MACROSOCIOLINGUISTICS AND MINORITY LANGUAGES 2025;3(2):131–145
ISSN 2949-5997 (Online) https://macrosociolingusictics.ru/MML/

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 ПОЖИДАЕВ[1] ✉, **Елена Сергеевна ТЕПЛЫХ**[1], **Сергей Ильич ДАНИЛОВ**[2]

[1]Национальный исследовательский Томский государственный университет, *Томск, Российская Федерация*
[2]Российский университет дружбы народов, *Москва, Российская Федерация*
✉ msp@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

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: HHYPZI

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

Mikhail S. POZHIDAEV[1] ✉, **Elena S. TEPLYKH**[1], **Sergey I. DANILOV**[2]

[1]National Research Tomsk State University, *Tomsk, Russian Federation*
[2]RUDN 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 stady 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/110.22363/2949-5997-2025-3-2-131-145 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; Arulprakashet 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'l' dölod» (рус. «Четыре ветра»)[8] — сказки для детей на вепском языке;
- аудиокниги на чувашском языке, среди которых притча «Вĕçекен кӳлĕ» (рус. Летающее озеро»), рассказ «Кăтра хĕвелçаврăнăш» (рус. «Кудрявый подсолнушек»), аудиоиздание «Ачасем валли Тăван çĕршывăн аслă вăрçи çинчен 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 тĕтĕ 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 дәүләт 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 «Нойрhоо hэриһэн 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'l' 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 выпустят 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://mastera.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 Бурятии): URL: https://soyol.ru/tag/?q=Аудиобиблиотека (дата обращения: 10.09.2025).
[14] Проект «Родные голоса» РОО «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. № 314[15], 2024: 5; Распоряжение Правительства РФ от 12 июня 2024 г. № 1481-р[16], 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 текста[17], 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. Баҡсасы 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 корпусе[18] Национального корпуса русского языка; прагматические маркеры определенной реплики, как, например, в корпусе устной речи Один речевой день[19] (маркеры аппроксиматор, рефлексив, хезитатив, ксенопоказатель, метакомуникатив, 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 пэ*], а не *вто, мкс, дз, ввп*, 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 *ВВС* 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 стен**а**м*
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 раздельными 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 автомата[22], 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 языка[23], где собраны как специальные сокращения: «н. п.» — *нэрын падеж* (рус. *именительный падеж*), 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 распознания 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 q0 попадает в конечное из множества F» (Лобарёв, Лобарёв, 2025: 10).

[23] Толковый словарь бурятского языка. URL: https://edbl.ru/pomety-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 *доро'ги* 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 смыслоразличительную 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 *ааангли*.

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» (Миниі 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 предоставление 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

Дрожащих 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 угроведения. 2025. Т. 15. № 1 (60). С. 80–89. https://doi.org/10.30624/2220-4156-2025-15-1-80-89 EDN: EMPAHK

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/IJARCCE.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, RJ, Stanton, D., Wu, Y., Weiss, R.J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Zh., Bengio, S., Le, Q., Agiomyrgiannakis, 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. [Sovremennyi 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/IJARCCE.2015.4812

Punegova, 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, RJ, Stanton, D., Wu, Y., Weiss, R.J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Zh., Bengio, S., Le, Q., Agiomyrgiannakis, 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