НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
май–июнь 2026 Том 26 № 3 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
May–June 2026 Vol. 26 No 3 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2026-26-3-475-485
УДК 004.855.5

Метод автоматического машинного перевода текстов с вербального языка в последовательность глосс

Арсений Михайлович Поляков[1], Дмитрий Александрович Рюмин[2]✉

[1,2] Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация

[1] arseney02@mail.ru, https://orcid.org/0000-0002-8681-988X
[2] ryumin.d@iias.spb.su✉, https://orcid.org/0000-0002-7935-0569

Аннотация

Введение. Рассмотрена задача машинного перевода с вербального языка на русский жестовый язык в виде промежуточного текстового представления — последовательности глосс. Разработан метод подготовки данных и автоматического перевода из вербального текста в последовательность глосс жестового языка. **Метод.** В предложенном методе формируется размеченный параллельный корпус «вербальный текст-последовательность глосс» на основе ручной разметки. Словарь глосс задается по примерам из корпуса жестового языка и используется в качестве ограничения допустимых выходных токенов. Для перевода сравниваются два класса моделей: трансформеры архитектуры «энкодер-декодер», адаптируемые к целевой задаче на параллельном корпусе; большие языковые модели архитектуры «только декодер», применяемые в режиме контекстного обучения по нескольким примерам с промптом, содержащим инструкцию, словарь глосс и ограничения на формат ответа. Качество перевода оценивается по метрике Bilingual Evaluation Understudy на тестовой выборке параллельного корпуса. **Основные результаты.** Экспериментальные результаты показали, что трансформерные модели обеспечивают более высокое качество машинного перевода по сравнению с большими языковыми моделями; наилучший результат среди трансформеров достигается моделью mT5-small (0,84). Среди больших языковых моделей максимальное значение 0,60 получено для GPT-5.2. **Обсуждение.** Предложенный метод может быть применен как часть системы по воспроизведению цифровой двусторонней коммуникации между носителями жестовых языков и вербальных. Метод позволяет переводить текст на вербальном языке в последовательность глосс, которые в дальнейшем могут быть синтезированы с помощью цифровых аватаров для понимания носителями жестовых языков информации, сказанной или написанной носителями вербальных языков. Исходные материалы, параллельный корпус и инструкции по воспроизведению экспериментов размещены в открытом репозитории, посвященному методу автоматического машинного перевода текстов с вербального языка в последовательность глосс.

Ключевые слова

вербальный текст, словарь глосс, параллельный корпус, русский жестовый язык, трансформеры, большие языковые модели, машинный перевод

Благодарности

Исследование выполнено при финансовой поддержке Российского научного фонда, проект № 24-71-00083.

Ссылка для цитирования: Поляков А.М., Рюмин Д.А. Метод автоматического машинного перевода текстов с вербального языка в последовательность глосс // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 3. С. 475–485. doi: 10.17586/2226-1494-2026-26-3-475-485

© Поляков А.М., Рюмин Д.А., 2026

Automatic machine translation from spoken-language text to sign language gloss sequences

Arseniy M. Polyakov[1], Dmitry A. Ryumin[2]✉

[1,2] St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, 199178, Russian Federation

[1] arseney02@mail.ru, https://orcid.org/0000-0002-8681-988X
[2] ryumin.d@iias.spb.su✉, https://orcid.org/0000-0002-7935-0569

Abstract

This article addresses the task of machine translation from a spoken-language to Russian Sign Language in the form of an intermediate textual representation — a gloss sequence. The goal of this work is to develop a data preparation method and an automatic translation method for mapping verbal text to a sign-language gloss sequence. A manually annotated parallel corpus of "spoken-language text–gloss sequence" pairs is constructed. A gloss vocabulary is defined based on examples from a sign-language corpus and is used to constrain the set of admissible output tokens. Two model classes are compared: Transformers with an encoder-decoder architecture, adapted to the target task on the parallel corpus; and Large Language Models with a decoder-only architecture applied via In-Context Learning with a few examples and a prompt that includes instructions, the gloss vocabulary, and output-format constraints. Translation quality is evaluated using the BLEU metric on the test split of the parallel corpus. Experimental results show that Transformer-based models provide higher machine translation quality than Large Language Models; the best Transformer result is achieved by mT5-small (0.84). Among Large Language Models, the best value of 0.60 is obtained for GPT-5.2. The proposed method can be applied as part of a system for enabling digital bidirectional communication between sign-language users and spoken-language users. The method translates spoken-language text into a gloss sequence which can subsequently be synthesized using digital avatars to allow sign-language users to understand information that is spoken or written by spoken-language users. Source materials, the parallel corpus, and instructions for reproducing the experiments are available in the public repository dedicated to the method of automatic machine translation of texts from verbal language into a sequence of glosses.

Keywords

spoken-language text, gloss vocabulary, parallel corpus, Russian Sign Language, Transformers, Large Language Models, machine translation

Acknowledgements

This research is financially supported by the Russian Science Foundation, project No. 24-71-00083.

For citation: Polyakov A.M., Ryumin D.A. Automatic machine translation from spoken-language text to sign language gloss sequences. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 3, pp. 475–485 (in Russian). doi: 10.17586/2226-1494-2026-26-3-475-485

Введение

В последние годы в Российской Федерации и за рубежом заметно расширяется эмпирическая база исследований жестовых языков и смежных задач компьютерного зрения. Формируются крупные визуальные корпуса, используемые для распознавания жестов и разработки интеллектуальных систем машинного перевода. Для русского жестового языка доступны корпуса Bukva[1], Slovo [1], TheRuSLan [2, 3], Logos[2], а также специализированный корпус для поликлинического применения [4]. Они используются для распознавания дактиля, отдельных жестов и жестовых высказываний по визуальным данным, а также для разметки и аннотирования жестовой речи. Среди зарубежных ресурсов широко применяются WLASL [5], AUTSL [6] и CSL[3], которые стали стандартными корпусами для сопоставимых экспериментов в области обработки жестовых высказываний. Наряду с ними существуют наборы данных жестов, не являющиеся корпусами жестовых языков в лингвистическом смысле. Например, HaGRID [10] содержит статические управляющие жесты для человеко-машинного взаимодействия и ориентирован на задачи классификации, а не на машинный перевод связных высказываний. Однако даже при росте числа визуальных ресурсов для обучения и сопоставимой оценки моделей машинного перевода требуются параллельные пары «вербальный текст-последовательность глосс». Под глоссами далее понимаются нормализованные текстовые обозначения жестов, используемые как промежуточное представление жестового высказывания. Отсутствие в открытом доступе стандартизованных параллельных пар «вербальный текст-последовательность глосс», особенно для русского жестового языка, ограничивает воспроизводимость и усложняет корректное сравнение нейросетевых моделей. Следует подчеркнуть, что аннотации вида «видеофрагмент – метка класса» для изолированных жестов не заменяют параллельные данные «вербальный текст-последовательность глосс», поскольку не отражают грамматическую структуру высказывания и не позволяют обучать нейросетевые модели последовательного машинного перевода на уровне предложения.

