СЛОВО МОЛОДЫМ

НАУЧНАЯ СТАТЬЯ

УДК: 81’322.4:81’33

DOI: 10.55959/MSU2074-6636-22-2026-19-2-131-156

МЕЖАГЕНТНАЯ ОЦЕНКА КАЧЕСТВА МАШИННОГО ПЕРЕВОДА В ПРЕДМЕТНЫХ ОБЛАСТЯХ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ GOOGLE TRANSLATE, YANDEX TRANSLATE И DEEPL

Виктор Олегович Савостьянов
Максимилиан Вячеславович Жидков
Дарья Романовна Инякина
Алина Васильевна Чугункина

Пензенский государственный университет, г. Пенза, Россия

Для контактов: victsav@gmail.com

Аннотация. **Цель:** сравнительный анализ качества машинного перевода трёх ведущих систем (Google Translate, Yandex Translate, DeepL) в трёх предметных областях (медицина, юриспруденция, публицистика) с использованием оригинального межагентного подхода на базе YandexGPT.

Методы: исследование выполнено на материале 60 англоязычных текстов (по 20 каждой тематики), переведённых тремя движками (всего 180 переводов). Для оценки привлечена система из шести виртуальных экспертов-агентов на базе YandexGPT: лингвист, семантик, терминолог, прагматик, культурный адаптер и тематический эксперт (медицинский/юридический/эксперт по медиа дискурсу). Агенты оценивали переводы по 10-балльной шкале; итоговое решение выносил агент-модератор. Статистическая обработка включала дисперсионный анализ (ANOVA), t-критерий Стьюдента и корреляционный анализ Пирсона/Спирмена.

Результаты: DeepL показал наилучший общий результат (8,57±0,69), особенно в юридической сфере (8,47). Google Translate неожиданно оказался лучшим в медицинской тематике (8,80) и публицистике (8,59). Yandex Translate уступил конкурентам во всех областях, продемонстрировав наибольшую вариативность (7,91±1,06) и критические ошибки в юридиче-

ских текстах (минимум 5,0). Выявлены статистически значимые различия между движками ($p < 0,05$). Анализ корреляций между агентами показал высокую согласованность лингвистической и семантической оценок ($r = 0,82$).

Выводы: Предложенный межагентный подход позволяет получать объяснимые, дифференцированные по аспектам оценки. DeepL является универсальным лидером, особенно в юриспруденции; Google предпочтителен для медицины и публицистики; Yandex требует доработки для специальных областей. Результаты могут быть использованы для оптимизации постредактирования и выбора систем перевода в профессиональных задачах.

Ключевые слова: машинный перевод, межагентные системы, YandexGPT, оценка качества, Google Translate, DeepL, Yandex Translate, медицинский перевод, юридический перевод

Для цитирования: *Савостьянов В.О., Жидков М.В., Инякина Д.Р., Чугункина А.В.* Межагентная оценка качества машинного перевода в предметных областях: сравнительный анализ Google Translate, Yandex Translate и DeepL // Вестник Московского университета. Серия 22. Теория перевода, 2026. № 2. С. 131–156. DOI: 10.55959/MSU2074-6636-22-2026-19-2-131-156

Статья поступила в редакцию 20.02.2026;
одобрена после рецензирования 05.05.2026;
принята к публикации 26.05.2026.

MULTI-AGENT ASSESSMENT OF MACHINE TRANSLATION QUALITY ACROSS SUBJECT DOMAINS: A COMPARATIVE ANALYSIS OF GOOGLE TRANSLATE, YANDEX TRANSLATE AND DEEPL

Viktor O. Savostyanov
Maksimilian V. Zhidkov
Daria R. Inyakina
Alina V. Chugunkina

Penza State University, Penza, Russia

For contacts: victsav@gmail.com

Abstract. **Object:** this study presents a comparative analysis of the quality of machine translation (MT) produced by three leading systems (Google Translate, Yandex Translate, DeepL) across three specialized domains (medicine, law, and

news articles). The analysis employs an original multi-agent approach based on the YandexGPT large language model.

Methods: the research material comprises 60 English-language texts (20 per domain), each translated by the three MT engines, resulting in a total of 180 translations for evaluation. A system of six virtual expert agents powered by YandexGPT was developed for assessment: a linguist, a semanticist, a terminologist, a pragmatist, a cultural adaptor, and a domain-specific expert (medical/legal/media discourse). Each agent evaluated translations on a 10-point scale. An agent-moderator then synthesized these individual assessments to produce a final score. Statistical analysis included ANOVA, Student's t-test, and Pearson/Spearman correlation analysis.

Findings: DeepL achieved the highest overall score (8.57±0.69), demonstrating particular strength in the legal domain (8.47). Google Translate unexpectedly excelled in the medical domain (8.80) and also led in media texts (8.59). Yandex Translate ranked third across all domains, exhibiting the highest score variability (7.91±1.06) and critical errors in legal translations (minimum score of 5.0). Statistically significant differences were found between the MT engines ($p < 0.05$). Correlation analysis of agent scores revealed high consistency between the linguist and semanticist ($r = 0.82$).

Conclusions: the proposed multi-agent approach provides explainable, multi-aspect evaluations. DeepL emerges as a versatile leader, particularly for legal translation; Google is preferable for medical and media texts; Yandex requires further development for specialized domains. These findings can inform post-editing strategies and guide the selection of MT systems for professional tasks.

Keywords: machine translation, multi-agent systems, YandexGPT, quality assessment, Google Translate, DeepL, Yandex Translate, medical translation, legal translation

For citation: *Savostyanov V.O., Zhidkov M.V., Inyakina D.R., Chugunkina A.V.* (2026) Multi-Agent Assessment of Machine Translation Quality Across Subject Domains: A Comparative Analysis of Google Translate, Yandex Translate and DeepL. Vestnik Moskovskogo Universiteta. Seriya 22. Teorija Perevoda — Moscow University Bulletin on Translation Studies. 2. P. 131–156. DOI: 10.55959/MSU2074-6636-22-2026-19-2-131-156

The article was received on February 20, 2026;
approved after reviewing received on May 05, 2026;
accepted for publication on May 26, 2026.

Введение

Современный мир невозможно представить без автоматизированного перевода. Ежедневно миллионы людей используют системы машинного перевода для работы, учёбы и общения. Три

гиганта индустрии — Google Translate, Yandex Translate и DeepL — стали неотъемлемыми инструментами в глобальной коммуникации. Однако вопрос о том, какой из движков лучше справляется с переводом специальных текстов (медицинских, юридических, публицистических), остаётся открытым и крайне актуальным для профессионального сообщества.

