Armageddon» /
A line from the film “Armageddon” as it was being shown**

Оригинал / Original	Звучащий перевод / A sounding translation
<i>And after seven minutes, we may never hear from them again. But we're trying to bounce a signal off a Russian military satellite, reestablish contact that way.</i>	<i>Возможно по истечении семи минут восстановить ее не удастся. Попытаемся передать отраженный военный сигнал через русский военный спутник, восстановить контакт.</i>

Источник: составлено А.А. Роговец

Source: compiled by A.A. Rogovets.

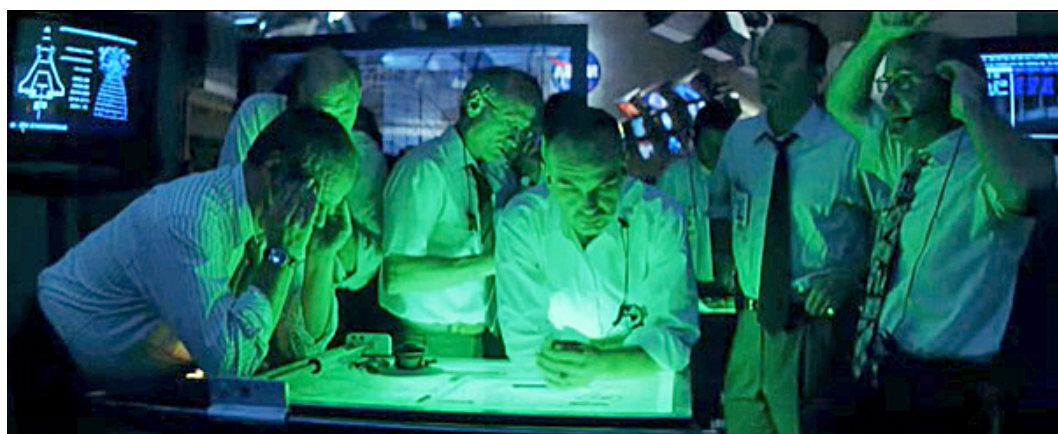


Рис. 8. Кадр из кинокартины «Armageddon», 1998

Источник: Кинопоиск. URL: <https://www.kinopoisk.ru/film/2941/> (дата обращения: 12.08.2025).

Fig. 8. A frame from the movie “Armageddon”, 1998

Source: Kinopoisk. Retrieved 12 August 2025, from <https://www.kinopoisk.ru/film/2941/>

Таблица 9 / Table 9

**Реплика в момент демонстрации кадра в фильме «Armageddon» /
A line from the film “Armageddon” as it was being shown**

Оригинал / Original	Звучащий перевод / A sounding translation
<i>We're bouncing off Russian and French satellites. It may have a faint signal.</i>	<i>Отраженный сигнал от русского и французского спутников.</i>

Источник: составлено А.А. Роговец.

Source: compiled by A.A. Rogovets.

В рамках той же сцены в кинокартине «Armageddon» кадр меняется, а ситуация получает развитие: сотрудникам космического агентства удастся поймать сигнал, хоть и слабый. В данном контексте цвет имеет аналогичное значение, спокойствие и безопасность, однако направлено это не на сравнение надежности спецслужб России и США, а на убеждение зрителя в том, что ситуация стабилизируется и сигнал будет окончательно восстановлен.

Заключение

Результаты проведенного исследования показали, что символическая роль цвета в репрезентации образа России в изученных американских фильмах актуализируется в передаче настроения или намерения определенного персонажа, ассоциируемого с Россией, маркировании тех или иных сюжетных поворотов, участниками которых являются русскоязычные персонажи. Наиболее частотным в данной функции оказался красный цвет. В 70 % случаев его использование оказывалось связано с государственной символикой, в частности с флагами СССР и России. В 30 % наблюдается использование красного цвета для передачи силы, власти и агрессии. Зеленый использовался в ассоциативной связи с надежностью и безопасностью, однако для усиления посредством сравнения образа ненадежности государственных служб России. Синий в равном числе сцен ассоциировался грустью и символизировал понимание и диалог. Кроме того, рассмотрение контекста сцен и учет звучащих реплик позволяет подтвердить гипотезу, выдвинутую на основании численных показателей. Действительно, в большинстве случаев цвет дополнял идею, выражаемую вербально или посредством контекста ситуации, а не передавал значение самостоятельно. Специфика кино как аудиовизуального произведения определяет возможность использования цвета в кадре для усиления сюжетного значения и применения его в качестве важного инструмента в формировании комплексного представления о репрезентируемых субъектах, в частности о России. Установлена невозможность получения объективных результатов анализа при изолированном изучении текстовых отрывков или визуальной составляющей кинокартин. Текст может содержать иронию, которую полноценно идентифицировать можно лишь при расширенном контексте. Визуальные же элементы поддаются интерпретации только в связи с идеями, переданными сюжетно и вербально, поскольку являются абстрактными, а следовательно, способны передавать спектр различных смыслов. Таким образом, для проведения комплексных исследований кино требуется сочетать изучение визуальной и вербальной сторон произведений.