Настоящая работа посвящена задаче автоматического машинного перевода текстов на вербальном языке в

[1] Bukva: Russian Sign Language Alphabet Dataset. 2024 [Электронный ресурс]. URL: https://github.com/ai-forever/bukva (дата обращения: 20.02.2026).

[2] Logos as a Well-Tempered Pre-Train for Sign Language Recognition. 2025 [Электронный ресурс]. URL: https://github.com/ai-forever/Logos (дата обращения: 20.02.2026).

[3] Chinese Sign Language (CSL) Dataset. 2024 [Электронный ресурс]. URL: https://github.com/woshisad159/TFNet (дата обращения: 20.02.2026).

последовательность глосс. Предложен метод подготовки и согласования данных на основе существующих визуальных корпусов, позволяющий сформировать параллельный корпус «вербальный текст-последовательность глосс» и зафиксировать единые правила аннотирования. На сформированном корпусе проводится экспериментальное сравнение современных методов генерации глосс, включая модели на основе трансформеров и больших языковых моделей (БЯМ), при единых настройках обучения и оценки качества. Полученные результаты позволяют сформулировать практические рекомендации по выбору и настройке нейросетевых моделей для перевода «вербальный текст-последовательность глосс» и создают основу для дальнейшей разработки интеллектуальных систем двустороннего машинного перевода между жестовыми и вербальными языками.

Обзор литературы

Современные методы машинного перевода для жестовых языков в основном основаны на последовательных нейросетевых моделях архитектуры трансформеров. Дополнительно развиваются методы, использующие БЯМ в качестве генератора выходной последовательности или как компонент декодирования вербального текста в последовательность глосс и обратно. В работе [8] проведено систематическое сравнение трансформерных архитектур и стратегий машинного обучения для постановки задачи «вербальный текст-последовательность глосс». Показано, что качество определяется конфигурацией энкодера и декодера, выбором методов оптимизации, регуляризации и декодирования. Следовательно, решающее значение имеют единые стратегии машинного обучения и правила оценки, а не только объем данных. В задачах перевода жестовой речи заметное развитие получили методы предварительного обучения с последующим дообучением [9, 10], при которых сначала формируется переносимое представление на крупных объемах неразмеченных или слабо размеченных визуальных данных, а затем выполняется адаптация на параллельных корпусах. Такие методы особенно эффективны при недостатке размеченных данных и сохраняют применимость при ограниченном объеме разметки [9, 10], включая задачу перевода вербального текста в последовательность глосс [10]. Для уменьшения зависимости от промежуточного аннотирования, требующего значительных трудозатрат, применяются методы машинного обучения с использованием векторных представлений предложений [11–13], в которых промежуточная языковая информация задается не последовательностью глосс, а обучаемыми векторными представлениями предложений. Это повышает устойчивость нейросетевой модели при несогласованной разметке и улучшает качество при применении на корпусах с иным методом разметки [11, 12].

Развиваются методы с использованием БЯМ, основанные на предварительном преобразовании входных данных в дискретное представление, пригодное для языкового декодирования. В работе [14] предложен метод SignLLM, в котором последовательность кадров преобразуется в дискретное представление, пригодное для языкового декодирования, после чего генерация выполняется БЯМ, что позволяет использовать ее способность учитывать контекстные и семантические зависимости без жесткой привязки к глоссам; в работе [15] представлен метод SpaMo, где в БЯМ подаются признаки пространственной конфигурации и движения рук, что повышает информативность входного представления и снижает зависимость от отдельного дообучения визуального кодировщика. Для сохранения преимуществ промежуточного представления при недостатке разметки глосс предложен метод PGG-SLT [16], в котором последовательность глосс формируется с использованием БЯМ, а затем выполняется уточнение порядка с применением полуконтролируемого машинного обучения для согласования с жестовой последовательностью. В работах по прикладным интеллектуальным системам двустороннего машинного перевода [17–19] показано, что при объединении модулей распознавания и синтеза жестовой речи необходимы единое промежуточное текстовое представление и унифицированные правила формирования выходных последовательностей, иначе нарушается согласование между компонентами системы и снижается качество перевода.

К ключевым ограничениям современных методов относятся:

— различия в правилах формирования и нормализации последовательности глосс, а также в правилах оценки, что затрудняет сопоставление результатов и воспроизводимость экспериментов даже при использовании однотипных нейросетевых архитектур;
— недостаток крупных согласованно размеченных параллельных корпусов «вербальный текст-последовательность глосс», из-за чего качество существенно зависит от стратегии машинного обучения, а при применении на других корпусах с иным методом разметки наблюдается заметное снижение качества;
— сохраняющаяся зависимость от промежуточного аннотирования, требующего значительных трудозатрат, а также необходимость специальных приемов для снижения влияния неполноты и несогласованности разметки. Дополнительно отмечается, что при использовании БЯМ необходимо контролировать формирование выходных последовательностей и согласовывать промежуточное представление с заданным форматом глосс, иначе возрастает вариативность вывода и снижается сопоставимость результатов.

Основная трудность состоит в построении параллельных корпусов и единых правил проведения экспериментов, позволяющих корректно сравнивать модели при фиксированном формате глосс. Настоящая работа направлена на преодоление этого ограничения за счет согласованной подготовки параллельного корпуса «вербальный текст-последовательность глосс» и проведения сопоставимого экспериментального анализа трансформерных архитектур и БЯМ при единых правилах обучения и оценки.

Метод

Функциональная схема предлагаемого метода приведена на рис. 1. Метод включает следующие этапы.

Этап 1. Формирование размеченного параллельного корпуса «вербальный текст-последовательность глосс» на основе корпуса жестового языка и ручной разметки.

Этап 2. Подготовка входных представлений, включающая токенизацию, векторизацию, добавление позиционного кодирования и, при необходимости, нормализацию представлений.

Этап 3. Выполнение машинного перевода с использованием моделей двух типов, трансформера архитектуры «энкодер-декодер» и БЯМ с архитектурой «только декодер».

Этап 4. Декодирование выходных оценок моделей, преобразование их в вероятности и выбор следующего токена с формированием выходной последовательности глосс.

Этап 5. Оценивание качества перевода на тестовой выборке с использованием выбранной метрики.

На этапе 1 выполняется формирование размеченного параллельного корпуса выполняется на основе корпуса жестового языка и набора предложений на вербальном языке. Корпус жестового языка используется не для непосредственного обучения нейросетевых моделей машинного перевода, а как источник для формирования словаря допустимых глосс.

Пусть $\mathcal{C}$ — корпус жестового языка, содержащий видеофрагменты и соответствующие аннотации. На основе $\mathcal{C}$ формируется словарь глосс согласно формуле:

$$V_g = g_1, g_2, \ldots, g_{|Vg|},$$

где $|V_g|$ — число различных глосс в словаре V_g. Таким образом, множество V_g задает все допустимые значения выходных токенов при формировании последовательностей глосс.

Использование фиксированного словаря V_g обеспечивает единообразие ручной разметки и позволяет формально описывать перевод как отображение из множества вербальных предложений в пространство конечных последовательностей.