Традиционные метрики оценки качества машинного перевода, такие как BLEU (Papineni et al., 2002) и chrF (Popović, 2015), обладают существенными ограничениями. Они основаны на n-граммном совпадении с эталонным переводом и не учитывают содержательные, стилистические и прагматические аспекты. Более того, для многих языковых пар и предметных областей эталонные переводы просто отсутствуют. Это создаёт потребность в разработке новых методов оценки, которые были бы одновременно объективными, многомерными и объяснимыми.

С появлением больших языковых моделей (LLM) открылись новые возможности для автоматической оценки качества перевода. Модели типа GPT, Claude и YandexGPT способны не только генерировать текст, но и выступать в роли экспертов, оценивая переводы по различным критериям (Kocmi & Federmann, 2023). Однако использование одной модели как универсального судьи также имеет недостатки: односторонность оценки, потенциальные систематические ошибки, зависимость от конкретной модели.

В данной работе мы предлагаем принципиально новый подход — межагентную систему оценки качества перевода. Вместо одного эксперта мы создаём команду виртуальных специалистов, каждый из которых оценивает перевод с определённой профессиональной позиции: лингвист анализирует грамматику и стиль, семантик — сохранение смысла, терминолог — точность специальной лексики, прагматик — коммуникативную эффективность, культурный адаптер — адекватность для русскоязычной аудитории. Дополнительно привлекается тематический эксперт (медицинский, юридический или эксперт по медиадискурсу), а итоговое решение выносит агент-модератор, синтезирующий все оценки.

Актуальность исследования обусловлена несколькими факторами: 1) практической потребностью в объективных данных о качестве перевода в специальных областях для переводчиков, редакторов и организаций; 2) необходимостью оценить, насколько современные нейросетевые движки (особенно DeepL, позиционируемый как специализированный) действительно превосходят универсальных конкурентов; 3) методологическим вызовом — раз-

работкой и валидацией нового подхода к оценке качества перевода с использованием LLM в качестве экспертов.

Выбор предметных областей не случаен. Медицинские тексты требуют абсолютной точности терминологии, юридические — однозначности формулировок, публицистические — сохранения стиля и эмоциональной окраски. Эти три сферы покрывают основные вызовы, стоящие перед машинным переводом.

Цель настоящего исследования — провести сравнительный анализ качества перевода трёх систем (Google Translate, Yandex Translate, DeepL) в трёх предметных областях (медицина, юриспруденция, публицистика) с использованием разработанной межагентной системы на базе YandexGPT. Для достижения цели были поставлены следующие задачи: 1) сформировать корпус из 60 англоязычных текстов (по 20 каждой тематики) и получить их переводы тремя движками; 2) разработать архитектуру межагентной системы, включающую шесть специализированных экспертов и модератора; 3) создать систему промптов для каждого агента, обеспечивающих объективную многокритериальную оценку; 4) провести оценку 180 переводов и собрать детальные данные по всем аспектам качества; 5) выполнить статистический анализ результатов, выявить значимые различия и корреляции; 6) интерпретировать полученные данные и сформулировать практические рекомендации.

Научная новизна работы заключается во впервые применённом межагентном подходе к оценке качества перевода с использованием YandexGPT, в разработке архитектуры из шести специализированных агентов, обеспечивающей многомерную оценку, в эмпирическом сравнении трёх ведущих движков на репрезентативном материале трёх предметных областей, а также в выявлении неожиданных закономерностей (лидерство Google в медицине, провал Yandex в юриспруденции).

Обзор литературы

Методы оценки качества машинного перевода: от автоматических метрик к экспертным системам

Проблема оценки качества машинного перевода (МП) имеет долгую историю, восходящую к первым системам rule-based перевода 1950-х годов. Традиционно выделяют два подхода: автоматические метрики и оценка человеком.

Автоматические метрики доминируют в исследованиях благодаря своей воспроизводимости и низкой стоимости. Наиболее

известная метрика — BLEU (Bilingual Evaluation Understudy), предложенная Papineni et al. (2002), основана на подсчёте совпадений n-грамм между машинным и эталонным переводами. Несмотря на широкое распространение, BLEU подвергается критике за слабую корреляцию с человеческой оценкой, особенно в языках с богатой морфологией (Callison-Burch et al., 2006). Альтернативные метрики, такие как chrF (Popović, 2015), оперирующая символами вместо слов, и TER (Translation Edit Rate) (Snover et al., 2006), измеряющая количество редактирований, также не лишены недостатков: все они требуют наличия эталонных переводов и не учитывают содержательные аспекты.

Человеческая оценка остаётся золотым стандартом, но страдает от субъективности, высокой стоимости и трудностей масштабирования. Классические схемы включают ранжирование переводов (Koehn, 2004) и оценку по шкалам адекватности и беглости. Исследования показывают, что даже обученные эксперты демонстрируют значительные расхождения в оценках (Graham et al., 2013), что ставит вопрос о надёжности человеческого суждения.

Новый рубеж: LLM как судьи. С появлением больших языковых моделей (LLM) возник подход LLM-as-a-judge, где модель выступает в роли эксперта. Kocmi и Federmann (2023) первыми систематически исследовали способность GPT-4 оценивать качество перевода, показав высокую корреляцию с человеческими оценками ($r = 0{,}85$). Последующие работы подтвердили эффективность этого подхода для разных языковых пар. Однако использование одной модели создаёт риск систематических ошибок и зависимости от конкретного LLM.

Межагентные системы в обработке естественного языка

Теоретические основы межагентных систем были заложены в работах Wooldridge и Jennings (1995), определивших агента как компьютерную систему, способную к автономному действию в динамической среде. В применении к NLP межагентный подход позволяет моделировать сложные коммуникативные процессы и распределённое решение задач.

Современные исследования демонстрируют эффективность MAS в различных задачах NLP. Tan et al. (2024) предложили архитектуру Multi-Agent Debate, где несколько агентов обсуждают решение, приходя к более точному результату. Wu et al. (2024) применили межагентный подход к задаче реферирования текстов, показав, что коллаборация специализированных агентов превосходит

единую модель. Liang et al. (2024) разработали систему агентов для оценки креативности текстов, где каждый агент отвечал за отдельный критерий.

В области перевода межагентные системы применялись преимущественно для генерации, а не для оценки. Zhang et al. (2024) предложили систему Multi-Agent Translation, где агенты-переводчики и агенты-редакторы совместно улучшают качество перевода. Однако применение межагентного подхода для оценки перевода остаётся малоисследованной областью, что определяет новизну настоящей работы.

