

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра романо-германской филологии
и переводоведения

**Оценка качества машинного перевода научных аннотаций:
использование автоматических метрик и лингвистического
анализа**

АВТОРЕФЕРАТ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ БАКАЛАВРА

студента 4 курса 421 группы
направления 45.03.01 «Филология»
профиля «Зарубежная филология»
(Английский и немецкий / французский языки;
Немецкий и английский языки)
Института филологии и журналистики

Глинской Ольги Сергеевны

Научный руководитель
к.ф.н., доцент

Харламова Т.В.

Зав. кафедрой
к.ф.н., доцент

Харламова Т.В.

Саратов 2026

Введение. Стремительное развитие технологий машинного перевода (МП) ставит перед исследователями задачу объективной оценки качества получаемых текстов. Научные аннотации выполняют ключевую роль в международной коммуникации, обеспечивая оперативный доступ к результатам исследований. Однако даже современные системы МП не всегда гарантируют точную передачу содержания, соблюдение терминологии и стилистических особенностей научного стиля. В связи с этим возникает необходимость в комплексном подходе к оценке качества перевода, который объединял бы количественные методы автоматического анализа и качественную лингвистическую интерпретацию.

Актуальность данной темы обусловлена постоянным увеличением объёма научной информации, требующей распространения среди международной аудитории, а также отсутствием унифицированного подхода к оценке машинного перевода научных аннотаций.

Научная новизна работы состоит в сопоставительном анализе качества перевода научных аннотаций, выполненного несколькими современными системами МП, с использованием как автоматических метрик, так и лингвистического анализа. В исследовании выявляется корреляция между результатами автоматической оценки и реальными языковыми характеристиками текста, что позволяет уточнить возможности и ограничения существующих методов оценки качества МП.

Цель данной работы – проведение комплексного анализа качества машинного перевода научных аннотаций с использованием автоматических метрик и лингвистического анализа. Для достижения этой цели будут решены следующие задачи:

1. изучить эволюцию подходов к оценке качества перевода и уточнить содержание категорий адекватности и эквивалентности;
2. проанализировать теоретические основы переводческих соответствий и трансформаций, предложенные Я.И. Рецкером;

3. рассмотреть принципы работы автоматических метрик оценки качества перевода, включая BERTScore и METEOR;
4. провести эмпирическое исследование переводов научных аннотаций с использованием современных систем МП и выявить корреляцию между автоматическими оценками и лингвистическим качеством текста.

Предметом данного исследования является качество машинного перевода научных аннотаций, оценённое с помощью автоматических метрик и лингвистического анализа.

Объектом исследования являются научные аннотации как специфический жанр научного текста, характеризующийся высокой степенью сжатости, терминологической насыщенности и строгими стилистическими требованиями.

Материалом для исследования послужили 150 аннотаций на русском языке, отобранные из научных публикаций, размещённых в электронной библиотеке eLibrary.ru, а также их авторские англоязычные версии, выступающие в качестве референтных переводов. С помощью трёх систем МП, основанных на нейросетевых технологиях (Yandex, Google, DeepSeek), были получены переводы с русского на английский язык. Всего было сформировано 450 текстов для последующего анализа.

В работе применялись следующие **методы** исследования: описательный метод, метод сопоставительного анализа, метод количественной обработки данных, метод автоматической оценки качества перевода с применением метрик BERTScore и METEOR, а также метод лингвистической интерпретации полученных результатов, основанный на теории соответствий и трансформаций.

Структура работы. Работа состоит из введения, двух глав, заключения и списка использованных источников.

Основное содержание работы. В первой главе исследуются теоретические основы изучения качества перевода. Вопросы теории и практики перевода относятся к числу актуальных направлений

лингвистической мысли, что подтверждается неизменным интересом к ним на протяжении более двух с половиной тысячелетий. В XX веке перевод вышел за её пределы, став вопросом международной безопасности и технологического прогресса, что положило начало исследованиям в области машинного перевода. Сегодня выделяют три основных типа систем МП: системы прямого перевода, трансферные системы и системы семантического перевода, однако любой машинный перевод требует последующего вмешательства человека. Рынок машинного перевода демонстрирует устойчивую положительную динамику, что обусловлено глобализацией, развитием электронной коммерции и достижениями в области искусственного интеллекта. От античных принципов к автоматизированным системам, развитие теории перевода и технологический прогресс привели к тому, что вопрос о качестве перевода стал ещё более актуальным.