Обозначим через $S = \{s_i\}_{i=1}^{M}$ множество предложений на вербальном языке, используемых для разметки, где s_i — i-е предложение, а M — число предложений. Ручная разметка заключается в сопоставлении каждому предложению s_i размеченной последовательности глосс y_i, составленной из элементов словаря V_g. Разметка задается в виде:

$$A: S \to V_g^*, \quad A(s_i) = y_i,$$

где V_g^* — множество всех конечных последовательностей глосс, составленных из элементов словаря V_g. Размеченная последовательность глосс имеет вид:

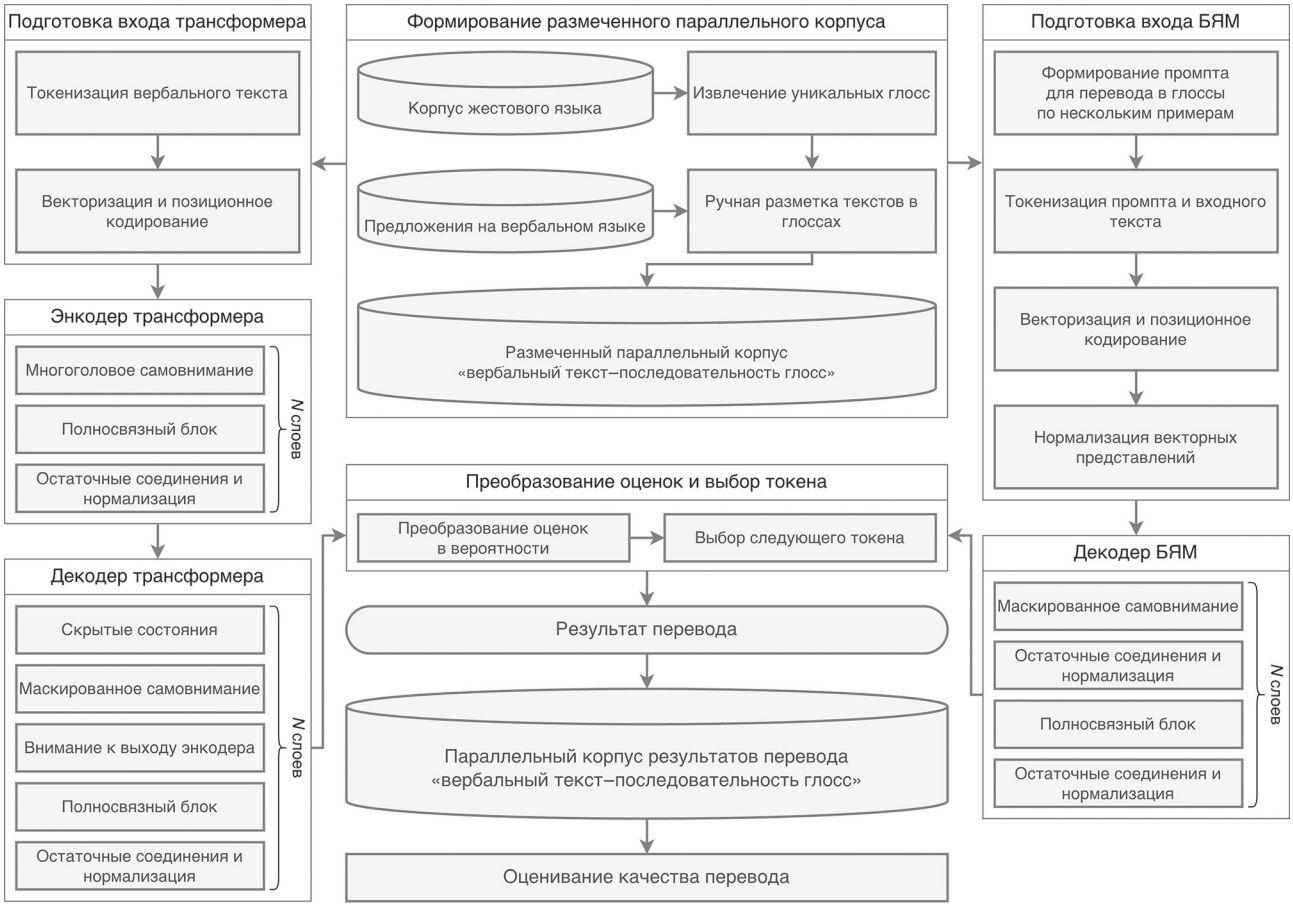

Рис. 1. Функциональная схема предлагаемого метода

Fig. 1. Pipeline of the proposed method

$$y_i = \left(g_1^{(i)}, g_2^{(i)}, \ldots, g_{T_i}^{(i)}\right), \quad g_t^{(i)} \in V_g,$$

где T_i — длина последовательности глосс для i-го предложения. В результате формируется размеченный параллельный корпус:

$$D = \{(s_i, y_i)\}_{i=1}^{M},$$

представляющий собой набор пар вида «вербальный текст-последовательность глосс» и используемый на последующих этапах обучения и оценки моделей перевода.

На этапах 2–4 метода выполняется подготовка входных данных для нейросетевых моделей отдельно для трансформера и для БЯМ.

Подготовка входа трансформера начинается с токенизации вербального текста. Для каждого предложения s_i формируется токенизированная последовательность, как показано в формуле вида:

$$X_i = \left(x_1^{(i)}, x_2^{(i)}, \ldots, x_{n_i}^{(i)}\right),$$

где $x_k^{(i)}$ — k-й токен предложения s_i; n_i — длина X_i. Далее выполняются векторизация и позиционное кодирование, в результате чего строятся входные векторные представления токенов:

$$E_i = \left(e_1^{(i)}, e_2^{(i)}, \ldots, e_{n_i}^{(i)}\right), \quad e_k^{(i)} = \mathrm{Emb}\left(x_k^{(i)}\right) + \mathrm{Pos}(k),$$

где $\mathrm{Emb}(\cdot)$ — отображение токена в пространство эмбеддингов; $\mathrm{Pos}(k)$ — позиционное кодирование для позиции k. Последовательность E_i подается на вход энкодера трансформера.

«Энкодер-декодер» трансформера используется в стандартной постановке задачи. Входная последовательность токенов E_i кодируется энкодером, а декодер авторегрессионно формирует выходную последовательность глосс.