Список литературы

- Александрова Е.В. Вербализация визуального в аудиовизуальных произведениях // Вестник Нижегородского университета им. Н.И. Лобачевского. 2023. № 1. С. 154–160 https://doi.org/10.52452/19931778_2023_1_154 EDN: SKISUU
- Алюнина Ю.М. Структурные компоненты образа России в американском кино и методика их выявления // Метаморфозы истории. 2024. № 34. С. 20. <https://doi.org/10.37490/S241436770032774-1> EDN: VPMATQ
- Алюнина Ю.М., Роговец А.А. Вербальная репрезентация образа России в американском кино: возможности и ограничения анализа // Славистика. 2024. Т. 28. № 2. С. 236–254. <https://doi.org/10.18485/slavistika.2024.28.2.15> EDN: FULUEW
- Алюнина Ю.М., Роговец А.А. Интермедиальное воплощение образа русского персонажа в американском фильме: портреты мужчины и женщины // Научный диалог. 2025. Т. 14. № 9. С. 162–188. <https://doi.org/10.24224/2227-1295-2025-14-9-162-188> EDN: IOGDPV
- Барило С.А., Христенко А.О., Рахимов Б.А. Роль каналов восприятия в коммуникативных процессах с клиентами // Столыпинский Вестник. 2023. Т. 5. № 1. С. 263–271. EDN: IDKXEA
- Бурдые П. Формы капитала / перевод М.С. Добряковой // Экономическая социология. 2002. Т. 3. № 5. С. 60–74. EDN: OYUVRD
- Дмитриев А.В., Коган М.С., Вдовина Е.К. Теоретико-прикладное значение корпусов в компьютерной лингводидактике // Litera. 2020. № 1. С. 200–216. EDN: JITBYW
- Ерохина Ю.В. Цветовое конструирование действительности в романе Достоевского «Преступление и наказание»: семантика бледного цвета // Философия. Журнал Высшей школы экономики. 2021. Т. 5. № 3. С. 140–158. <https://doi.org/10.17323/2587-8719-2021-3-140-158> EDN: HMFHMW
- Леонова О.Г. Джозеф най и «мягкая сила»: попытка нового прочтения // Социально-гуманитарные знания. 2018. № 1. С. 101–114. EDN: YPLPTX
- Мусаева А.М. Символика желтого цвета в современной русской поэзии // Litera. 2024. № 8. С. 168–176. <https://doi.org/10.25136/2409-8698.2024.8.71615> EDN: SVLMPO
- Пастуро М. Синий. История цвета. М. : Новое литературное обозрение, 2017. 154 с.
- Пастуро М. Красный. История цвета. М. : Новое литературное обозрение, 2019. 160 с.
- Пастуро М. Зеленый. История цвета. М. : Новое литературное обозрение, 2024. 146 с.
- Савченко П.А. Понятие социального капитала у П. Бурдые: трудности в интерпретации // Социология и право. 2020. № 2 (48). С. 62–68. <https://doi.org/10.35854/2219-6242-2020-2-62-68> EDN NPBWVG
- Bourdieu P. Ce que parler veut dire: L'économie des échanges linguistiques, Paris : Fayard, 1982. 239 p.
- Bourdieu P. The Forms of Capital // Handbook of Theory and Research for the Sociology of Education / ed. by J.G. Richardson. New York : Greenwood Press, 1986. P. 241–258.
- Nye J.S., Jr. Bound to Lead: The Changing Nature of American Power. New-York : Basic Books, 1991. 336 p.
- Nye J.S., Jr. Soft Power // Foreign Policy. 1990. № 80. P. 153–171. <https://doi.org/10.2307/1148580>

