

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

Кафедра теории, истории языка и прикладной лингвистики

**Проблемы автоматического перевода научных текстов**

АВТОРЕФЕРАТ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ МАГИСТРА

студентки 2 курса 252 группы  
направления 45.04.01 – «Филология»  
профиля «Теория языка и современные лингвистические технологии»  
Института филологии и журналистики

Макаровой Анастасии Алексеевны

Научный руководитель

к.ф.н., доцент

\_\_\_\_\_

А.И. Буранова

подпись, дата

Зав. кафедрой, д.ф.н., профессор

\_\_\_\_\_

О. Ю. Крючкова

подпись, дата

Саратов 2025

Современное развитие науки и технологий сопровождается появлением большого количества научных публикаций на различных языках. Это создает значительные барьеры для международного обмена знаниями, особенно когда речь идет о сложных текстах, насыщенных терминологией и специализированной лексикой. Автоматический перевод научных текстов является одним из перспективных решений этой проблемы, однако его реализация связана с рядом трудностей.

Научные тексты обладают особыми характеристиками: строгая структура, высокая степень формализации, использование специфической терминологии, а также необходимость точной передачи смысла. Это предъявляет особые требования к качеству перевода, поскольку даже незначительные искажения могут повлиять на правильное понимание содержания.

**Актуальность исследования** обусловлена необходимостью изучения и анализа известных систем машинного перевода с целью выявления их достоинств и недостатков при работе с научными текстами.

**Объектом** нашего исследования являются переводы научных текстов, выполненные с помощью систем машинного перевода Google Translate, Яндекс Переводчик, DeepL Translate.

**Предметом** исследования стало качество машинного перевода научных текстов.

**Цель** данной работы – определение эффективности систем машинного перевода Google Translate, Яндекс Переводчик и DeepL Translate при переводе с английского языка на русский язык научных текстов из области лингвистики и информационных технологий.

Для реализации поставленной цели необходимо было решить следующие **задачи**:

- 1) произвести обзор литературы по теме данного исследования; рассмотреть понятие «перевод», виды и норму перевода, классификации перевода, особенности машинного перевода;

2) собрать материал исследования – переводы научных текстов из области лингвистики и информационных технологий;

3) проанализировать собранный материал с целью его дальнейшей оценки;

4) произвести оценку качества машинного перевода разных типов научных текстов;

5) выявить типы ошибок, допускаемых системами машинного перевода при работе с научными текстами;

4) определить, какая система машинного перевода справилась с переводом лучше остальных и является наиболее оптимальной программой для работы с научными текстами.

**Материалом** исследования послужили переводы на русский язык 18 научных публикаций: 9 статей из области лингвистики и 9 статей по информационным технологиям (общим объемом 31 466 словоупотреблений). Также в качестве материала исследования выступают переводы аннотаций к выбранным нами статьям и переводы автореферата диссертации Моктар Алия «Перевод и семантизация названий кинофильмов (на материале английского, французского и арабского языков)».

В работе использовались следующие **методы**: общенаучные методы, современный описательный метод.

**Структура** работы обусловлена поставленными задачами исследования. Магистерская работа состоит из введения, двух глав, заключения, списка использованной литературы, приложения.

**Основное содержание работы.** В первой главе «Теоретические основы исследования» данной работы нами были проанализированы научные труды, посвященные проблемам теории перевода, классификации переводческих ошибок и машинного перевода.

Исследователи отмечают двойственную природу понятия «перевод». Перевод представляет собой процесс и результат, передающий содержание с одного языка на другой с учетом языковых и культурных особенностей.

Ученые выделяют несколько видов перевода: устный последовательный перевод, синхронный, синхронизация видеотекста, перевод с листа, коммунальный перевод, письменный и машинный переводы.

В работах, посвященных особенностям перевода научных текстов и терминов, отмечается необходимость соблюдения точности, знания терминологии и понимания предметной области со стороны переводчика.

Под нормой перевода понимается соблюдение эквивалентности, стилистических и жанровых норм, а также учет прагматических требований и культурных различий. Переводческие ошибки разделяют по степени их влияния на содержание. Учеными делается акцент на важности точности и соответствия нормам языка перевода.