Подготовка входа БЯМ начинается с формирования промпта для перевода в глоссы по нескольким примерам. Для каждого предложения s_i формируется промпт p_i, включающий инструкцию к задаче перевода, описание допустимого словаря глосс V_g и несколько примеров корректного перевода. Объединенный вход задается конкатенацией:

$$u_i = \mathrm{concat}(p_i, s_i), \quad Z_i = \left(z_1^{(i)}, z_2^{(i)}, \ldots, z_{m_i}^{(i)}\right),$$

где $\mathrm{concat}(\cdot)$ — операция конкатенации; Z_i — токенизированная последовательность объединенного входа; m_i — длина последовательности. Выполняются векторизация и позиционное кодирование, в результате чего формируются векторные представления вида:

$$H_i = \left(h_1^{(i)}, h_2^{(i)}, \ldots, h_{m_i}^{(i)}\right), \quad h_k^{(i)} = \mathrm{Emb}\left(z_k^{(i)}\right) + \mathrm{Pos}(k).$$

При необходимости применяется нормализация компонент векторных представлений токенов:

$$\tilde{H}_i = \mathrm{Norm}(H_i).$$

Полученная последовательность $\tilde{H}_i$ используется как вход БЯМ. Декодер БЯМ состоит из N последовательно соединенных слоев и реализует авторегрессионное формирование выходной последовательности. На вход подается последовательность нормализованных представлений $\tilde{H}_i$, полученная на этапе подготовки входа БЯМ. Вход первого слоя задается по формуле:

$$R_1 = \tilde{H}_i.$$

Каждый слой декодера БЯМ содержит маскированное самовнимание и полносвязный блок, причем после каждого из них применяются остаточные соединения и нормализация.

Преобразование оценок и выбор токена выполняются одинаково для трансформера и для БЯМ. На шаге t декодер формирует вектор выходных оценок по словарю глосс V_g, который интерпретируется как логиты:

$$o_t \in \mathbb{R}^{|V_g|}.$$

Преобразование логитов в вероятности выполняется с помощью функции $\mathrm{softmax}(\cdot)$. Для $g \in V_g$ преобразование логитов имеет вид:

$$p_t(g) = \mathrm{soft\,max}\,(o_t)(g) = \frac{\exp(o_t(g))}{\sum\limits_{g' \in V_g} \exp(o_t(g'))}. \qquad (1)$$

где $p_t(g)$ — вероятность выбора глоссы g на шаге t. Далее выбирается следующий токен по правилу максимума вероятности:

$$\hat{g}_t = \arg\max_{g \in V_g} p_t(g). \qquad (2)$$

Последовательное применение формул (1) и (2) формирует результат перевода для входного предложения s_i в виде предсказанной последовательности глосс:

$$\hat{y}_i = \left(\hat{g}_1^{(i)}, \hat{g}_2^{(i)}, \ldots, \hat{g}_{\hat{T}_i}^{(i)}\right), \quad \hat{g}_t^{(i)} \in V_g,$$

где $\hat{T}_i$ — длина предсказанной последовательности. Генерация завершается при предсказании специального токена конца последовательности либо при достижении заданной максимальной длины $T_{\max}$. Для набора входных предложений формируется параллельный корпус результатов перевода:

$$\hat{D} = \{(s_i, \hat{y}_i)\}_{i=1}^{M}.$$

Параллельный корпус $\hat{D}$ представляет собой набор пар вида «вербальный текст-последовательность глосс», полученных в результате работы модели.

На этапе 5 происходит оценивание качества перевода, которое выполняется путем сравнения предсказанных последовательностей $\hat{y}_i$ с размеченными последовательностями y_i на тестовой выборке. Пусть $D_{\text{test}} \subset \mathrm{D}$ — тестовая выборка параллельного корпуса. Тогда качество перевода определяется значением выбранной метрики $M(\cdot,\cdot)$, вычисляемой по формуле:

$$\mathrm{Score} = \frac{1}{|D_{\text{test}}|} \sum_{(s_i, y_i) \in D_{\text{test}}} M(y_i, \hat{y}_i).$$

В качестве $M(\cdot,\cdot)$ может использоваться любая метрика, определенная на парах дискретных последо-

вательностей, что обеспечивает возможность замены критерия оценки без изменения описания этапов построения корпуса и декодирования.

Экспериментальные исследования метода

Экспериментальные исследования выполнены для количественной оценки качества машинного перевода в задаче «вербальный текст-последовательность глосс» на параллельном корпусе, в котором разметка последовательностей глосс произведена вручную. Качество перевода оценено сопоставлением предсказанных последовательностей глосс с разметкой для каждого предложения. В настоящем исследовании проведено сравнение двух классов моделей. Первый класс включали трансформерные модели машинного перевода архитектуры «энкодер-декодер», предобученные на мультиязычных параллельных корпусах и далее адаптируемые в постановке задачи «вербальный текст-последовательность глосс». Второй класс состоит из БЯМ архитектуры «только декодер», применяемые в режиме контекстного обучения по нескольким примерам.

Для БЯМ использован способ решения задачи через формирование промпта по нескольким примерам, содержащего инструкцию к переводу, ограничение на формат ответа и демонстрационные примеры. Было выбрано 5 примеров, включающих в себя, как и короткие (три слова), так и длинные (5 слов) предложения на различные тематики, не совпадающие с примерами в обучающей и тестовой выборках. Данный выбор обусловлен потенциальным увеличением точности машинного перевода БЯМ благодаря тематической и лексико-грамматической вариативности и уникальности примеров промпта, добавляющих больший контекст на вход декодера. Варианты промптов и оптимальный из них приведены на рис. 2.

Для обеспечения сопоставимости условий параметры генерации использованы в значениях по умолчанию для каждой модели. Для трансформерных моделей рассмотрены архитектуры, широко применяемые в мультиязычном машинном переводе и предобученные на параллельных корпусах, включая mBART[1] [20], NLLB[2], OPUS-MT[3] [21] и mT5[4] [22], после чего модели адаптированы к задаче «вербальный текст-последовательность глосс» на имеющемся корпусе. Вычислительные эксперименты проведены на графическом ускорителе NVIDIA RTX 4090 (24 ГБ).

Для количественной оценки качества машинного перевода использована метрика Bilingual Evaluation Understudy (BLEU) [23]. Применение BLEU в рассматриваемой задаче обосновано тем, что выход модели представляет собой последовательность дискретных символов из фиксированного словаря глосс, а также необходимостью строгого соблюдения формата ответа. Различия BLEU значений обусловлены настройками токенизации, сглаживания и агрегирования *n*-грамм в разных реализациях, однако качественные выводы при сравнении моделей сохраняются [24].

В качестве параллельного корпуса применен вручную размеченный набор коротких русскоязычных предложений, охватывающих различные тематики. Предложения сформированы с использованием БЯМ ГигаЧат (GigaChat) [25]. При генерации выполнен контроль длины (5–6 слов) и тематическая вариативность, что обеспечивает репрезентативность корпуса для оценки качества машинного перевода на коротких высказываниях. Для сохранения возможности последующей разметки в глоссах при формировании данных использован фиксированный промпт, в котором заданы уникальные глоссы корпуса русского жестового языка Slovo [1], сформированного ПАО Сбербанк, и зафиксированы ограничения на применение заданного словаря. Дополнительно в промпт включены требования к построению семантически близких формулировок, что повышает согласованность получаемых предложений с разметкой и снижает вероятность появления лексики, не покрываемой словарем глосс.

Для обоснования выбора используемого ресурса (корпус Slovo) и сопоставления его характеристик с другими корпусами жестового языка в табл. 1 приведен сравнительно-сопоставительный анализ, учитывающий язык, тип аннотаций, объем и доступность данных.