Сравнительные исследования систем машинного перевода

Сравнительный анализ коммерческих систем МП — активно развивающаяся область. DeepL, появившийся в 2017 году, быстро завоевал репутацию лидера благодаря использованию нейросетевых архитектур и специализированных обучающих данных. Исследования подтверждают превосходство DeepL для европейских языков (Ahrenberg, 2017; Macketanz et al., 2022), однако для русского языка данные противоречивы.

Google Translate, основанный на архитектуре Transformer (Vaswani et al., 2017), выигрывает от огромного объёма обучающих данных и постоянного обновления. Сравнительные исследования показывают, что Google особенно силён в переводах с английского на языки с ограниченными ресурсами (Popović, 2020).

Yandex Translate, российская система, имеет преимущество в понимании русскоязычных реалий и морфологии. Однако систематических сравнений по предметным областям ранее не проводилось.

Проблемы перевода в специальных областях

Медицинский перевод предъявляет особые требования к точности терминологии. Karwacka (2014) подчёркивает, что ошибка в переводе медицинского термина может иметь фатальные последствия. Montalt и González-Davies (2014) выделяют необходимость учёта жанровых особенностей (история болезни, инструкция к препарату, научная статья). Автоматические системы часто допускают ошибки в переводе аббревиатур и названий препаратов (Dew, 2020).

Юридический перевод требует абсолютной однозначности формулировок. Šarčević (1997) в своём фундаментальном труде по-

казала, что правовые понятия редко имеют прямые эквиваленты в разных правовых системах. Biel (2017) исследует проблему терминологической согласованности в юридических текстах. Машинный перевод часто порождает двусмысленности, недопустимые в юридическом дискурсе (Mattila, 2013).

Публицистический перевод ставит во главу угла сохранение стиля и эмоциональной окраски. Bani (2006) анализирует проблемы передачи идеологически окрашенной лексики в новостных текстах. Valdeón (2015) исследует, как машинный перевод нивелирует культурные особенности и авторскую интонацию. Сохранение прагматического потенциала текста остаётся вызовом для современных систем МП.

Проблемная ситуация и место настоящего исследования

Анализ литературы выявляет следующие противоречия и лакуны:

1. Методологический разрыв: автоматические метрики нечувствительны к содержательным аспектам, а человеческая оценка субъективна и дорога. LLM-as-a-judge предлагает компромисс, но единая модель не может учесть все критерии.

2. Отсутствие многомерных оценок: существующие сравнения движков обычно дают один интегральный балл, не раскрывая, по каким именно аспектам один движок превосходит другой.

3. Недостаток данных по предметным областям: хотя существуют отдельные исследования для медицины (Dew, 2020; Karwacka, 2014) и юриспруденции (Biel, 2017; Šarčević, 1997), систематического сравнения трёх ведущих движков на одном материале не проводилось.

4. Отсутствие российского контекста: большинство исследований выполнены на западноевропейских языках; сравнение DeepL, Google и Yandex для перевода на русский в специальных областях не проводилось.

Настоящее исследование заполняет эти лакуны, предлагая:

- межагентный подход с шестью специализированными экспертами;
- многомерную оценку по разным критериям;
- систематическое сравнение на материале трёх предметных областей;
- использование российской модели YandexGPT в качестве основы для агентов.

Материалы и методы

Материал исследования

Корпус исследования составили 60 англоязычных текстов, равномерно распределённых по трём тематическим областям: медицина (20 текстов), юриспруденция (20 текстов) и публицистика (20 текстов). Тексты были отобраны из открытых источников: Directory of Open Access Journals (DOAJ) и публицистических изданий.

Тексты отбирались по наличию переводческих трудностей, релевантных для каждой области. Медицинские тексты содержат специализированную анатомическую и клиническую терминологию, аббревиатуры, сложные номинативные цепочки. Юридические тексты насыщены модальными глаголами, клишированными формулировками, терминами с несовпадающим объёмом понятий в системах общего и континентального права. Публицистические тексты отличаются прагматической направленностью, элементами авторской оценки и культурными референциями.

Объём каждого текста составляет примерно 900 знаков с пробелами или половину переводческой страницы. Такой размер позволяет агентам удерживать контекст и не выходить за пределы эффективного окна памяти YandexGPT: современные языковые модели всё ещё ограничены в фиксации активного контекста.

Жанровая структура внутри тематик сформирована следующим образом:

- Медицина: аннотации к научным статьям, описания клинических случаев.
- Юриспруденция: отрывки статей по праву, извлечения из договоров и исковых заявлений.
- Публицистика (медийные тексты): отрывки новостных статей разных тематик (внутренняя политика США; внешняя политика США и конфликты; экономика, бизнес и технологии).

Каждый текст был переведён тремя системами машинного перевода: Google Translate, Yandex Translate и DeepL. Переводы выполнялись в феврале 2026 года для обеспечения актуальности результатов. Таким образом, общий объём анализируемого материала составил 180 переводов (60 исходных текстов × 3 системы).

Данные были организованы в формате CSV со следующими полями:

Domain — тематика (medical/legal/media);

segment_id — уникальный идентификатор текста;

source — исходный текст на английском;
google, yandex, deepl — соответствующие переводы.

Архитектура межагентной системы оценки

Для оценки качества переводов была разработана оригинальная межагентная система на базе языковой модели YandexGPT. Система включает семь виртуальных агентов, каждый из которых выполняет роль эксперта в определённой области.

Базовые агенты (присутствуют во всех оценках):

1. Лингвист (Linguist Agent) — оценивает грамматическую корректность, синтаксическую правильность, лексическое разнообразие, стилистическую естественность и отсутствие «машинного» звучания.
2. Семантик (Semantic Agent) — анализирует полноту и точность передачи смысла, сохранение логических связей и модальности.
3. Терминолог (Terminologist Agent) — проверяет корректность перевода специальных терминов, их согласованность и адекватность для профессиональной аудитории.
4. Прагматик (Pragmatist Agent) — оценивает сохранение интенции автора, адекватность для целевой аудитории, соответствие регистра и достижение прагматического эффекта.
5. Культурный адаптер (Cultural Adaptor Agent) — анализирует адаптацию культурных реалий, отсутствие культурного диссонанса, естественность для русскоязычного читателя.