Мы изучили историю появления и развития МП. Фактически история МП началась с Джорджтаунского эксперимента в 1954 г., когда впервые был осуществлен машинный перевод текстов с русского языка на английский. В этом же году был проведен сходный эксперимент в СССР И.К. Бельской и Д.Ю. Пановым. Последующее десятилетие было отмечено повышенным интересом к МП, велись активные работы над разработкой разных программ, но в 1966 г. темпы работ в данной области заметно снизились. В 1970-е гг. внимание исследователей вновь вернулось к проблеме МП, активная разработка систем МП продолжается в наше время.

Исследователи предлагают следующую классификацию систем машинного перевода: системы на основе правил, на основе примеров, статистические системы, нейронный МП и гибридные системы.

Исследователи выделяют разную степень участия человека в процессе МП: предредактирование, интерредактирование, постредактирование и смешанное участие человека.

Существуют разные стратегии МП: пословный, трансфертный и интерлингвистический переводы.

**Во второй главе «Анализ качества машинного перевода научных текстов»** нашего исследования нами были проанализированы переводы

Google Translate, DeepL Translate и Яндекс Переводчика 18 научных статей из таких научных областей, как лингвистика и информационные технологии: 9 статей по лингвистике и 9 статей по информационным технологиям (общим объемом 31 466 словоупотреблений).

Единой системы оценивая качества машинного перевода на данный момент нет. Поэтому мы проанализировали несколько работ на данную тему и выбрали классификацию, предложенную И. А. Кириченко и Н. А. Сигачевой. Исследователи предлагают производить оценку содержательной идентичности текста (тождественность перевода содержанию оригинального текста), лексического аспекта перевода и правильности перевода профессиональных терминов, грамматических аспектов (орфографические, пунктуационные и др. ошибки), стилистической идентичности текста [Кириченко, Сигачева: 2020].

В процессе работы мы не считали за ошибки неточности перевода, которые были связаны с ошибками в оригинальном тексте.

В переводах выбранных нами СМП было выявлено 166 нарушений, среди которых больше половины составили лексические ошибки, вторую по численности группу составили ошибки содержательной идентичности, меньше всего было обнаружено грамматических ошибок, а нарушений научного стиля нами выявлено не было.

По критерию содержательной идентичности ошибки были обнаружены только в переводах Яндекс Переводчика. Из-за технических неточностей, некоторые предложения были переведены некорректно. Переводчик часто упустил некоторые слова, менял их местами, вставлял в текст лишние знаки препинания, а именно точки, из-за чего значительно страдало качество переводов, поэтому говорить об их полной тождественности оригинальным текстам мы не можем:

Оригинал	Перевод
<i>So, according to A.S. Gerd, for example, you can see the process of the scientific and</i>	<i>Итак, по словам А.С. Например, вы можете увидеть процесс формирования</i>

<i>technical terminology...</i>	<i>научно-технической терминологии...</i>
---------------------------------	---

Анализ лексических аспектов и правильности перевода профессиональных показал, что проблемой для всех программ является транслитерирование фамилий:

Оригинал	Перевод Google Translate
<i>Angrave, Charlwood, Kirkpatrick, Lawrence, and Stuart (2016)[1] referencing a study...</i>	<i>Angrave, Charlwood, Kirkpatrick, Lawrence u Stuart (2016)[1] ссылаясь на исследование...</i>
Оригинал	Перевод DeepL Translate
<i>Baesens, Winne, and Sels (2017)[4] talk about factors...</i>	<i>В работе Baesens, Winne u Sels (2017)[4] говорится о факторах...</i>
Оригинал	Перевод Яндекс Переводчика
<i>... and cited in the work of Kurylowicks and especially of Martinet.</i>	<i>...и процитированы в работах Куриловича и особенно Мартине.</i>

Во всех переводах были обнаружены непереуведенные слова или словосочетания, а также непереуведенные расшифровки аббревиатур:

Оригинал	Перевод Google Translate
<i>But his interesting book entitled <b>The Latin Cooks: vegetable nutrition, lexical study</b> is exemplary.</i>	<i>Но его интересная книга под названием <b>The Latin Cooks: Vegetable nutrition, lexical study</b> является образцовой.</i>
Оригинал	Перевод DeepL Translate
<i>By stating the numerous uses of <b>People Analytics</b>...</i>	<i>Указывая на многочисленные сферы применения <b>People Analytics</b>...</i>
Оригинал	Перевод Яндекс Переводчика
<i>This obviously increases the cost of managing data in <b>onpremisses</b> databases...</i>	<i>Очевидно, что это увеличивает стоимость управления данными в базах данных <b>onpremisses</b></i>

Также нами были выявлены неточности в переводе слов и словосочетаний, а также профессиональных терминов:

Оригинал	Перевод Google Translate
<i>Amongst 50 respondents, it was observed that 10% were <b>interns</b> or <b>trainees</b> in a company.</i>	<i>Среди 50 респондентов было отмечено, что 10% были <b>стажерами</b> или</i>

	<i>стажерами</i> (в первом случае в оригинале используется слово «interns» – тот, кто проходит практику, а во втором «trainees» – стажер, работающий в компании) в компании.
Оригинал	Перевод DeepL Translate
<i>For this reason, the creation of text corpora would not be approached with such enthusiasm...</i>	<i>По этой причине к созданию текстовых <b>корпораций</b> (корректный перевод: корпусов) не отнеслись бы с таким энтузиазмом...</i>
Оригинал	Перевод Яндекс Переводчика
<i>...conceptology is trying to unravel the substantive mystery of the concept through its primary meaning.</i>	<i>...<b>концептуология</b> (корректный перевод: концептология) пытается разгадать содержательную тайну понятия через его первичное значение.</i>

Переводы Google Translate отличаются большим количеством речевых ошибок, по сравнению с результатами DeepL Translate и Яндекс Переводчика:

Оригинал	Перевод
<i>Jie (2008) mentioned several important criticisms about interlanguage studies...</i>	<i>Jie (2008) <b>упомянул</b> (корректный перевод: сделал) несколько важных критических замечаний по поводу межъязыковых исследований...</i>

В результатах, полученных с помощью DeepL Translate, было обнаружено больше проблем с переводом аббревиатур, чем в переводах Google Translate и Яндекс Переводчика. Проблема заключается в том, что аббревиатуры, которые были ранее переведены, далее в тексте могли встретиться в исходном виде, то есть на английском языке, или же переведенными, но ошибочно:

Оригинал	Перевод
<i>The scope of the present article assumes a massive computer attack (MCA).../ ...MCA-</i>	<i>В рамках данной статьи под массовой компьютерной атакой</i>

<i>exposed BD can be divided into 3 subgroups...</i>	(МКА) понимается.../ ...подверженных МСА, можно разделить на 3 подгруппы...
--	---

Были обнаружены ошибки в грамматическом аспекте в переводах Google Translate и Яндекс Переводчика. Типичными являются ошибки при склонении фамилий:

Оригинал	Перевод Google Translate
<i>The preliminary date for the appearance of cognitive linguistics can be recognized as that of the publication by Y.S. Stepanov of a small book entitled Semiotics [Stepanov, 1971]...</i>	Предварительной датой появления когнитивной лингвистики можно считать публикацию <b>Ю.С. Степанов</b> (вместо «Ю.С. Степановым») <i>небольшой</i> книги «Семиотика»...
Оригинал	Перевод Яндекс Переводчика
<i>Following A.A. Potebnja, they argue that "abstract nouns...</i>	Следуя за <b>А.А. Потенбня</b> , они утверждают, что "абстрактные существительные...

Частыми были нарушения связи между подлежащим и сказуемым:

Оригинал	Перевод Google Translate
<i>This group included many Romance linguists...</i>	В эту группу <b>входило</b> (вместо «входили») <i>многие романские лингвисты...</i>
Оригинал	Перевод Яндекс Переводчика
<i>In 1951, M. Minsky and D. Edmonds created the first network computer based on a neural network...</i>	В 1951 году <b>М. Мински и Д. Эдмондс</b> <i>создал</i> первый сетевой компьютер на основе нейронной сети...