Результаты экспериментов для БЯМ представлены в табл. 2, а для трансформерных моделей — в табл. 3, включая значения BLEU разных порядков. Полученные значения BLEU показывают, что трансформерные модели обеспечивают более высокое качество машинного перевода по сравнению с БЯМ. Это согласуется с тем, что трансформеры в рассматриваемых экспериментах адаптируются к целевой задаче на параллельном корпусе и используют предобучение на параллельных данных, тогда как БЯМ решают задачу без дообучения, опираясь на постановку по нескольким примерам.

Полученные результаты (табл. 2 и 3) демонстрируют преимущество трансформерных моделей над БЯМ по метрике BLEU в рассматриваемой постановке. Для трансформеров значения BLEU находятся в диапазоне 0,70–0,84 (табл. 3), при этом наилучший результат достигается моделью mT5-small (0,84), а сопоставимые значения демонстрируют модели nllb-200-distilled-600M и mbart-large-50 (0,83). Кроме того, модель mT5-small показала лучшее качество выполнения задачи на уровне *n*-грамм высокого ранга: достигнуто максимальное значение 0,65 для 4-грамм. Для БЯМ наблюдается существенно более широкий разброс качества, а максимальное значение BLEU достигается моделью GPT-5.2 (0,60) (табл. 2), что ниже значений всех рассматриваемых трансформерных моделей.

Среди БЯМ более высокие значения BLEU в основном продемонстрировали рассуждающие модели по сравнению с диалоговыми. В табл. 2 максималь-

[1] mBART-50. 2021 [Электронный ресурс]. URL: https://huggingface.co/facebook/mbart-large-50 (дата обращения: 20.02.2026).

[2] NLLB-200. 2022 [Электронный ресурс]. URL: https://huggingface.co/facebook/nllb-200-distilled-600M (дата обращения: 20.02.2026).

[3] Opus-mt-ru-en. 2020 [Электронный ресурс]. URL: https://huggingface.co/Helsinki-NLP/opus-mt-ru-en (дата обращения: 20.02.2026).

[4] mT5. 2020 [Электронный ресурс]. URL: https://huggingface.co/google/mt5-small (дата обращения: 20.02.2026).

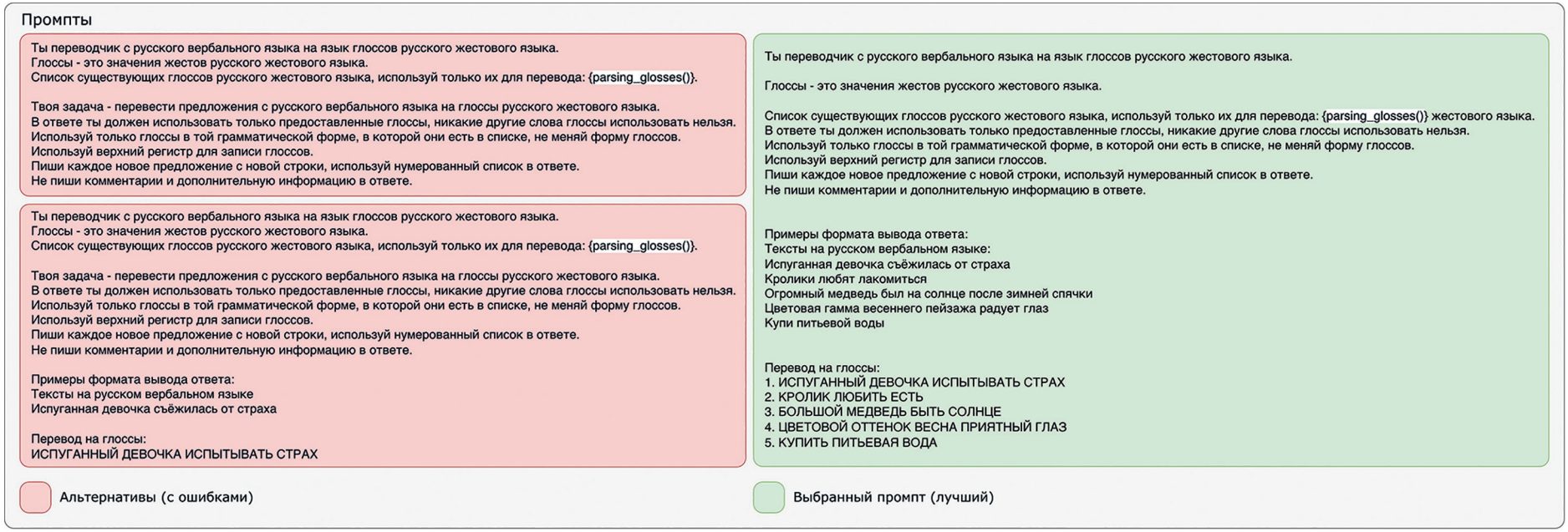

Рис. 2. Варианты промптов для задачи машинного перевода «вербальный текст-последовательность глосс»

Fig. 2. Prompt variants for machine translation “verbal text-gloss sequence”

Таблица 1. Сравнительно-сопоставительный анализ корпусов жестовых языков
Table 1. Comparative analysis of sign language corpora

Корпус	Жестовый язык	Наполнение	Количество видео	Классы	Открытый доступ
Bukva[1]	Русский	Дактиль	3757	33	Да
Slovo [1]	Русский	Дактиль, слова и словосочетания	20 400	1000	Да (частично)
TheRuSLan [2, 3]	Русский	Слова и словосочетания	10 660	164	Нет
Logos[2]	Русский	Слова и словосочетания	199 668	2863	Нет
Поликлинический [4]	Русский	Слова и словосочетания	5100	85	Нет
WLASL [5]	Американский английский	Слова и словосочетания	21 083	2000	Да
AUTSL [6]	Турецкий	Слова и словосочетания	38 336	226	Да (частично)
CSL[3]	Китайский	Слова и словосочетания	25 000	178	Нет

Таблица 2. Результаты экспериментов по задаче машинного перевода «вербальный текст-последовательность глосс» с использованием БЯМ
Table 2. Experimental results for "spoken-language text-gloss sequence" translation task using LLMs

Модель	КТ	Режим вывода	BLEU-1	BLEU-2	BLEU-3	BLEU-4	BLEU
Olmo-3.1-32B-Instruct[4]	27 460	Диалоговая	0,06	0,15	0,05	0,00	0,06
Llama-3.3-70B-Instruct[5]	20 686	Диалоговая	0,12	0,21	0,08	0,03	0,12
Llama-3.1-8B-Instruct[6]	20 685	Диалоговая	0,15	0,21	0,10	0,06	0,16
Gpt-oss-120B[7]	30 676	Рассуждающая	0,11	0,18	0,08	0,02	0,17
GigaChat[8]	17 127	Диалоговая	0,23	0,31	0,18	0,09	0,24
GigaChat-2[8]	17 112	Диалоговая	0,24	0,32	0,19	0,09	0,25
Gemma-3-27B-It[9]	18 416	Диалоговая	0,36	0,47	0,31	0,16	0,27
Qwen3-32B[10]	28 245	Рассуждающая	0,26	0,37	0,20	0,10	0,27
YandexGPT-5-Lite-8B-instruct[11]	13 592	Диалоговая	0,26	0,35	0,20	0,11	0,27
GigaChat-2-Pro[8]	16 978	Диалоговая	0,26	0,36	0,20	0,10	0,27
Алиса AI[12]	13 570	Диалоговая	0,33	0,42	0,29	0,16	0,33
GigaChat-2-Max[8]	17 088	Диалоговая	0,32	0,43	0,27	0,13	0,33
Command-a-reasoning-08-2025[13]	24 788	Рассуждающая	0,35	0,46	0,29	0,16	0,36
DeepSeek-V3-0324[14]	20 582	Рассуждающая	0,38	0,48	0,31	0,19	0,39
GPT-5.2[15]	27 238	Рассуждающая	0,58	0,66	0,54	0,37	0,60

Примечание: КТ – количество токенов.