References

- Aleksandrova, E.V. (2023). Verbal description of visual elements in audiovisual content. *Vestnik of Lobachevsky University of Nizhny Novgorod*, (1), 154–160. (In Russ.). https://doi.org/10.52452/19931778_2023_1_154 EDN: SKISUU
- Alyunina, Yu.M. (2024). Structural components of Russian image in American cinema and methodology of their identification. *Metamorphoses of History*, (34), 20. (In Russ.). <https://doi.org/10.37490/S241436770032774-1> EDN: VPMATQ
- Alyunina, Yu.M., & Rogovets, A.A. (2024). Verbal representation of the image of Russia in American cinema: Possibilities and limitations of analysis. *Slavistics*, 28(2), 236–254. (In Russ.). <https://doi.org/10.18485/slavistika.2024.28.2.15> EDN: FULUEW
- Alyunina, Yu.M., & Rogovets, A.A. (2025). Intermedial representation of Russian character in American cinema: Portraits of men and women. *Nauchnyi dialog*, 14(9), 162–188. (In Russ.). <https://doi.org/10.24224/2227-1295-2025-14-9-162-188> EDN: IOGDPV
- Barilo, S.A., Khristenko, A.O., & Rakhimov, B.A. (2023). The role of perception channels in communicative processes with clients. *Stolypin Vestnik*, 5(1), 263–271. (In Russ.). EDN: IDKXEA
- Bourdieu, P. (1982). *Ce que parler veut dire: L'économie des échanges linguistiques*. Fayard.
- Bourdieu, P. (1986). The Forms of Capita. In J.G. Richardson (Ed.), *Handbook of Theory and Research for the Sociology of Education* (pp. 241–258). Greenwood Press.

- Bourdieu, P.; Dobryakova, M.S. (ed.). (2002). The forms of capital. *Journal of Economic Sociology*, 3(5), 60–74. (In Russ.). EDN: OYUVRD
- Dmitriev, A.V., Kogan, M.S., & Vdovina, E.K. (2020). Theoretical-applied significance of corpora in computer linguodidactics. *Litera*, 200–216. (In Russ.). EDN: JITBYW
- Leonova, O.G. (2018). Joseph Nye and “Soft Power”: An attempt to read it again. *Social and Humanitarian Knowledge*, (1), 101–114. (In Russ.). EDN: YPLPTX
- Musaeva, A.M. (2024). The symbolism of yellow in modern Russian poetry. *Litera*, (8), 168–176. (In Russ.). <https://doi.org/10.25136/2409-8698.2024.8.71615> EDN: SVLMPO
- Nye, J.S., Jr. (1991). *Bound to Lead: The Changing Nature of American Power*. Basic Books.
- Nye, J.S., Jr. (1990). Soft Power. *Foreign Policy*, 80, 153–171. <https://doi.org/10.2307/1148580>
- Pasturo, M. (2017). *Sinii. Istoriya tsveta* [Blue: The History of a Color]. Novoe literaturnoe obozrenie. (In Russ.).
- Pasturo, M. (2019). *Krasnyi. Istoriya tsveta* [Red: The History of a Color]. Novoe literaturnoe obozrenie. (In Russ.).
- Pasturo, M. (2024). *Zelenyi. Istoriya tsveta* [Green: The History of a Color]. Novoe literaturnoe obozrenie. (In Russ.).
- Savchenko, P.A. (2020). The concept of social capital in the works of P. Bourdieu: Interpretation Issues. *Sociology and Law*, 2(48), 62–68. (In Russ.). <https://doi.org/10.35854/2219-6242-2020-2-62-68> EDN: NPBWVG
- Yerokhina, Yu.V. (2021). Color construction of reality in the novel ‘Crime and Punishment’ by Dostoevsky: Semantics of pale colour. *Philosophy. Journal of the Higher School of Economics*, 5(3), 140–158. (In Russ.). <https://doi.org/10.17323/2587-8719-2021-3-140-158> EDN: HMFHMW

Информация об авторе:

РОГОВЕЦ Арина Анатольевна, лаборант-исследователь кафедры иностранных языков филологического факультета, Российский университет дружбы народов, Российская Федерация, 117198, Москва, ул. Миклухо-Маклая, д. 6. *Научные интересы:* имагология, социолингвистика, компьютерная лингвистика, корпусная лингвистика, когнитивная лингвистика.

E-mail: RinaRogovets@yandex.ru

ORCID: 0009-0005-8401-3590

SPIN-код: 4460-2906

Bionote:

Arina A. ROGOVETS, research assistant at the Department of Foreign Languages, Faculty of Philology, RUDN University, 6 Miklukho-Maklaya st., Moscow, 117198, Russian Federation. *Research interests:* imagology, sociolinguistics, computer linguistics, corpus linguistics, cognitive linguistics.