В переводах Google Translate были обнаружены пунктуационные ошибки:

Оригинал	Перевод
<i>The category of the Cause in formal linguistics serves as a category connecting all linguistic relations, as the basis of knowledge....</i>	Категория Причины в формальной лингвистике выступает как категория (пропущена запятая в начале причастного оборота) <u>связывающая</u> все языковые отношения, как основа знания...

Анализ стилистической идентичности переводов показал, что программы хорошо справились со стилистическими особенностями научных текстов.

Таким образом, все переводчики справились с поставленной задачей хорошо. Но с точки зрения критериев, по которым производился анализ, наименьшее количество ошибок было обнаружено в переводах Google Translate, на втором месте оказался DeepL Translate, наибольшее число ошибок было выявлено в переводах Яндекс Переводчика. Но в некоторых моментах DeepL Translate и Яндекс Переводчик показали лучшие варианты переводов некоторых фрагментов, то есть выбрали наиболее подходящие по контексту синонимы и конструкции в предложениях:

Оригинал	Google Translate	DeepL Translate
<i>The similarity of utilities and services lies in the fact that...</i>	<i>Сходство коммунальных услуг и услуг заключается в том, что...</i>	<i>Сходство коммунальных услуг и сервисов заключается в том, что...</i>
Оригинал	Google Translate	Яндекс Переводчик
<i>...characteristics of the structure of common and non-common languages, commonalities common to all languages...</i>	<i>...общие черты и особенности структуры распространенных и нераспространенных языков, общие черты, общие для всех языков...</i>	<i>...общих черт и характеристик структуры распространенных и нераспространенных языков, общих черт, присущих всем языкам...</i>

Большой разницы в качестве переводов статей из областей лингвистики и информационных технологий не было обнаружено, а все ошибки были распределены по текстам переводов приблизительно в равном количестве.

Нами также был проведён анализ переводов аннотаций к научным статьям, послужившим материалом для данного исследования. Из всего количества, выбранных нами текстов, только 12 сопровождалась аннотациями. Целью анализа является определение качества переводов небольшого фрагмента научного текста. Анализ показал, что больших различий в результатах переводов нет. Некоторые переводы аннотаций были идентичны:

Оригинал	Google Translate	DeepL Translate
<i>The growing interest to the social linguistics turned out impossible without correlation establishment of language and social phenomena.</i>	<i>Растущий интерес к социальной лингвистике оказался невозможным без установления взаимосвязи языка и социальных явлений.</i>	<i>Растущий интерес к социальной лингвистике оказался невозможен без установления взаимосвязи языка и социальных явлений.</i>

Грамматических ошибок в переводах аннотаций нами не было обнаружено. Нами были выявлены только лексические ошибки, а именно проблемы с переводом аббревиатур:

Оригинал	Яндекс Переводчик
<i>... the development of solutions to ensure the stability of the <b>BD</b> function under the conditions of object-orientated aggressive actions.</i>	<i>... для разработки решений по обеспечению стабильности функционирования <b>БПЛА</b> в условиях целенаправленных агрессивных действий.</i>

В целом все программы справились с переводом аннотаций научных текстов. Можно сделать вывод, что СМП хорошо справляются с переводом небольших фрагментов текстов.

Кроме анализа качества переводов научных текстов и аннотаций к ним, мы провели сравнительный анализ переводов, полученных с помощью выбранных нами программ и эталонным переводом. Материалом послужил автореферат Моктар Алиа «Перевод и семантизация названий кинофильмов (на материале английского, французского и арабского языков)». Анализ должен был показать, насколько отличаются результаты машинного перевода от текста, написанного человеком. Переводы оценивались по тем же критериям, что и переводы научных статей.

Анализ показал, что по количеству ошибок DeepL Translate справился с переводом лучше, чем Google Translate и Яндекс Переводчик.