[1] Bukva: Russian Sign Language Alphabet Dataset. 2024 [Электронный ресурс]. URL: https://github.com/ai-forever/bukva (дата обращения: 20.02.2026).

[2] Logos as a Well-Tempered Pre-Train for Sign Language Recognition. 2025 [Электронный ресурс]. URL: https://github.com/ai-forever/Logos (дата обращения: 20.02.2026).

[3] Chinese Sign Language (CSL) Dataset. 2024 [Электронный ресурс]. URL: https://github.com/woshisad159/TFNet (дата обращения: 20.02.2026).

[4] Olmo-3.1-32B-Instruct. 2025 [Электронный ресурс]. URL: https://huggingface.co/allenai/Olmo-3.1-32B-Instruct (дата обращения: 20.02.2026).

[5] Llama-3.3-70B-Instruct. 2024 [Электронный ресурс]. URL: https://huggingface.co/meta-llama/Llama-3.3-70B-Instruct (дата обращения: 20.02.2026).

[6] Llama-3.1-8B-Instruct. 2022 [Электронный ресурс]. URL: https://huggingface.co/meta-llama/Llama-3.1-8B-Instruct (дата обращения: 20.02.2026).

[7] Gpt-oss-120B. 2025 [Электронный ресурс]. URL: https://huggingface.co/openai/gpt-oss-120b (дата обращения: 20.02.2026).

[8] Модели GigaChat. 2026 [Электронный ресурс]. URL: https://developers.sber.ru/docs/ru/gigachat/models/updates (дата обращения: 20.02.2026).

[9] Gemma-3-27B-It. 2025 [Электронный ресурс]. URL: https://huggingface.co/google/gemma-3-27b-it (дата обращения: 20.02.2026).

[10] Qwen3-32B. 2025 [Электронный ресурс]. URL: https://huggingface.co/Qwen/Qwen3-32B (дата обращения; 20.02.2026).

[11] YandexGPT-5-Lite-8B-instruct. 2025 [Электронный ресурс]. URL: https://huggingface.co/yandex/YandexGPT-5-Lite-8B-instruct (дата обращения: 20.02.2026).

[12] Алиса AI. 2026 [Электронный ресурс]. URL: https://alice.yandex.ru/ (дата обращения: 20.02.2026).

[13] Command-a-reasoning-08-2025. 2025 [Электронный ресурс]. URL: https://huggingface.co/CohereLabs/command-a-reasoning-08-2025 (дата обращения: 20.02.2026).

[14] DeepSeek-V3-0324. 2025 [Электронный ресурс]. URL: https://huggingface.co/deepseek-ai/DeepSeek-V3-0324 (дата обращения: 20.02.2026).

[15] GPT 5.2. 2025 [Электронный ресурс]. URL: https://openai.com/ru-RU/index/introducing-gpt-5-2 (дата обращения: 20.02.2026).

Таблица 3. Результаты экспериментов по задаче машинного перевода «вербальный текст-последовательность глосс» с использованием трансформерных моделей

Table 3. Experimental results for “spoken-language text-gloss sequence” translation task using transformer models

Модель	ЧЭ	НГ	ВО, мин	BLEU-1	BLEU-2	BLEU-3	BLEU-4	BLEU
opus-mt-ru-en[1]	1000	0,00053	около 122	0,66	0,71	0,65	0,50	0,70
nllb-200-distilled-600M[2]	250	0,05741	около 82	0,77	0,82	0,76	0,62	0,83
mbart-large-50[3]	250	0,00324	около 84	0,76	0,81	0,77	0,64	0,83
mT5-small[4]	1000	2,25835	около 91	0,77	0,80	0,76	0,65	0,84

Примечание: ЧЭ — число эпох; НГ — норма градиента; ВО — время обучения.

ные значения метрики получены для моделей GPT-5.2, DeepSeek-V3-0324 и Command-a-reasoning-08-2025, тогда как большинство диалоговых моделей формируют результаты 0,33 и ниже. Наблюдаемое различие согласуется с тем, что в рассматриваемой задаче требуется строгое следование ограничениям на словарь и формат ответа, а также точное воспроизведение дискретной последовательности глосс, что в большей степени обеспечивается при использовании специализированных стратегий вывода.

В целом полученные результаты показали, что при наличии вручную размеченного параллельного корпуса наилучшее качество перевода достигается трансформерными моделями после адаптации, тогда как БЯМ целесообразны в качестве базового сравнения при решении задачи без дообучения на целевых парах.

Заключение

В работе выполнено экспериментальное сравнение трансформерных моделей архитектуры «энкодер-декодер» и больших языковых моделей архитектуры «только декодер» в задаче машинного перевода «вербальный текст-последовательность глосс» при единых правилах подготовки данных и оценки качества. Полученные результаты показали преимущество трансформерных моделей по метрике Bilingual Evaluation Understudy в постановке задачи «вербальный текст-последовательность глосс». Дополнительно установлено, что среди больших языковых моделей более высокие значения Bilingual Evaluation Understudy преимущественно достигаются моделями рассуждающего типа, что согласуется с повышенными требованиями задачи к строгому соблюдению ограничений на словарь и формат ответа.

Предложенный метод ориентирован на использование последовательности глосс как промежуточного текстового представления и может рассматриваться как компонент интеллектуальных систем двусторонней коммуникации между носителями жестовых и вербальных языков, включая сценарии последующего синтеза жестовой речи с применением цифровых аватаров. При этом результаты подчеркивают практическую целесообразность применения предобученных мультиязычных трансформеров с последующей адаптацией на параллельном корпусе, когда требуется максимально точное воспроизведение дискретной последовательности из фиксированного словаря глосс.

Исходные материалы, параллельный корпус и инструкции по воспроизведению экспериментов размещены в репозитории[5].

В дальнейшей работе планируется переход к обратному направлению перевода «последовательность глосс-вербальный текст». Для данной постановки наряду с *n*-граммными метриками целесообразно использовать семантические показатели качества для учета допустимой синонимической вариативности вербального вывода.

[1] Opus-mt-ru-en. 2020 [Электронный ресурс]. URL: https://huggingface.co/Helsinki-NLP/opus-mt-ru-en (дата обращения: 20.02.2026).