Во втором разделе первой главы изучаются категории качества перевода – адекватность и эквивалентность. Основной проблемой оценки качества перевода выступает проблема выделения параметров этой оценки, а объём и содержание терминов «адекватность» и «эквивалентность» не определены. Использование терминов «эквивалентный» и «адекватный» для оценки качества перевода затруднено из-за множественных трактовок, поэтому необходим иной, более операциональный понятийный аппарат. Таким аппаратом в данной работе выступает теория переводческих соответствий и трансформаций.

В третьем разделе первой главы рассматриваются переводческие соответствия и трансформации как средство достижения адекватности и эквивалентности. Теория закономерных соответствий, представленная в 1950 г. Я.И. Рецкером, стала одной из первых лингвистических теорий в отечественном переводоведении. Автор выделил три группы соответствий: эквиваленты, аналоги (вариативные соответствия) и адекватные замены. В работе анализируются различные определения переводческих трансформаций – от Я.И. Рецкера («приёмы логического мышления») до И.С. Алексеевой

(«межъязыковые преобразования, требующие перестройки на лексическом, грамматическом или текстовом уровне»). Несмотря на множественность подходов к определению переводческих трансформаций, можно выделить их межъязыковую природу и направленность на достижение эквивалентности при переводе. В качестве рабочей классификации принимается типология И.С. Алексеевой (перестановки, замены, добавления, опущения, антонимический перевод, компенсация, описательный перевод).

В четвёртом разделе первой главы исследуется экспертная оценка качества перевода. Описываются ранние методики ALPAC, шкала М. Нагао, типология ошибок DQF–MQM и подход Д.М. Бузаджи, выделяющий 4 основные группы ошибок с весовыми коэффициентами. Субъективность эксперта, размытость границ критериев и сложность учёта всех аспектов текста сохраняют ограничения ручной оценки. В связи с этим всё более востребованным становится альтернативный подход, позволяющий минимизировать влияние человеческого фактора. Данный подход реализуется в методах автоматической оценки качества перевода.

В пятом разделе первой главы рассматривается автоматическая оценка качества перевода. Выделяют традиционные метрики, основанные на методе n-грамм (BLEU, ROUGE, METEOR и др), нейросетевые метрики, основанные на эмбедингах (BERTScore, COMET), а также безреференсные метрики (COMET-Kiwi, MATESE-QE, REUSE). Выбор метода автоматической оценки зависит от целей исследования: для быстрого бенчмаркинга систем возможно применение BLEU или ROUGE, тогда как для более точного анализа качества перевода, приближенного к человеческому восприятию, предпочтительнее использование семантических метрик (COMET, BERTScore).

Во второй главе проводится эмпирическое исследование переводов научных аннотаций.

В первом разделе второй главы описывается методология исследования.

Далее рассматривается научная аннотация как объект перевода, её жанровые и стилистические особенности. В системе научной коммуникации

аннотация занимает особое место как жанр, сопровождающий основную публикацию. Успешный перевод аннотации предполагает не просто передачу содержания, но и воспроизведение структурно-стилистических характеристик, присущих целевому научному дискурсу.

В третьем разделе второй главы исследуется термин и способы его перевода. Согласно определению К.А. Мякшина, термин – это «общепринятое частотное слово (словесный комплекс), характеризующееся единством звукового облика и соотнесенного с ним соответствующего понятия в системе понятий данной области знания и деятельности». Существуют различные способы перевода терминов: транслитерация, калькирование и транскрибирование, перестановки, замены, добавления и опущения, описательный приём. Анализ теоретических подходов к определению термина и способов его перевода позволяет заключить, что терминологическая точность является одним из ключевых факторов достижения эквивалентности при переводе научных текстов.

В четвёртом разделе представлен сопоставительный лингвистический анализ переводов, получивших наиболее высокие баллы автоматических метрик. Данный анализ показал, что высокие значения BERTScore и METEOR не всегда полностью совпадают с реальным качеством текста. При переводе аннотации о нейросетях все системы корректно передали устоявшиеся термины («*machine translation systems*», «*neural networks*»). Однако в переводе DeepSeek были обнаружены неточности: вместо «*for quite some time*» использовано «*long*» с сохранением русского порядка слов, фраза «*значительно упрощая жизнь*» переведена калькой «*greatly simplifying life*» (вместо идиоматичного «*making life much easier*»), а временная форма «*улучшилось*» передана Past Simple («*improved*») вместо Present Perfect, как у автора и Google. Несмотря на эти стилистические и грамматические шероховатости, метрики оценили перевод DeepSeek выше, чем вариант Google, что противоречит лингвистическому анализу.