E-mail: RinaRogovets@yandex.ru

ORCID: 0009-0005-8401-3590

SPIN-code: 4460-2906



НАУЧНЫЕ СОБЫТИЯ. БРИФЫ. РЕЦЕНЗИИ SCIENTIFIC EVENTS. BRIEFS. BOOK REVIEWS

Научный семинар
«Динамика языков миноритарной ситуации»
6 марта 2024




Research Seminar
“Dynamics of Languages in Minority Situation”
March 6, 2024

DOI: 10.22363/2949-5997-2025-3-2-214-221


EDN HPQKFO

Тезисы доклада / Abstract

Геоинформационное картографирование в современных этнографических исследованиях

Инесса Евгеньевна СИДОРINA¹  ,
Арсений Алексеевич СЮЗЮМОВ^{1,2} 

¹Санкт-Петербургский государственный университет, Санкт-Петербург, Российская Федерация

²Музей антропологии и этнографии имени Петра Великого (Кунсткамера) Российской академии наук (МАЭ РАН), Санкт-Петербург, Российская Федерация
 innablank@mail.ru

Появление первых этнических или этнографических карт относят к XVII в. При этом размещение этносов отображали на картах еще древние греки, а в эпоху Великих географических

© Сидорина И.Е., Сюзюмов А.А., 2025



This work is licensed under a Creative Commons Attribution 4.0 International License
<https://creativecommons.org/licenses/by-nc/4.0/legalcode>

открытый отобразить размещение и расселения народов было очень важной новой информацией. С середины XIX в. начался процесс исчезновения элементов традиционной культуры в развитых промышленных странах — появилась необходимость в сохранении информации об этносах. Благодаря переписям населения этнографическое картографирование стало развиваться как самостоятельное направление, однако по-прежнему выполняло иллюстративную функцию: на картах с помощью ареалов (выделяемых цветом, штриховкой или подписями) и значков (чаще всего точек), отображалась информация, в основном, только о расселении народов.

С середины XX в., с развитием картографического метода исследований, картографирование стало особым методом и в этнографии. Используя приемы математической статистики, теории информации, аналитические методы картографии и т.д., стало возможным обнаруживать изменения различных объектов и явлений в пространстве и во времени, а также взаимосвязи между географическими характеристиками территории и закономерностями расселения народов.

В 1990-е гг. с бурным развитием геоинформационных технологий процессы хранения, обработки и анализа данных значительно упростились. Географические информационные системы (ГИС) стали главным инструментом для создания и анализа карт. Тем не менее, если в современной археологии методы геоинформатики давно стали необходимым и эффективным инструментом, в этнографических исследованиях ГИС — пока слабо реализованы (Сидорина и др., 2024).

В России этнографическим картографированием занимаются ученые Российского этнографического музея (РЭМ) и Музея антропологии и этнографии имени Петра Великого (Кунсткамера) РАН (МАЭ РАН)¹ (Головнёв, Куканов, Первалова, 2018). В сотрудничестве с преподавателями и студентами кафедры картографии и геоинформатики Санкт-Петербургского государственного университета стали эффективно применяться методы геоинформационного картографирования в этнографических исследованиях Ленинградской области и для работы с малыми коренными народами Российской Федерации и народами Севера (приложение, рис. 1). Традиции геоинформационного картирования продолжают в исследовании «Коренные малочисленные народы России: виртуальная этничность и сетевые опыты»², целью которого является изучить этническую идентичность коренных малочисленных народов в киберпространстве с применением новых методов полевой этнографии и аналитической этнологии, а также с учетом актуальной ситуации.

Основными задачами геоинформационного картографирования в этнографических исследованиях:

- поиск автоматизированных способов и инструментов ГИС, максимально подходящих для анализа и отображения информации;
- нахождение взаимосвязей между разными явлениями, характеризующими расселение тех или иных народов;
- выявление зависимости этнографических данных от рельефа, климата, экономических и исторических событий и др.

Помимо создания ГИС, электронных карт, появляются новые приемы, такие как киберэтнография: изучение онлайн-коммуникации внутри этнических сообществ и между ними (Белорусова, Сюзюмов, 2024). Геоинформатика и геоинформационное картографирование помогают в изучении влияния Интернета на современные этносы и в свою очередь активно используют различные онлайн-сервисы для решения технологических задач в обработке и визуализации данных (приложение, рис. 2).

¹ Фишман О.М., Засецкая М.Л., Исаченко Г.А. и др. Этноконфессиональный иллюстрированный атлас Ленинградской области. СПб.: Издательский дом «Инкери», 2017. EDN: ZTNXXL

² Проект № 23-78-10079 «Коренные малочисленные народы: виртуальная этничность и сетевые опыты» // РНФ. URL: <https://rscf.ru/project/23-78-10079/> (дата обращения: 21.04.2024).

Одним из популярных открытых программных комплексов является геоинформационная система QGIS³, с помощью которой можно создавать геоинформационные проекты, карты, анализировать данные исследований. С применением этой технологии была разработана геоинформационная база данных, позволяющая проследить, как коренные малочисленные народы России пользуются социальными сетями: какие темы поднимаются в социальных сетях, насколько широкий охват пользователей и потребителей контента, как социальные сети используются для сохранения и репрезентации культурного разнообразия коренных малочисленных народов России и др.