Тематические агенты
(подключаются в зависимости от домена):

6. Медицинский эксперт (Medical Expert Agent) — для медицинских текстов оценивает точность перевода терминов, названий препаратов и процедур, соответствие стандартным формулировкам.
7. Юридический эксперт (Legal Expert Agent) — для юридических текстов проверяет точность правовых терминов, однозначность формулировок, корректность передачи модальных глаголов.
8. Эксперт по медиа дискурсу (Media Discourse Analyst) — для медийных текстов оценивает жанровую адекватность, сохранение прагматической интенции автора, передачу идеологических оттенков, дискурсивных стратегий, ритори-

ческих приёмов и стилистических особенностей, а также адаптацию культурных референций для русскоязычной аудитории.

Координирующий агент:

9. Модератор (Moderator Agent) — получает оценки всех экспертов, анализирует их согласованность и выносит итоговое решение: финальную оценку перевода по 10-балльной шкале, уровень согласованности экспертов, ключевые сильные и слабые стороны.

Промпты и процедура оценки

Для каждого агента был разработан детальный промпт (системное сообщение), задающий роль, критерии оценки и требуемый формат ответа. Все промпты были унифицированы по структуре и требовали вывода строго в формате JSON для автоматической обработки.

Оценка каждого перевода проходила в два этапа:

1. Параллельная оценка экспертами: все шесть агентов (пять базовых + один тематический) независимо оценивали перевод, получая на вход исходный текст и перевод.
2. Синтез модератором: модератор получал все оценки экспертов, анализировал их и выносил итоговое решение.

Все запросы выполнялись через YandexGPT API с температурой 0.1 для обеспечения воспроизводимости результатов. Между запросами устанавливалась пауза 1 секунда для соблюдения ограничений API.

Организация эксперимента

Эксперимент проводился в два этапа:

1. Пилотный этап — оценка 9 переводов (3 текста × 3 движка) для отладки системы и промптов.
2. Основной этап — оценка всех 180 переводов с полным набором агентов.

Обработка выполнялась с помощью Python-скриптов с использованием библиотек pandas, requests, tqdm. Результаты сохранялись в двух форматах:

– Полные данные (JSON) — все оценки экспертов, включая детальные комментарии;
– Сводная таблица (CSV) — итоговые оценки модератора по каждому переводу.

Статистическая обработка

Статистический анализ проводился с использованием языка Python 3.10 (библиотеки scipy, statsmodels, pandas). Применялись следующие методы:

- описательная статистика (среднее, стандартное отклонение, медиана, минимум, максимум);
- однофакторный дисперсионный анализ (ANOVA) для сравнения тематик и движков;
- t-критерий Стьюдента для попарного сравнения движков;
- корреляционный анализ Пирсона и Спирмена для оценки согласованности агентов.

Уровень статистической значимости был принят $p < 0,05$.

Этические аспекты и ограничения

Все тексты были взяты из открытых источников и использовались исключительно в исследовательских целях. В процессе оценки 4 публицистических текста или 12 переводов (6,7% выборки) были исключены из анализа в связи со срабатыванием встроенных механизмов модерации YandexGPT, отказавшейся оценивать тексты определённой тематики. Данный факт отмечен как ограничение исследования и особенность работы модели. После обнаружения систематических отказов YandexGPT (постфактум, когда переводы тремя движками уже были получены и эксперимент начат) замена текстов была невозможна без нарушения единых временных условий (фиксированная дата перевода — февраль 2026 г.).

Результаты

Общая статистика

Всего проанализировано 168 переводов (исключены 12 с отказами модерации). Средняя оценка составила 8,33±0,89, медиана — 8,50, минимум — 5,0, максимум — 9,7.

Сравнение движков по тематикам

В Таблице 1 представлены средние оценки и стандартные отклонения. DeepL показал наилучший общий результат (8,57±0,69), особенно в юридической сфере (8,47). Google Translate лидирует в медицинской тематике (8,80) и публицистике (8,59). Yandex Translate уступает во всех областях, особенно в юриспруденции (7,51), демонстрируя максимальную вариативность.

Таблица 1

Средние оценки (mean ± std) качества перевода по тематикам и движкам

Тематика/Domain	DeepL	Google	Yandex
Медицина/Medical	8,70±0,50	8,80±0,65	8,21±1,02
Юриспруденция/Legal	8,47±0,61	8,15±0,70	7,51±1,15
Публицистика/Media	8,54±0,48	8,59±0,52	8,02±0,65
Общая/Overall	8,57±0,69	8,51±0,73	7,91±1,06

Статистическая значимость различий

ANOVA выявил значимые различия между движками ($p < 0,05$). Попарное сравнение (t-критерий) показало статистически значимое превосходство DeepL над Yandex ($p < 0,01$) и Google над Yandex ($p < 0,05$). Различие между DeepL и Google не является статистически значимым на уровне $p < 0,05$.

Анализ согласованности агентов

Корреляционный анализ выявил высокую согласованность между лингвистом и семантиком ($r = 0,82$) и низкую корреляцию между прагматиком и терминологом ($r = 0,31$), что подтверждает независимость оцениваемых аспектов.

Примеры расхождения оценок агентов:

Тема: исследование TTP (тромботическая тромбоцитопеническая пурпура) с упоминанием PLT (тромбоциты).

Агент	Оценка
Лингвист (linguist)	9 (overall 8,6 в JSON)
Семантик (semantic)	9 (overall 8,75)
Терминолог (terminologist)	9
Прагматик (pragmatist)	7,75
Культурный адаптер (cultural_adaptor)	9,75
Медицинский эксперт (medical_expert)	8,75
Итог модератора	8,0

Суть расхождения: перевод оценён терминологически точно и семантически полно. Однако прагматик снизил балл из-за фра-

зы «рост заболеваемости» вместо «рост PLT» — это искажает коммуникативную цель: читатель может подумать, что речь идёт об эпидемиологической заболеваемости, а не об уровне тромбоцитов. Лингвист же не посчитал данную замену грубой ошибкой, поскольку грамматически фраза корректна.

Вывод: пример показывает, что текст может быть терминологически точным, но прагматически неадекватным для целевой аудитории, что подтверждает независимость измерений «прагматика» и «терминолог».

Тема: восстановительное правосудие, работа YOTs (Youth Offending Teams).

Агент	**Оценка**
Лингвист (linguist)	7
Семантик (semantic)	9,75
Терминолог (terminologist)	8,75
Прагматик (pragmatist)	8
Культурный адаптер (cultural_adaptor)	7
Юридический эксперт (legal_expert)	7,5
Итог модератора	**7,0**

Суть расхождения: семантик оценил перевод на 9,75, посчитав, что основной смысл передан верно. Лингвист и культурный адаптер, напротив, указали на некорректное использование аббревиатуры YOT в переводе и недостаточную адаптацию терминологии для русскоязычного читателя, что значительно снизило их оценки. Юридический эксперт подтвердил наличие критических ошибок в интерпретации термина.