Другие примеры подтвердили, что автоматические метрики чувствительны прежде всего к формальному лексико-семантическому сходству с референтным текстом, но «слепы» к контексту и терминологической точности. В одной аннотации Google перевёл «задачи» как «*problems*» вместо уместного «*tasks*», что снизило качество, однако оценки остались высокими. В другом случае авторский референтный перевод содержал терминологические ошибки («*corpuses*» вместо научного «*corpora*», «*trees of syntactic analysis*» вместо «*parse trees*»), тогда как системы машинного перевода предложили более корректные варианты.

В пятом разделе второй главы представлен анализ переводов, получивших низкие оценки. Часто причиной снижения оценок становилась не грубая ошибка перевода, а терминологическая сложность текста или расхождение с референтным переводом. Например, при переводе термина «индекс TQI» автор использовал «*quality assurance*», что неверно, поскольку TQI — это инструмент оценки, а не процесс обеспечения качества. Машинные системы оказались точнее, но их оценки всё равно были низкими из-за несовпадения с референсом. В другом случае при переводе «актанты и циркуманты глагола» все системы МП применили калькирование («*actants and circumstants*»), тогда как в референсе использовался устоявшийся термин «*complements and adjuncts*». Это также привело к резкому снижению баллов.

Примеры подтвердили, что оценки метрик оказываются неверными, если в переводах используются комплексные трансформации. В авторских переводах применялись такие трансформации как членение сложного предложения и синтаксическая перестройка, что обусловлено требованиями английского научного стиля, однако системы МП сохраняли при переводе на английский язык русскую структуру предложения. Несмотря на то, что смысл был передан верно, низкие оценки метрик отражали лишь формальное несовпадение с референсом, а не реальное качество перевода. В случае с терминами контаминации фразеологизмов («перекрещивание», «присоединение», «наложение») выяснилось, что устоявшиеся английские

эквиваленты отсутствуют в словарях; каждый переводчик (человек или машина) вынужден конструировать термин заново, что порождает естественную вариативность (*crossing/intersection, joining/adjunction, superimposition/overlay*). Метрики интерпретируют эту вариативность как ошибку, хотя объективно правильного варианта не существует.

Современные автоматические метрики (BERTScore и METEOR) демонстрируют определенную корреляцию с качеством перевода, однако их оценки не всегда адекватно отражают лингвистические особенности и стилистические нюансы переводимых текстов. Лингвистический анализ подтвердил критическую зависимость результатов от качества эталонных переводов: системы МП (Google, Yandex, DeepSeek) зачастую предлагали более точные или более идиоматичные варианты, чем человек. Автоматические метрики оказались «слепы» к нарушению актуального членения, идиоматичности, грамматической когерентности и учёту контекста. Наиболее объективные результаты достигаются при комбинировании методов: количественной оценки (автоматические метрики) для первичного ранжирования систем; качественного лингвистического анализа для проверки результатов и выявления зон нечувствительности метрик; критического анализа датасета (качества референсов).

В заключении подводятся итоги проведённого исследования. Теоретическая часть показала ограниченность категорий адекватности и эквивалентности для эмпирических исследований, что обусловило выбор теории переводческих соответствий и трансформаций в качестве основного инструмента. Практически установлено, что качество перевода определяется не только точностью передачи содержания, но и соблюдением жанрово-стилистических норм языка перевода и корректностью терминологии. Автоматические метрики часто демонстрируют высокие оценки при наличии стилистических и грамматических отклонений, что свидетельствует об их недостаточной чувствительности к данным аспектам.

Среди ограничений исследования – ограниченная выборка, рассмотрение только одного жанра (научная аннотация) и неполный охват автоматических метрик. Практическое значение работы состоит в возможности использования результатов для совершенствования систем оценки качества перевода. Перспективы дальнейших исследований связаны с разработкой метрик, чувствительных к стилистическим особенностям, и созданием специализированных терминологических баз.