



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/110.22363/2949-5997-2025-3-2-117-130> 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 вепского 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 вепского 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 ▾ 2025 ▾
Фольклорный жанр:	Былички и бывальщины / Демонологические рассказы ▾
Сюжет/Мотив:	Водяной боится грома ▾
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 (аренальная, 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: IPVMVI
- 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: YBZRHV

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: YBZRHV
- 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: IPVMVI
- 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 archetypal 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