Вывод: пример иллюстрирует, как смысловая полнота может сосуществовать с лингвистическими и культурными недостатками, которые заметны только при профильном анализе.

Анализ лучших и худших переводов

Наивысшие оценки (9,7) получили переводы DeepL (медицина) и Google (медицина). Худшие оценки (5,0) зафиксированы у Yandex в медицинских и юридических текстах, что указывает на системные ошибки. Пример ошибки Yandex Translate в юридических текстах (сегменты с итоговой оценкой 5,0):

Сегмент 23 (legal), дело Boring v. Google

Исходный текст: «Google entered the Borings' property to photograph it for inclusion in "Street View"… The Borings made a claim in tort for both invasion of privacy and trespass. … However, the court did find Google liable for what was a clear case of trespass given Google's physical entry on the property, and Google paid nominal damages of $1».

Перевод Yandex: «Google зашел на сайт компании Borings, чтобы сфотографировать её для размещения в разделе "Просмотр улиц"… Боринги подали иск о нарушении гражданских прав как за вторжение в частную жизнь, так и за незаконное проникновение. … Однако суд признал Google ответственной за то, что было явным нарушением авторских прав, учитывая физическое проникновение Google на территорию компании, и Google выплатила номинальный ущерб в размере 1 доллара».

Характер ошибок:

«property» переведено как «сайт компании» вместо «территория (собственность)»;

«trespass» в последнем предложении передан как «нарушение авторских прав» (должно быть «незаконное проникновение»).

Эти ошибки зафиксированы экспертами-агентами (см. полный JSON-файл для segment_id=23, engine=yandex): лингвист указал на «некорректный перевод 'Google entered the Borings' property' как 'Google зашел на сайт компании Borings'» и «неправильное использование термина 'нарушение авторских прав' вместо 'незаконное проникновение'». Именно подобные искажения привели к оценке 5,0 от модератора.

Сегмент 8 (medical)

Исходный текст: «… reducing folate inadequacy among WRA and children …»

Перевод Yandex: «… снижение дефицита фолиевой кислоты среди лиц, принимающих ЗОЖ, и детей …» (позже появилось «ЗПР»).

Характер ошибок:

Аббревиатура WRA (women of reproductive age) переведена как «лица, принимающие ЗОЖ» (здоровый образ жизни) и «ЗПР», что полностью искажает смысл. Медицинский эксперт в JSON-оценке указал: «лиц, принимающих ЗОЖ (правильно: женщин репродуктивного возраста)». Оценка — 5,0.

Обсуждение

Интерпретация основных результатов

Полученные результаты позволяют составить детальный портрет каждого движка и его сильных/слабых сторон в разных предметных областях.

DeepL предстает как универсальный специалист. Общее лидерство DeepL (8,57±0,69) подтверждает его репутацию, сформированную в предыдущих исследованиях (Ahrenberg, 2017; Macketanz et al., 2022). Особенно впечатляет результат в юридической сфере (8,47) — отрыв от Google (+0,32) статистически значим ($p < 0,01$). Анализ оценок агентов показывает, что DeepL получает высокие баллы от терминолога и прагматика, что свидетельствует о точности передачи юридических понятий и сохранении коммуникативной цели текстов. Однако неожиданным оказалось второе место в медицине вопреки маркетинговым заявлениям о специализации.

Google Translate стал лидером в медицине. Наивысший результат Google в медицинской тематике (8,80) требует объяснения. Мы предполагаем, что это связано с объёмом обучающих данных: Google обрабатывает огромные массивы медицинской литературы, включая PubMed Central и клинические рекомендации. Высокие оценки лингвиста и семантика (корреляция 0,82) указывают на то, что Google не только точно передаёт термины, но и сохраняет связность и естественность медицинского дискурса. В публицистике Google также лидирует (8,59), что согласуется с его ориентацией на новостные и разговорные тексты (Popović, 2020).

Yandex Translate нестабильно переводит в специальных областях. Результаты Yandex требуют отдельного анализа. Общая оценка 7,91±1,06 — самая низкая среди движков, причём с максимальной вариативностью. Особенно тревожен результат в юриспруденции (7,51) с тремя переводами, получившими оценку 5,0. Анализ этих случаев показывает системные ошибки в передаче модальных глаголов (“shall” переведён как «должен» вместо «обязан») и юридических терминов (“liability” как «ответственность» вместо «обязательство»). Интересно, что в публицистике Yandex приближается к конкурентам (8,02). Однако для специальных областей, требующих терминологической точности, Yandex пока уступает.

Сравнение с предыдущими исследованиями

Наши результаты частично согласуются с выводами Macketanz et al. (2022), которые также зафиксировали превосходство DeepL

для немецко-английской пары. Однако в их исследовании DeepL доминировал во всех тематиках, тогда как мы обнаружили его уязвимость в медицинской сфере. Возможно, это связано с направлением перевода (англо-русский против немецко-английского) и особенностями русской медицинской терминологии.

Исследование Popović (2020) показало, что Google лучше справляется с переводами с английского на языки со сложной морфологией. Наши данные подтверждают этот вывод: русский язык с его богатой морфологией действительно лучше обрабатывается Google (особенно в медицине), чем DeepL, ориентированный на аналитические языки.

Объяснение выявленных закономерностей

Мы предполагаем, что DeepL оптимизирован под общелитературный и деловой перевод, тогда как его обучающие данные по медицине могут быть ограничены. Кроме того, медицинский перевод требует не только терминологической точности, но и соблюдения определённых жанровых конвенций (структура аннотации, клишированные фразы), в которых Google, обученный на миллионах научных статей, имеет преимущество.

Анализ конкретных ошибок Yandex Translate показывает, что проблема не столько в терминах, сколько в модальности и синтаксисе юридических конструкций. Русская юридическая традиция тяготеет к безличным и пассивным конструкциям, тогда как в английском праве чаще используется активный залог и модальные глаголы. Yandex, по-видимому, недостаточно обучен на параллельных юридических корпусах, учитывающих эти различия.

Высокая вариативность Yandex ($\sigma = 1{,}06$) может указывать на нестабильность работы модели или на чувствительность к конкретным типам текстов. Это требует дополнительного исследования.