Ключевые слова: коренные малочисленные народы, народы России, народы Севера, этнографическое картографирование, киберэтнография, геоинформатика

Вклад авторов: Сидорина И.Е. — разработка концепции, формулировка и развитие ключевых целей и задач, а также проведение исследования, утверждение окончательного варианта результатов исследования; Сюзюмов А.А. — проведение исследования в геоинформационных системах, подготовка и редактирование текста тезисов для публикации.

Финансирование. Статья подготовлена в рамках проекта № 23–78–10079 «Коренные малочисленные народы: виртуальная этничность и сетевые опыты» по гранту РНФ.

Заявления о конфликте интересов. Авторы заявляют об отсутствии конфликта интересов.

История статьи: получена: 23 августа 2025; принята к публикации 20 ноября 2025.

Для цитирования: Сидорина И.Е., Сюзюмов А.А. Геоинформационное картографирование в современных этнографических исследованиях // *Macrosociolinguistics and Minority Languages*. 2025. Т. 3. № 2. С. 214–221. <https://doi.org/110.22363/2949-5997-2025-3-2-214-221> EDN: HPQKFO

GIS cartography in modern ethnography

Inessa E. SIDORINA¹  , Arsenii A. SIUZIUMOV^{1,2} 

¹Saint Petersburg State University, *Saint Petersburg, Russian Federation*

²Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) of the Russian Academy of Sciences (MAE RAS), *Saint Petersburg, Russian Federation*

 innablank@mail.ru

The first ethnic or ethnographic maps date back to the 17th century. The distribution of ethnic groups was already depicted on maps by the ancient Greeks, and during the Age of Discovery, mapping the distribution and settlement of peoples was a crucial new information. Beginning in the mid-19th century, elements of traditional culture began to disappear in developed industrial countries, necessitating the preservation of information about ethnic groups. Thanks to population censuses, ethnographic mapping began to develop as an independent discipline, but it still served an illustrative function:

³ Энтин А., Самсонов Т., Карпачевский А. Основы геоинформатики: практикум в QGIS // Основы геоинформатики. URL: <https://aentin.github.io/qgis-course/> (дата обращения: 20.04.2024).

Уроки по геоинформатике и картографии // ВКонтакте. URL: <https://vk.com/kartenx> (дата обращения: 20.04.2024).
Уроки QGIS // ВКонтакте. URL: <https://vk.com/qgisuroki> (дата обращения: 20.04.2024).

maps, using areas (highlighted by color, shading, or captions) and symbols (usually dots), primarily displayed information about the distribution of peoples.

Since the mid-20th century, with the development of cartographic research, mapping has become a specialized method in ethnography. Using techniques from mathematical statistics, information theory, analytical cartographic methods, and so on, it has become possible to identify changes in various objects and phenomena in space and time, as well as the relationships between the geographic characteristics of a territory and patterns of population distribution.

In the 1990s, with the rapid development of geoinformation technologies, the processes of storing, processing, and analyzing data were significantly simplified. Geographic information systems (hereinafter referred to as GIS) became the primary tool for creating and analyzing maps. However, while geoinformatics methods have long been a necessary and effective tool in modern archaeology, GIS remains poorly implemented in ethnographic research (Sidorina, Siuziumov, Rakova, Andreeva, Artemeva, 2024).

In Russia, ethnographic mapping is being conducted by scientists from the Russian Museum of Ethnography (REM) and the Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) of the Russian Academy of Sciences (MAE RAS)⁴ (Golovnev, Kukanov, Perevalova, 2018; Fishman et al., 2017). In collaboration with faculty and students from the Department of Cartography and Geoinformatics at Saint Petersburg State University, geoinformation mapping methods have been effectively applied in ethnographic research in the Leningrad Oblast and in work with indigenous minorities of the Russian Federation and the Northern peoples (Appendix, Fig. 1). The traditions of geoinformation mapping are continued in the study “Indigenoussmall-numbered Peoples: Virtual Ethnicity and Network Experiences”,⁵ the purpose of which is to study the ethnic identity of indigenous peoples in cyberspace using new methods of field ethnography and analytical ethnology, as well as taking into account the current situation.

The main objectives of geoinformation mapping in ethnographic research are:

- finding automated GIS methods and tools that are best suited for analyzing and displaying information;
- identifying relationships between various phenomena characterizing the distribution of certain peoples;
- identifying the relationship between ethnographic data and topography, climate, economic and historical events, etc.

In addition to the creation of GIS and electronic maps, new techniques are emerging, such as cyberethnography: the study of online communication within and between ethnic communities (Belorussova, Siuziumov, 2024). Geoinformatics and geoinformation mapping are helping to study the impact of the internet on modern ethnic groups, and in turn, actively utilize various online services to solve technological problems in data processing and visualization (Appendix, Fig. 2).