Были обнаружены ошибки по всем критериям. Так, по содержательной идентичности программы дублировали или вставляли в текст лишние слова, добавляли лишние знаки препинания, пропускали слова:

Оригинал на английском языке	Оригинал на русском языке	Перевод Google Translate	Перевод DeepL Translate
<i>...is an American 3D computer-animated comedy film which becomes "Zootropolis" in UK and Ireland "Zoomania" in Germany.</i>	<i>...американский 3D-компьютерный анимационный комедийный фильм, который выходит под названием "Zootropolis" в Великобритании и Ирландии, а в Германии – под названием "Zoomania".</i>	<i>...американский компьютерно-анимационный комедийный фильм в формате 3D, который в Великобритании и Ирландии становится «Зоотропом» в Германии.</i>	<i>...американский 3D компьютерно-анимационный комедийный фильм, который в Великобритании становится „Зоотрополисом“, а в Ирландии „Зооманией“, в Германии.</i>

Со стороны лексического аспекта было обнаружено несколько ошибок в переводе терминов, неверно были транслитерированы некоторые фамилии, были случаи, когда слова или словосочетания остались непереведенными, а также встретились неточности перевода:

Оригинал на английском языке	Оригинал на русском языке	Перевод Яндекс Переводчик
<i>... of the original text within the semantic constraints and linguistic repertoire of the target language.</i>	<i>... как и исходный текст, в рамках семантических ограничений и лексикона языка перевода.</i>	<i>... что и исходный текст, с учетом семантических ограничений. лингвистический репертуар изучаемого языка.</i>

В грамматическом аспекте было обнаружено несколько ошибок в согласовании, склонении фамилий:

Оригинал на английском языке	Оригинал на русском языке	Перевод Яндекс Переводчика
<i>... and the titles of films intended for French-speaking</i>	<i>... с названиями фильмов, предназначенных для</i>	<i>...и названиями фильмов, предназначенных для</i>

<i>audiences in Quebec...</i>	<b>франкоязычной аудитории</b> <i>Квебека...</i>	<b>Франкоязычная аудитория</b> <i>в Квебеке...</i>
Оригинал на английском языке	Оригинал на русском языке	Перевод Google Translate
<i>...(according to Jean Paul Vinay and Jean Derbelnet)...</i>	<i>...(по Жану Полю Вине и Жану Дербельне)...</i>	<i>...(по Жану Полю Винею и Жану Дербельне)...</i>

По критерию стилистической идентичности была обнаружена одна ошибка в переводе Яндекс Переводчика:

Оригинал на английском языке	Оригинал на русском языке	Перевод Яндекс Переводчика
<i>France's <b>cinephilic</b> culture...</i>	<b>Синефильская</b> культура <i>Франции...</i>	<i>Французская</i> <b>киношная</b> <i>культура...</i>

**В заключении** говорится о том, что достаточно высокий уровень перевода научных текстов, который показали все СМП, обусловлен спецификой данного стиля. Объективность, нейтральность, унифицированный способ изложения, высокая степень формализации, строгая структура текста способствуют более качественному выполнению перевода компьютерной программой.

На наш взгляд, ошибки, возникающие в машинном переводе, могут быть следствием целого ряда факторов: недостаточная обученность СМП, отсутствие качественного лингвистического анализа, неумение компьютерных программ анализировать контекст, неоднозначность естественного языка.

В целом, СМП являются полезным инструментом для перевода научных текстов, однако требуют обязательного постредактирования профессиональными переводчиками для устранения ошибок и повышения точности переведенного текста.

Практическая значимость данной работы, на наш взгляд, состоит в том, что результаты, полученные в ходе исследования, могут быть использованы в образовательных программах для обучения будущих специалистов в области перевода, развития навыков постредактирования. Определение типичных

ошибок, допускаемых СМП, позволяет разработчикам улучшать алгоритмы машинного перевода.

**Список цитируемой литературы:**

1. Кириченко, И. А., Сигачева Н. А. Цифровизация перевода с английского на русский язык: анализ и сравнение качества машинного сервиса // Казанский вестник молодых учёных. 2020. Т. 4, № 3. С. 27–37.