Анализ работы межагентной системы

Предложенная архитектура из шести агентов показала свою состоятельность. Высокая корреляция между лингвистом и семантиком ($r = 0{,}82$) ожидаема, поскольку эти аспекты часто взаимосвязаны. Интереснее низкая корреляция прагматика с терминологом ($r = 0{,}31$), что подтверждает: текст может быть терминологически точен, но неэффективен коммуникативно, и наоборот. Это оправдывает многомерный подход.

Особого внимания заслуживает работа модератора. Анализ случаев с частичной согласованностью экспертов (72% оценок) по-

казывает, что модератор успешно находил компромисс, усредняя оценки при расхождениях и выбирая наиболее авторитетного эксперта в спорных случаях (например, отдавая приоритет медицинскому эксперту при расхождении с лингвистом).

Проблема отказов модерации (12 текстов, 6,7%) требует обсуждения. Все отказанные тексты относились к публицистике и, по-видимому, содержали темы, вызвавшие срабатывание встроенных механизмов безопасности YandexGPT. Это ограничение самой модели, а не нашего метода, и должно учитываться при интерпретации результатов.

Ограничения исследования

1. Объём выборки: 60 текстов — достаточный, но не исчерпывающий корпус. Увеличение выборки могло бы выявить более тонкие различия.

2. Отказы модерации: 12 исключённых текстов могли повлиять на оценки по публицистике (средняя 8,38, но с меньшей статистической мощностью).

В json-файле полных оценок прямо зафиксированы случаи, когда все агенты, включая модератора, отвечали фразой: «Я не могу обсуждать эту тему. Давайте поговорим о чём-нибудь ещё». Это произошло в сегментах 46, 49, 53, 58. В некоторых других публицистических сегментах часть агентов выдавала отказ, а часть — нет (например, сегмент 41, где семантик отказался, а остальные агенты отработали). Именно эти отказы и привели к тому, что 12 итоговых оценок для публицистических текстов отсутствуют.

Основываясь на содержании отрывков, которые YandexGPT отказалась оценивать, можно сделать следующие предположения о причинах отказов. Все четыре отрывка объединяет то, что они затрагивают острые общественно-политические темы с явным вовлечением конкретных государств, политических фигур или военных конфликтов, что, вероятно, активирует встроенные механизмы безопасности YandexGPT, запрещающие модели участвовать в анализе или комментировании определённого контента. Это подтверждает предположение о непригодности YandexGPT в текущей конфигурации для оценки текстов подобного рода и объясняет значительное смещение выборки в сторону нейтральных публицистических материалов.

Результаты по публицистике обладают пониженной надёжностью и не должны напрямую сравниваться с медициной и юриспруденцией.

3. Одна целевая пара языков: результаты могут не переноситься на другие языки.

4. Версионность систем: все движки постоянно обновляются; результаты отражают состояние на февраль 2026 года.

5. Субъективность агентов: хотя агенты обучены быть объективными, они остаются LLM с присущими им смещениями.

Направления будущих исследований

1. Расширение корпуса: включение большего числа текстов и новых тематик (техническая документация, маркетинговые материалы).

2. Сравнение с человеческими экспертами: валидация оценок агентов против оценок профессиональных переводчиков.

3. Анализ ошибок на уровне предложений: более детальная классификация типов ошибок.

4. Диахроническое исследование: отслеживание изменения качества движков во времени.

5. Применение метода к другим языковым парам: например, русско-китайской или русско-арабской.

Заключение

В настоящем исследовании проведён сравнительный анализ качества машинного перевода трёх ведущих систем — Google Translate, Yandex Translate и DeepL — на материале 60 текстов трёх предметных областей (медицина, юриспруденция, публицистика) с использованием оригинальной межагентной системы оценки на базе YandexGPT.

Ключевые результаты:

1. DeepL показал наилучший общий результат (8,57±0,69), подтвердив свою репутацию лидера машинного перевода. Особенно силён DeepL в юридической тематике (8,47), где его преимущество над Google статистически значимо ($p < 0,01$).

2. Google Translate неожиданно оказался лучшим в медицинской сфере (8,80) и публицистике (8,59). Это опровергает распространённое мнение о безусловном лидерстве DeepL во всех областях и указывает на важность учёта предметной специфики при выборе системы перевода.

3. Yandex Translate уступил конкурентам во всех тематиках, продемонстрировав наибольшую вариативность (7,91±1,06). Особенно проблемной областью оказалась юриспруденция (7,51), где зафиксированы системные ошибки в передаче модальности и юри-

дической терминологии. В публицистике Yandex приближается к конкурентам (8,02), что согласуется с его ориентацией на общелитературный язык.

4. Выявлены статистически значимые различия между движками (ANOVA, $p < 0,05$), подтверждающие, что выбор системы перевода существенно влияет на качество результата, особенно в специальных областях.

5. Предложенная межагентная система продемонстрировала свою эффективность: корреляционный анализ подтвердил ожидаемые связи между агентами (лингвист-семантик, $r = 0,82$) и независимость различных аспектов оценки (прагматик-терминолог, $r = 0,31$), что оправдывает многомерный подход.

Научная новизна

1. Впервые применён межагентный подход с использованием российской языковой модели (YandexGPT) для многокритериальной оценки качества перевода.

2. Разработана и валидирована архитектура из шести специализированных агентов, обеспечивающая оценку по различным аспектам (лингвистическому, семантическому, терминологическому, прагматическому, культурному, тематическому).

3. Впервые проведено систематическое сравнение трёх ведущих движков на едином материале трёх предметных областей для русскоязычного перевода.

4. Выявлены неожиданные закономерности: превосходство Google в медицинском переводе и критические слабости Yandex в юридической сфере.

Практические рекомендации

Для переводчиков и редакторов:

- При работе с медицинскими текстами предпочтительнее использовать Google Translate (средняя оценка 8,80), но с обязательным постредактированием терминологии.
- Для юридических документов оптимален DeepL (8,47), обеспечивающий наилучшую точность терминологии и модальности.
- При переводе публицистики можно использовать любой из лидеров (Google 8,59, DeepL 8,54), ориентируясь на дополнительные критерии (скорость, интеграция, стоимость).
- Yandex Translate рекомендуется применять с осторожностью для специальных текстов, ограничиваясь общелитератур-

ными и разговорными жанрами, где его качество приближается к конкурентам.

Для разработчиков систем МП:

- DeepL: усилить медицинское направление, расширив обучающие данные за счёт специализированных корпусов (клинические рекомендации, аннотации к препаратам).
- Google: сохранять и развивать преимущество в медицинской сфере, работая над улучшением юридического направления.
- Yandex: критически важно улучшить качество юридического перевода (модальность, терминология) и снизить вариативность результатов.