One of the popular opensource software packages is the QGIS⁶ geographic information system, which can be used to create geographic information projects, maps, and analyze research data. Using this technology, a geoinformation database was developed that makes it possible to track how indigenous peoples of Russia use social networks: what topics are raised on social networks, how broad the reach

⁴ Fishman, O.M., Zasetzkaya, M.L., Isachenko, G.A., et al. (2017). *Ehtnokonfessional'nyi illyustrirovannyi atlas Leningradskoi oblasti [The Ethnoconfessional Illustrated Atlas of Leningradskaya Oblast]*. Publishing House “Inkeri”. (In Russ.). EDN: ZTNXXL

⁵ Russian Science Foundation. *Korennyye malochislennyye narody: virtual'naya ehtnichnost' i setevyye opyty [Project № 23–78–10079 “Indigenoussmall-numbered Peoples: Virtual Ethnicity and Network Experiences”]*. Retrieved 21 April 2024, from: <https://rscf.ru/project/23-78-10079/> (In Russ.).

⁶ Ehntin, A., Samsonov, T., & Karpachevskii, A. *Osnovy geoinformatiki: praktikum v QGIS [Basics of geoinformatics: workshop in QGIS]*. *Osnovy geoinformatiki [Basics of geoinformatics]*. Retrieved 21 April 2024, from: <https://aentin.github.io/qgis-course/> (In Russ.).

Vkontakte. *Uroki po geoinformatike i kartografii [Geoinformatics and cartography lessons]*. Retrieved 20 April 2024, from: <https://vk.com/kartenx> (In Russ.).

Vkontakte. *Uroki QGIS [QGIS Lessons]*. Retrieved 20 April 2024, from: <https://vk.com/qgisuroki> (In Russ.).

of users and content consumers is, how social networks are used to preserve and represent the cultural diversity of indigenous peoples of Russia, etc.

Key words: indigenous peoples, peoples of Russia, Northern peoples, ethnographic mapping, cyberethnography, geoinformatics

Authors' contribution: Sidorina I.E. — developing the vision, defining and establishing key goals and objectives, and conducting the study, approving the final version of the research results; Siuziumov A.A. — conducting the research in GIS, preparing and editing the abstract for publishing.

Financing. The article was prepared within the project № 23–78–10079 “Indigenoussmall-numbered Peoples: Virtual Ethnicity and Network Experiences” which has been supported by Russian Science Foundation.

Conflict of interest. The authors declare no conflicts of interest.

Article history: received 23 August 2025; accepted 20 November 2025.

For citing: Sidorina, I.E., & Siuziumov, A.A. (2025). GIS cartography in modern ethnography. *Macrosociolinguistics and Minority Languages*, 3(2), 214–221. <https://doi.org/110.22363/2949-5997-2025-3-2-214-221> EDN: HPQKFO

Список литературы

- Белорусова С.Ю., Сюзюмов А.А. Картография в киберэтнографии (на примере коренных малочисленных народов России) // Сибирские исторические исследования. 2024. № 2. С. 93–119. <https://doi.org/10.17223/2312461X/44/5> EDN: NBQNJO
- Головнёв А.В., Куканов Д.А., Перевалова Е.В. Арктика : атлас кочевых технологий : монография. СПб. : МАЭ РАН, 2018. 352 с. EDN: NZWRUF
- Сидорина И.Е., Сюзюмов А.А., Ракова А.И., Андреева Т.А., Артемьева О.В. Геоинформационное картографирование в современных этнографических, конфессиональных и этноконфессиональных исследованиях // Кунсткамера. 2024. № 2 (24). С. 61–79. [https://doi.org/10.31250/2618-8619-2024-2\(24\)-61-79](https://doi.org/10.31250/2618-8619-2024-2(24)-61-79) EDN: RFNNDP

References

- Belorussova, S. Yu., & Siuziumov, A.A. (2024). Cartography in cyberethnography (using the example of the indigenous peoples of Russia). *Siberian Historical Research*, (2), 93–119. (In Russ.). <https://doi.org/10.17223/2312461X/44/5> EDN: NBQNJO
- Golovnev, A.V., Kukanov, D.A., & Perevalova, E.V. (2018). *Arktika: atlas kochevykh tekhnologii [Arctic: Atlas of Nomadic Technologies]*. MAE RAS Publ. (In Russ.). EDN: NZWRUF
- Sidorina, I.E., Siuziumov, A.A., Rakova, A.I., Andreeva, T.A., & Artemeva, O.V. (2024). GIS-based mapping in modern ethnographic, confessional, and ethno-confessional research. *Kunstkamera*, (2), 61–79. (In Russ.). [https://doi.org/10.31250/2618-8619-2024-2\(24\)-61-79](https://doi.org/10.31250/2618-8619-2024-2(24)-61-79) EDN: RFNNDP

Сведения об авторах:

СИДОРИНА Инесса Евгеньевна, кандидат географических наук, доцент кафедры картографии и геоинформатики, Санкт-Петербургский государственный университет, Российская Федерация, 199034, Санкт-Петербург, Университетская набережная, д. 7–9. *Научные интересы:* геоин-

формационное картографирование, геоинформационное и картографическое сопровождение этнографических, археологических, гидрологических исследований, изучение восприятия геоизображений.