[2] NLLB-200. 2022 [Электронный ресурс]. URL: https://huggingface.co/facebook/nllb-200-distilled-600M (дата обращения: 20.02.2026).

[3] mBART-50. 2021 [Электронный ресурс]. URL: https://huggingface.co/facebook/mbart-large-50 (дата обращения: 20.02.2026).

[4] mT5. 2020 [Электронный ресурс]. URL: https://huggingface.co/google/mt5-small (дата обращения: 20.02.2026).

[5] [Электронный ресурс]. Режим доступа: https://github.com/Arseniy-Polyakov/text-to-gloss-translation (дата обращения: 20.02.2026).

Литература

1. Kapitanov A., Karina K., Nagaev A., Elizaveta P. Slovo: Russian sign language dataset // Lecture Notes in Computer Science. 2023. V. 14253. P. 63–73. doi: 10.1007/978-3-031-44137-0_6
2. Kagirov I., Ivanko D., Ryumin D., Axyonov A., Karpov A. TheRuSLan: Database of Russian sign language // Proc. of the 12th Language Resources and Evaluation Conference. 2020. P. 6079–6085.
3. Ryumin D., Ivanko D., Axyonov A., Kagirov I., Karpov A., Zelezny M. Human-robot interaction with smart shopping trolley using sign language: data collection // Proc. of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). 2019. P. 949–954. doi: 10.1109/percomw.2019.8730886

References

1. Kapitanov A., Karina K., Nagaev A., Elizaveta P. Slovo: Russian sign language dataset. *Lecture Notes in Computer Science*, 2023, vol. 14253, pp. 63–73. doi: 10.1007/978-3-031-44137-0_6
2. Kagirov I., Ivanko D., Ryumin D., Axyonov A., Karpov A. TheRuSLan: Database of Russian sign language. *Proc. of the 12th Language Resources and Evaluation Conference*, 2020, pp. 6079–6085.
3. Ryumin D., Ivanko D., Axyonov A., Kagirov I., Karpov A., Zelezny M. Human-robot interaction with smart shopping trolley using sign language: data collection. *Proc. of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2019, pp. 949–954. doi: 10.1109/percomw.2019.8730886

4. Кагиров И.А., Рюмин Д.А. База данных русского жестового языка поликлинического предназначения: лингвистические особенности материала и аннотирования // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2022. Т. 20. № 3. С. 90–108. doi: 10.25205/1818-7935-2022-20-3-90-108
5. Li D., Opazo C.R., Yu X., Li H. Word-level deep sign language recognition from video: a new large-scale dataset and methods comparison // Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV). 2020. P. 1448–1458. doi: 10.1109/wacv45572.2020.9093512
6. Sincan O.M., Keles H.Y. AUTSL: A Large scale multi-modal Turkish sign language dataset and baseline methods // IEEE Access. 2020. V. 8. P. 181340–181355. doi: 10.1109/ACCESS.2020.3028072
7. Kapitanov A., Kvanchiani K., Nagaev A., Kraynov R., Makhliarchuk A. HaGRID-HAnd Gesture Recognition Image Dataset // Proc. of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2024. P. 4560–4569. doi: 10.1109/WACV57701.2024.00451
8. Zhu D., Czehmann V., Avramidis E. Neural machine translation methods for translating text to sign language glosses // Proc. of the 61st Annual Meeting of the Association for Computational Linguistics. 2023. P. 12523–12541. doi: 10.18653/v1/2023.acl-long.700
9. Rust P., Shi B., Wang S., Camgoz N.C., Maillard J. Towards privacy-aware sign language translation at scale // Proc. of the 62nd Annual Meeting of the Association for Computational Linguistics. 2024. P. 8624–8641. doi: 10.18653/v1/2024.acl-long.467
10. Zhang B., Tanzer G., Firat O. Scaling sign language translation // Proc. of the 38th International Conference on Neural Information Processing Systems. 2024. P. 114018–114047.
11. Li Z., Zhou W., Zhao W., Wu K., Hu H., Li H. Uni-Sign: Toward unified sign language understanding at scale // Proc. of the 13th International Conference on Learning Representations (ICLR). 2025. P. 1–20.
12. Hamidullah Y., van Genabith J., España-Bonet C. Sign language translation with sentence embedding supervision // Proc. of the 62nd Annual Meeting of the Association for Computational Linguistics. 2024. P. 425–434. doi: 10.18653/v1/2024.acl-short.40
13. Hamidullah Y., Yazdani S., Oguz C., van Genabith J., España-Bonet C. SONAR-SLT: Multilingual sign language translation via language-agnostic sentence embedding supervision // Proc. of the 10th Conference on Machine Translation. 2025. P. 301–313. doi: 10.18653/v1/2025.wmt-1.18
14. Toshpulatov M., Lee W., Jun J., Lee S. Deep learning pathways for automatic sign language processing // Pattern Recognition. 2025. V. 164. P. 111475. doi: 10.1016/j.patcog.2025.111475
15. Gong J., Foo L.G., He Y., Rahmani H., Liu J. LLMs are good sign language translators // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024. P. 18362–18372. doi: 10.1109/CVPR52733.2024.01738
16. Hwang E.J., Cho S., Lee J., Park J.C. An efficient gloss-free sign language translation using spatial configurations and motion dynamics with LLMs // Proc. of the Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies. 2025. P. 3901–3920. doi: 10.18653/v1/2025.naacl-long.197
17. Guo J., Li P., Cohn T. Bridging sign and spoken languages: pseudo gloss generation for sign language translation // Proc. of the 39th Conference on Neural Information Processing Systems (NeurIPS). 2025. P. 1–29.
18. Baltatzis V., Potamias R.A., Ververas E., Sun G., Deng J., Zafeiriou S. Neural sign actors: a diffusion model for 3D sign language production from text // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024. P. 1985–1995. doi: 10.1109/CVPR52733.2024.00194
19. Tang S., Xue F., Wu J., Wang S., Hong R. Gloss-driven conditional diffusion models for sign language production // ACM Transactions on Multimedia Computing, Communications, and Applications. 2025. V. 21. N 4. P. 105. doi: 10.1145/3663572
20. Ivanko D., Ryumin D. Intelligent system for automatic bidirectional sign language translation based on recognition and synthesis of audiovisual and sign speech // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2025. V. XLVIII-2/W9-2025. P. 131–136. doi: 10.5194/isprs-archives-xlviii-2-w9-2025-131-2025
21. Chipman H.A., George E.I., McCulloch R.E., Shively T.S. mBART: Multidimensional Monotone BART // Bayesian Analysis. 2022. V. 17. N 2. P. 515–544. doi: 10.1214/21-BA1259