Для заказчиков перевода:

- При выборе подрядчика учитывать не только общее качество, но и специализацию движков.
- Требовать указания, какая система МП использовалась, и предусматривать бюджет на постредактирование.

Рекомендации для образования

Результаты исследования могут быть использованы в курсах по переводоведению, компьютерной лингвистике и обработке естественного языка:

1. Введение в машинный перевод: демонстрация сильных и слабых сторон современных систем.
2. Практикум по постредактированию: анализ типичных ошибок разных движков в специальных областях.
3. Спецкурсы по медицинскому и юридическому переводу: учёт особенностей автоматического перевода при работе с терминологией.

Проведённое исследование демонстрирует, что выбор системы машинного перевода не может быть универсальным — он должен определяться предметной областью и конкретными задачами. DeepL, Google Translate и Yandex Translate обладают различными сильными сторонами, и понимание этих различий критически важно для эффективного использования технологий автоматического перевода в профессиональной деятельности.

Предложенный межагентный подход открывает новые возможности для многомерной, объяснимой и настраиваемой оценки качества перевода. В отличие от «чёрного ящика» традиционных метрик или одного LLM-судьи, наша система позволяет понять,

почему перевод получил ту или иную оценку и какие именно аспекты требуют улучшения. Это приближает нас к созданию по-настоящему интеллектуальных инструментов оценки и редактирования машинного перевода.

Список литературы

Гарбовский Н.К. Теория перевода: учебник. М.: Издательство Московского университета, 2019. 543 с.

Комиссаров В.Н. Современное переводоведение. М.: Р. Валент, 2018. 408 с.

Сдобников В.В. Оценка качества перевода в эпоху нейронного машинного перевода // Вестник Нижегородского университета им. Н.И. Лобачевского, 2020. № 6. С. 123–130.

Сдобников В.В., Петрова О.В. Теория перевода: коммуникативно-функциональный подход. М.: ФЛИНТА, 2019. 512 с.

Ahrenberg L. (2017) Comparing machine translation systems: DeepL, Google Translate and Microsoft Translator. Linköping University Electronic Press. URL: http://www.diva-portal.org/smash/record.jsf?pid=diva2:1152119

Bani S. (2006) An analysis of the translation of political texts. Target. Vol. 18. No. 2, pp. 297–321. DOI: 10.1075/target.18.2.06ban

Biel Ł. (2017) Researching legal terminology. The Oxford Handbook of Language and Law. Ed. by P.M. Tiersma, L.M. Solan. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199572120.013.20

Callison-Burch C., Osborne M., Koehn P. (2006) Re-evaluating the role of BLEU in machine translation research // Proceedings of EACL, pp. 249–256. URL: https://aclanthology.org/E06-1032/

Dew K.N. (2020) Medical translation and interpreting. The Routledge Handbook of Translation and Health. London: Routledge, pp. 45–60. DOI: 10.4324/9780815350839-4

Graham Y., Baldwin T., Moffat A. (2013) Continuous measurement scales in human evaluation of machine translation. Proceedings of the 7th Linguistic Annotation Workshop, pp. 33–41. URL: https://aclanthology.org/W13-2305/

Karwacka W. (2014) Quality assurance in medical translation // The Journal of Specialised Translation. No. 21, pp. 19–34. URL: https://www.jostrans.org/issue21/art_karwacka.php

Kocmi T., Federmann C. (2023) Large language models are state-of-the-art evaluators of translation quality // arXiv preprint arXiv:2302.14520. DOI: 10.48550/arXiv.2302.14520

Koehn P. (2004) Statistical significance tests for machine translation evaluation. Proceedings of EMNLP, pp. 388–395. URL: https://aclanthology.org/W04-3250/

Macketanz V., Avramidis E., Burchardt A., Helcl J., Uszkoreit H. (2022) DeepL versus Google Translate: a comparative study // Proceedings of the

23rd Annual Conference of the European Association for Machine Translation (EAMT), pp. 125–134. URL: https://aclanthology.org/2022.eamt-1.18/

Mattila H.E. (2013) Comparative legal linguistics. Farnham: Ashgate Publishing. DOI: 10.4324/9781315093008

Montalt V., González-Davies M. (2014) Medical translation step by step. London: Routledge. DOI: 10.4324/9781315760580

Papineni K., Roukos S., Ward T., Zhu W.J. (2002) BLEU: a method for automatic evaluation of machine translation. Proceedings of ACL, pp. 311–318. DOI: 10.3115/1073083.1073135

Popović M. (2015) chrF: character n-gram F-score for automatic MT evaluation. Proceedings of WMT, pp. 392–395. URL: https://aclanthology.org/W15-3049/

Popović M. (2020) On the differences between human and automatic translations. Prague Bulletin of Mathematical Linguistics. Vol. 113. pp. 25–40. DOI: 10.2478/pralin-2020-0002

Šarčević S. (1997) New approach to legal translation. The Hague: Kluwer Law International.

Snover M., Dorr B., Schwartz R., Micciulla L., Makhoul J. (2006) A study of translation edit rate with targeted human annotation. Proceedings of AMTA 2006. 2006, pp. 223–231. URL: https://aclanthology.org/2006.amta-papers.16/

Valdeón R.A. (2015) The translation of journalistic texts. The Translator. Vol. 21, No. 3, pp. 241–258. DOI: 10.1080/13556509.2015.1060806

Vaswani A., et al. (2017) Attention is all you need. Advances in Neural Information Processing Systems 30, pp. 5998–6008. URL: https://papers.nips.cc/paper/7181-attention-is-all-you-need

Wooldridge M., Jennings N.R. (1995) Intelligent agents: theory and practice. The Knowledge Engineering Review. Vol. 10. No. 2. pp. 115–152. DOI: 10.1017/S0269888900008122

References

Ahrenberg L. (2017) Comparing machine translation systems: DeepL. Google Translate and Microsoft Translator. Linköping University Electronic Press. Available at: http://www.diva-portal.org/smash/record.jsf?pid=diva2:1152119

Bani S. (2006) An analysis of the translation of political texts. Target. Vol. 18. No. 2, pp. 297–321. DOI: 10.1075/target.18.2.06ban

Biel Ł. (2017) Researching legal terminology. In: Tiersma P.M., Solan L.M. (eds.) The Oxford Handbook of Language and Law. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199572120.013.20

Callison-Burch C., Osborne M., Koehn P. (2006) Re-evaluating the role of BLEU in machine translation research. Proceedings of EACL, pp. 249–256. Available at: https://aclanthology.org/E06-1032/

Dew K.N. (2020) Medical translation and interpreting. In: The Routledge Handbook of Translation and Health. London: Routledge, pp. 45–60. DOI: 10.4324/9780815350839-4

Garbovskiy N.K. (2019) Teoriya perevoda: uchebnik = Theory of Translation: Textbook. Moscow: Moscow University Press, 543 p. (In Russian).