E-mail: innablank@mail.ru

ORCID: 0000-0002-6373-0639

SPIN-код: 5118-3160

СЮЗИУМОВ Арсений Алексеевич, аспирант 3 курса кафедры картографии и геоинформатики, Санкт-Петербургский государственный университет, Российская Федерация, 199034, Санкт-Петербург, Университетская набережная, д. 7–9; младший научный сотрудник лаборатории музейных технологий, Музей антропологии и этнографии имени Петра Великого (Кунсткамера) Российской академии наук, Российская Федерация, 199034, Санкт-Петербург, Университетская набережная, д. 3. *Научные интересы:* тематическая картография, картография, теория картографии, этнографическая картография, этническая картография, цифровые методы в гуманитарных исследованиях, изучение восприятия геоизображений.

E-mail: SiuziumovAA@kunstkamera.ru

ORCID: 0000-0002-8178-8439

SPIN-код: 8414-1933

Bionotes

Inessa E. SIDORINA, Candidate of Geographical Sciences, associate professor, Department of Cartography and Geoinformatics, Saint Petersburg State University, 7–9 Universitetskaya Embankment, Saint Petersburg, 199034, Russian Federation. *Research interests:* geoinformation cartography, geoinformation and cartographic assistance in ethnographic, archeological, hydrological researches, geoimage perception.

E-mail: innablank@mail.ru

ORCID: 0000-0002-6373-0639

SPIN-code: 5118-3160

Arsenii A. SIUZUMOV, 3d year Ph.D. student, Department Cartography and Geoinformatics, Saint Petersburg State University, 7–9 Universitetskaya Embankment, Saint Petersburg, 199034, Russian Federation; unior research fellow, Laboratory of Museum Technologies, Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) of the Russian Academy of Sciences, 3 Universitetskaya Embankment, Saint Petersburg, 199034, Russian Federation. *Research interests:* thematic cartography, cartography, theory of cartography, ethnographic cartography, ethnic cartography, digital humanities, geoimage perception.