4. Kagirov I.A., Ryumin D.A. Russian sign language database for clinical use: data and annotation peculiarities. *NSU Vestnik. Series: Linguistics and Intercultural Communication*, 2022, vol. 20, no. 3, pp. 90–108. (in Russian). doi: 10.25205/1818-7935-2022-20-3-90-108
5. Li D., Opazo C.R., Yu X., Li H. Word-level deep sign language recognition from video: a new large-scale dataset and methods comparison. *Proc. of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2020, pp. 1448–1458. doi: 10.1109/wacv45572.2020.9093512
6. Sincan O.M., Keles H.Y. AUTSL: A Large scale multi-modal Turkish sign language dataset and baseline methods. *IEEE Access*, 2020, vol. 8, pp. 181340–181355. doi: 10.1109/ACCESS.2020.3028072
7. Kapitanov A., Kvanchiani K., Nagaev A., Kraynov R., Makhliarchuk A. HaGRID-HAnd Gesture Recognition Image Dataset. *Proc. of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024, pp. 4560–4569. doi: 10.1109/WACV57701.2024.00451
8. Zhu D., Czehmann V., Avramidis E. Neural machine translation methods for translating text to sign language glosses. *Proc. of the 61st Annual Meeting of the Association for Computational Linguistics*, 2023, pp. 12523–12541. doi: 10.18653/v1/2023.acl-long.700
9. Rust P., Shi B., Wang S., Camgoz N.C., Maillard J. Towards privacy-aware sign language translation at scale. *Proc. of the 62nd Annual Meeting of the Association for Computational Linguistics*, 2024, pp. 8624–8641. doi: 10.18653/v1/2024.acl-long.467
10. Zhang B., Tanzer G., Firat O. Scaling sign language translation. *Proc. of the 38th International Conference on Neural Information Processing Systems*, 2024, pp. 114018–114047.
11. Li Z., Zhou W., Zhao W., Wu K., Hu H., Li H. Uni-Sign: Toward unified sign language understanding at scale. *Proc. of the 13th International Conference on Learning Representations (ICLR)*, 2025, pp. 1–20.
12. Hamidullah Y., van Genabith J., España-Bonet C. Sign language translation with sentence embedding supervision. *Proc. of the 62nd Annual Meeting of the Association for Computational Linguistics*, 2024, pp. 425–434. doi: 10.18653/v1/2024.acl-short.40
13. Hamidullah Y., Yazdani S., Oguz C., van Genabith J., España-Bonet C. SONAR-SLT: Multilingual sign language translation via language-agnostic sentence embedding supervision. *Proc. of the 10th Conference on Machine Translation*, 2025, pp. 301–313. doi: 10.18653/v1/2025.wmt-1.18
14. Toshpulatov M., Lee W., Jun J., Lee S. Deep learning pathways for automatic sign language processing. *Pattern Recognition*, 2025, vol. 164, pp. 111475. doi: 10.1016/j.patcog.2025.111475
15. Gong J., Foo L.G., He Y., Rahmani H., Liu J. LLMs are good sign language translators. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 18362–18372. doi: 10.1109/CVPR52733.2024.01738
16. Hwang E.J., Cho S., Lee J., Park J.C. An efficient gloss-free sign language translation using spatial configurations and motion dynamics with LLMs. *Proc. of the Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2025, pp. 3901–3920. doi: 10.18653/v1/2025.naacl-long.197
17. Guo J., Li P., Cohn T. Bridging sign and spoken languages: pseudo gloss generation for sign language translation. *Proc. of the 39th Conference on Neural Information Processing Systems (NeurIPS)*, 2025, pp. 1–29.
18. Baltatzis V., Potamias R.A., Ververas E., Sun G., Deng J., Zafeiriou S. Neural sign actors: a diffusion model for 3D sign language production from text. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 1985–1995. doi: 10.1109/CVPR52733.2024.00194
19. Tang S., Xue F., Wu J., Wang S., Hong R. Gloss-driven conditional diffusion models for sign language production. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2025, vol. 21, no. 4, pp. 105. doi: 10.1145/3663572
20. Ivanko D., Ryumin D. Intelligent system for automatic bidirectional sign language translation based on recognition and synthesis of audiovisual and sign speech. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2025, vol. XLVIII-2/W9-2025, pp. 131–136. doi: 10.5194/isprs-archives-xlviii-2-w9-2025-131-2025
21. Chipman H.A., George E.I., McCulloch R.E., Shively T.S. mBART: Multidimensional Monotone BART. *Bayesian Analysi*s, 2022, vol. 17, no. 2, pp. 515–544. doi: 10.1214/21-BA1259

22. Tiedemann J., Aulamo M., Bakshandaeva D., Boggia M., Grönroos S.-A., Nieminen T., et al. Democratizing neural machine translation with OPUS-MT // Language Resources and Evaluation. 2024. V. 58. N 2. P. 713–755. doi: 10.1007/s10579-023-09704-w
23. Xue L., Constant N., Roberts A., Kale M., Al-Rfou R., Siddhant A., et al. mT5: A Massively multilingual pre-trained text-to-text transformer // Proc. of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL). 2021. P. 483–498. doi: 10.18653/v1/2021.naacl-main.41
24. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a method for automatic evaluation of machine translation // Proc. of the 40th Annual Meeting on Association for Computational Linguistics. 2002. P. 311–318. doi: 10.3115/1073083.1073135
25. Post M. A call for clarity in reporting BLEU scores // Proc. of the 3rd Conference on Machine Translation: Research Papers. 2018. P. 186–191. doi: 10.18653/v1/W18-6319

22. Tiedemann J., Aulamo M., Bakshandaeva D., Boggia M., Grönroos S.-A., Nieminen T., et al. Democratizing neural machine translation with OPUS-MT. *Language Resources and Evaluation*, 2024, vol. 58, no. 2, pp. 713–755. doi: 10.1007/s10579-023-09704-w
23. Xue L., Constant N., Roberts A., Kale M., Al-Rfou R., Siddhant A., et al. mT5: A Massively multilingual pre-trained text-to-text transformer. *Proc. of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, 2021, pp. 483–498. doi: 10.18653/v1/2021.naacl-main.41
24. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a method for automatic evaluation of machine translation. *Proc. of the 40th Annual Meeting on Association for Computational Linguistics*, 2002, pp. 311–318. doi: 10.3115/1073083.1073135
25. Post M. A call for clarity in reporting BLEU scores. *Proc. of the 3rd Conference on Machine Translation: Research Papers*, 2018, pp. 186–191. doi: 10.18653/v1/W18-6319

Авторы

Поляков Арсений Михайлович — стажер-исследователь, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, https://orcid.org/0000-0002-8681-988X, arseney02@mail.ru
Рюмин Дмитрий Александрович — кандидат технических наук, старший научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, sc 57191960214, https://orcid.org/0000-0002-7935-0569, ryumin.d@iias.spb.su

Authors

Arseniy M. Polyakov — Intern Developer, St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Saint Petersburg, 199178, Russian Federation, https://orcid.org/0000-0002-8681-988X, arseney02@mail.ru
Dmitry A. Ryumin — PhD, Senior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg (SPC RAS), 199178, Russian Federation, sc 57191960214, https://orcid.org/0000-0002-7935-0569, ryumin.d@iias.spb.su

Статья поступила в редакцию 22.02.2026
Одобрена после рецензирования 09.03.2026
Принята к печати 20.05.2026

Received 22.02.2026
Approved after reviewing 09.03.2026
Accepted 20.05.2026

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»