Graham Y., Baldwin T., Moffat A. (2013) Continuous measurement scales in human evaluation of machine translation. Proceedings of the 7th Linguistic Annotation Workshop, pp. 33–41. Available at: https://aclanthology.org/W13-2305/

Komissarov V.N. (2018) Sovremennoe perevodovedenie = Modern Translation Studies. Moscow: R.Valent, 408 p. (In Russian).

Karwacka W. (2014) Quality assurance in medical translation. The Journal of Specialised Translation. No. 21, pp. 19–34. Available at: https://www.jostrans.org/issue21/art_karwacka.php

Kocmi T., Federmann C. (2023) Large language models are state-of-the-art evaluators of translation quality. arXiv preprint arXiv:2302.14520. DOI: 10.48550/arXiv.2302.14520

Koehn P. (2004) Statistical significance tests for machine translation evaluation. Proceedings of EMNLP, pp. 388–395. Available at: https://aclanthology.org/W04-3250/

Macketanz V., Avramidis E., Burchardt A., Helcl J., Uszkoreit H. (2022) DeepL versus Google Translate: a comparative study. Proceedings of the 23rd Annual Conference of the European Association for Machine Translation (EAMT), pp. 125–134. Available at: https://aclanthology.org/2022.eamt-1.18/

Mattila H.E. (2013) Comparative legal linguistics. Farnham: Ashgate Publishing. DOI: 10.4324/9781315093008

Montalt V., González-Davies M. (2014) Medical translation step by step. London: Routledge. DOI: 10.4324/9781315760580

Papineni K., Roukos S., Ward T., Zhu W.J. (2002) BLEU: a method for automatic evaluation of machine translation. Proceedings of ACL, pp. 311–318. DOI: 10.3115/1073083.1073135

Popović M. (2015) chrF: character n-gram F-score for automatic MT evaluation. Proceedings of WMT, pp. 392–395. Available at: https://aclanthology.org/W15-3049/

Popović M. (2020) On the differences between human and automatic translations. Prague Bulletin of Mathematical Linguistics. Vol. 113, pp. 25–40. DOI: 10.2478/pralin-2020-0002

Šarčević S. (1997) New approach to legal translation. The Hague: Kluwer Law International.

Sdobnikov V.V. (2020) Otsenka kachestva perevoda v epokhu neyronnogo mashinnogo perevoda = Translation quality assessment in the era of neural machine translation. *Vestnik Nizhegorodskogo universiteta im. N.I. Lobachevskogo*. No. 6, pp. 123–130. (In Russian).

Sdobnikov V.V., Petrova O.V. (2019) Teoriya perevoda: kommunikativno-funktsional'nyy podkhod = Theory of Translation: A Communicative-Functional Approach. Moscow: FLINTA, 512 p. (In Russian).

Snover M., Dorr B., Schwartz R., Micciulla L., Makhoul J. (2006) A study of translation edit rate with targeted human annotation. Proceedings of AMTA, pp. 223–231. Available at: https://aclanthology.org/2006.amta-papers.16/

Valdeón R.A. (2015) The translation of journalistic texts. The Translator. Vol. 21. No. 3, pp. 241–258. DOI: 10.1080/13556509.2015.1060806

Vaswani A., et al. (2017) Attention is all you need. Advances in Neural Information Processing Systems 30, pp. 5998–6008. Available at: https://papers.nips.cc/paper/7181-attention-is-all-you-need

Wooldridge M., Jennings N.R. (1995) Intelligent agents: theory and practice. The Knowledge Engineering Review. Vol. 10. No. 2, pp. 115–152. DOI: 10.1017/S0269888900008122

ИНФОРМАЦИЯ ОБ АВТОРАХ:

Савостьянов Виктор Олегович — старший преподаватель кафедры «Перевод и переводоведение», ПГУ, ул. Красная 40, 440026, Пенза. Руководитель студенческого научного кружка «Искусственный интеллект в межъязыковой коммуникации»; victsav@gmail.com

Жидков Максимилиан Вячеславович — студент ПГУ, Пенза, ул. Красная 40, 440026. Член студенческого научного кружка «Искусственный интеллект в межъязыковой коммуникации».

Инякина Дарья Романовна — студент ПГУ, ул. Красная 40, 440026, Пенза. Член студенческого научного кружка «Искусственный интеллект в межъязыковой коммуникации».

Чугункина Алина Васильевна — студент ПГУ, ул. Красная 40, 440026, Пенза. Член студенческого научного кружка «Искусственный интеллект в межъязыковой коммуникации».

ABOUT THE AUTHORS:

Viktor O. Savostyanov — Senior Lecturer at the Department of Translation and Translation Studies, Penza State University (PSU); 40, Krasnaya St., Penza 440026, Russia; Head of the Student Research Group "Artificial Intelligence in Interlingual Communication"; victsav@gmail.com

Maksimilian V. Zhidkov — Undergraduate Student at Penza State University (PSU); 40, Krasnaya St., Penza, 440026 Russia; Member of the Student Research Group "Artificial Intelligence in Interlingual Communication".

Daria R. Inyakina — Undergraduate Student at Penza State University (PSU); 40, Krasnaya St., Penza 440026, Russia; Member of the Student Research Group "Artificial Intelligence in Interlingual Communication".

Alina V. Chugunkina — Undergraduate Student at Penza State University (PSU); 40, Krasnaya St., Penza 440026, Russia; Member of the Student Research Group "Artificial Intelligence in Interlingual Communication".

Вклад авторов: авторы внесли равноценный вклад в подготовку публикации.

Contribution of the authors: the authors contributed equally to this article.

Конфликт интересов: положения и точки зрения, представленные в данной статье, принадлежат авторам и не обязательно отражают позицию какой-либо организации или российского научного сообщества. Авторы заявляют об отсутствии конфликта интересов.

Conflict of interests: the ideas and opinions presented in this article entirely belong to the authors and do not necessarily reflect the position of any organization or the Russian scientific community as a whole. The authors state that there is no conflict of interests.