E-mail: SiuziumovAA@kunstkamera.ru

ORCID: 0000-0002-8178-8439

SPIN-code: 8414-1933

Приложение / Appendix

Распределение народов в сравнении с ландшафтными характеристиками

Рабочее окно ГИС-проекта

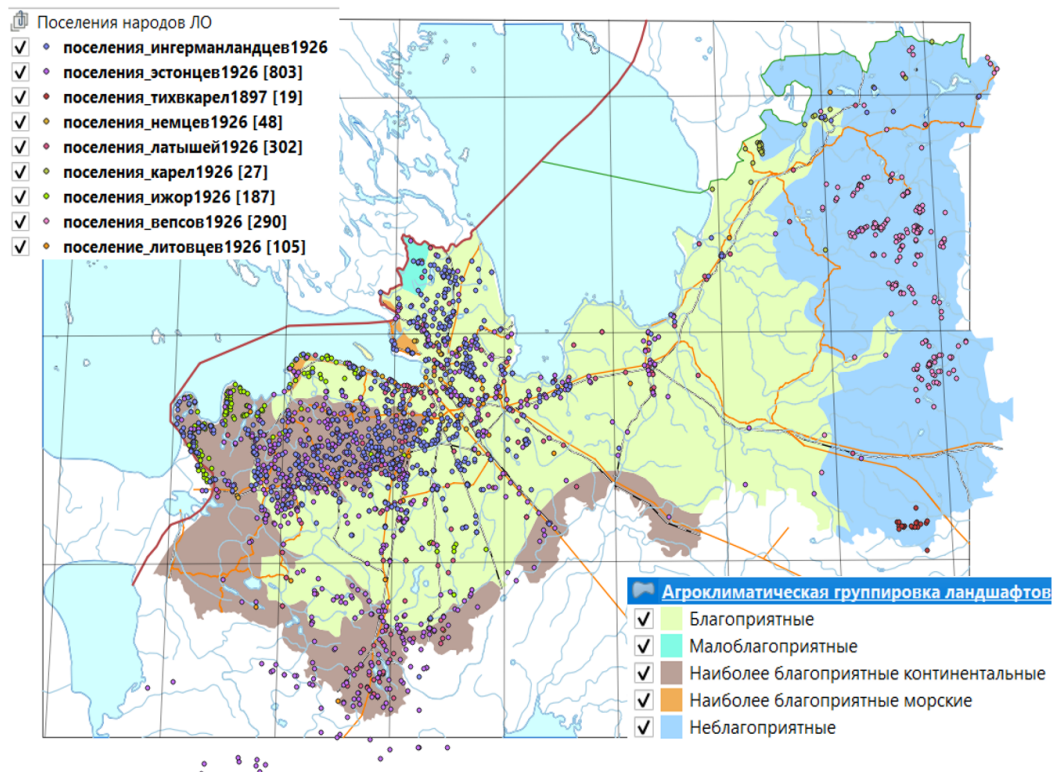


Рис. 1. Распределение коренных малочисленных народов России в Санкт-Петербурге и Ленинградской области

Источник: создано И.Е. Сидориной, А.А. Сюзюмовым с помощью QGIS.

Fig. 1. Distribution of indigenous peoples of Russia in Saint Petersburg and the Leningrad Oblast'

Source: made by I.E. Sidorina, A.A. Siuziumov using QGIS.

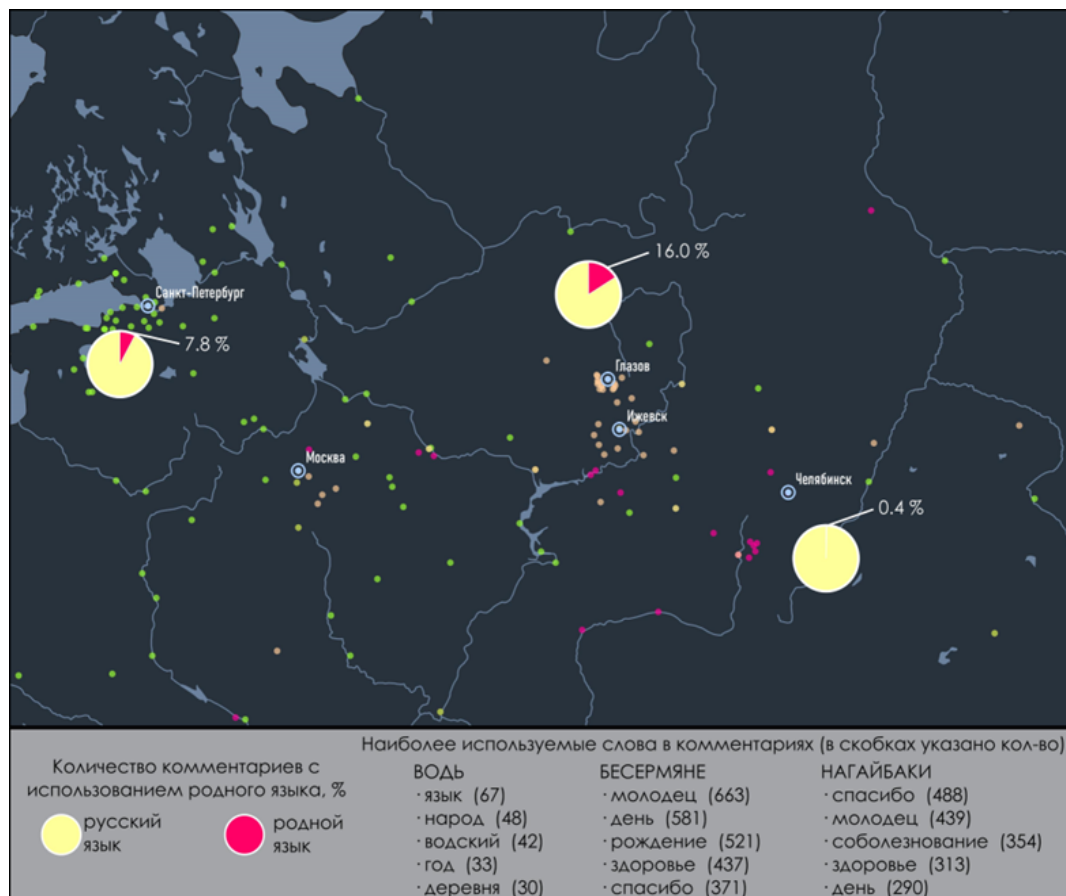


Рис. 2. Распределение использования родного языка и русского среди коренных малочисленных народов России в социальных сетях

Источник: создано И.Е. Сидориной, А.А. Сюзюмовым с помощью QGIS.

Fig. 2. Distribution of the use of native language and Russian among indigenous peoples of Russia in social networks

Source: made by I.E. Sidorina, A.A. Siuziumov using QGIS.