

Серова О.У.

**ФЕНОМЕН TRANSLATIONESE В МАШИННОМ ПЕРЕВОДЕ
НАУЧНЫХ СТАТЕЙ С АНГЛИЙСКОГО ЯЗЫКА НА РУССКИЙ
(КОЛИЧЕСТВЕННЫЕ ХАРАКТЕРИСТИКИ)[©]**

*Российский государственный педагогический университет им. А.И. Герцена,
Россия, Санкт-Петербург, serovaolesyau@gmail.com*

Аннотация. Статья посвящена исследованию феномена переводческого языка (*translationese*), выполненному на материале исследовательского сопоставимого корпуса научных текстов на русском языке и текстов машинных переводов на русский язык. Проанализированы такие показатели феномена, как упрощение, интерференция, нормализация и экспликация, оценены их количественные характеристики, установлено их действие в машинном переводе, подтверждающее отличие текстов переводов от оригинальных текстов. Предложены стратегии постредактирования, направленные на устранение индикаторов этих характеристик. Применение этих стратегий может повысить качество машинных переводов научных статей посредством минимизации влияния феномена переводческого языка на текст перевода.

Ключевые слова: машинный перевод; *translationese*; переводческие универсалии; научная статья; индикаторы *translationese*.

Поступила: 01.07.2024

Принята к печати: 08.09.2025

Serova O.U.

**The translationese phenomenon in English-to-Russian
machine translation of scientific papers (quantitative analysis)®**

*Herzen State Pedagogical University of Russia,
Russia, Saint Petersburg, serovaolesyau@gmail.com*

Abstract. The paper concerns the study of the translationese phenomenon based on the comparable corpus of Russian scientific papers and machine-translation results due to such translationese features as simplification, normalization, interference and explicitation. The quantitative analysis of these features is presented. The effect of simplification, interference and explicitation is established, confirming the difference between the Russian and machine-translated papers. Post-editing strategies aimed at eliminating indicators of these features are proposed. Implementing these strategies might improve the quality of machine translation of scientific papers by reducing the impact of translationese.

Keywords: machine translation; translationese; translation universals; scientific style; translationese indicators.

Received: 01.07.2024

Accepted: 08.09.2025

Введение

Результатом перевода научного текста должен быть соответствующий ему по содержанию научный текст, выдержанный в научном стиле, характерном для языка перевода [Комиссаров, 1990, с. 109–111; Пумпянский, 1965, с. 10–11], однако при чтении переводов часто возникает впечатление искусственности их языка. Эта особенность переводного текста стала объектом исследований только в последней четверти XX в.

М. Геллерстам считает, что отличие переводных текстов от оригиналов обусловлено языковой интерференцией со стороны системы исходного языка (ИЯ) [Gellerstam, 1986]. У. Фроули полагает, что текст перевода испытывает влияние как системы ИЯ, так и системы языка перевода (ЯП), а сам текст перевода не сводится ни к одной из них и может быть назван «третьим кодом» [Frawley, 1984]. Г. Тури предлагает рассматривать такой «промежуточный язык» перевода как особое образование диалектного типа, интегрирующее

черты, характерные для конкретной пары языков [Toury, 1979; Koppel, 2011].

Кроме того, существует точка зрения, согласно которой переводные тексты обладают особыми отличительными характеристиками в силу специфики самой переводческой деятельности [Baker, 1993, p. 243]. М. Бейкер называет такие характеристики универсальными характеристиками перевода, или *переводческими универсалиями* [ibid., p. 243–245], которые отображают определенные преобразования, присущие переводческой деятельности. В результате таких преобразований текст перевода приобретает ряд уникальных статистически значимых характеристик (индикаторов), позволяющих отличать его от оригинальных текстов. Это явление принято называть термином *translationese* [Gellerstam, 1986]. Наиболее изученными и эмпирически подтвержденными переводческими универсалиями, свойственными текстам перевода, являются такие их характеристики, как упрощение, нормализация, интерференция и экспликация [Volansky, Ordan, Wintner, 2013].

Феномен *translationese* исследовался на материале разных языковых пар (см., например [Volansky, Ordan, Wintner, 2013; Koppel, 2011; Краснопеева, 2015; Rabinovich, Wintner, 2015]). Значительный вклад в изучение феномена в контексте русского языка ручных переводов был сделан М.А. Куниловской [Kunilovskaya, Kutuzov, 2015; Kunilovskaya, Corpas Pastor, 2021; Kunilovskaya, 2023]. В последние годы набирают популярность исследования *translationese* в машинном переводе (МП) (см., например [Kotait, 2024, Zhou, 2024]). Однако анализ существующей литературы по теме показал лакуарность работ, посвященных изучению феномена на материале МП на русский язык.

Использование систем МП получило широкое распространение как в среде профессиональных переводчиков, так и среди специалистов в конкретных областях знаний [Беляева, 2022]. Переводчикам системы МП позволяют существенно сократить время на перевод, предоставляя так называемый черновой вариант перевода, который затем дорабатывается на этапе постредактирования. Частое применение МП находит место также в научной и учебной деятельности [Беляева, Камшилова, Шубина, 2023]. Исследователи и студенты часто сталкиваются с необходимостью перевода собственных научных статей и иностранных литературных источников. Одним словом, в условиях нехватки времени и/или недостаточного уровня владения языком МП становится незаменимым инструментом.

Однако, несмотря на очевидные преимущества МП, в большинстве случаев он требует тщательного постредактирования, особенно когда речь идет о переводе научных текстов – сфере, в которой вопрос качества перевода встает наиболее остро. Перевод низкого качества может не только затруднить восприятие информации, но и привести к искажению смысла и неверной интерпретации текста, что критично в научной коммуникации, где точность, ясность и логическая связность играют ключевую роль.

В недавнем исследовании [Kunilovskaya, 2023] было установлено, что степень выраженности феномена *translationese* тесно связана с оценкой качества перевода и что индикаторы *translationese* служат предикторами качества перевода на уровне всего текста. Потенциальное наличие *translationese* в МП определяет необходимость выявлять те индикаторы *translationese*, по которым может быть установлена необходимость постредактирования. Установление таких индикаторов в МП научных статей может поспособствовать выработке стратегий преодоления влияния *translationese* и повышению качества текста МП. Более того, в последние годы все больше внимания уделяется важности разработки таких стратегий, так как они становятся необходимой составляющей компетенций современных переводчиков и постредакторов [Бердняев, Светова, Коканова, 2024].

Таким образом, цель статьи заключается в установлении и описании количественных характеристик индикаторов феномена *translationese*, актуализирующих потребность в постредактировании МП, на материале англо-русского сопоставимого корпуса научных статей, а также в предварительном формулировании стратегий постредактирования их МП.

Материал и процедура исследования

Материал исследования – составленный автором сопоставимый корпус из 60 текстов, извлеченных из научных журналов по разным предметным областям (лингвистика, экология, агрохимия, микробиология, ботаника, фармакология, психиатрия), в него вошли 30 текстов, написанных на русском языке, и 30 текстов англо-русских МП. Отбирались статьи, объем которых приблизительно равен шести страницам формата А4 (без учета иллюстраций). Длина текстов была подогнана под среднюю длину всех отобранных статей и составила 114 предложений (для выравнивания длины текстов предложения удалялись случайным образом), тексты прошли предобработку: были

удалены заголовки, аннотации, ключевые слова, ссылки, таблицы, изображения и математические формулы. Перевод выполнялся с использованием системы МП Google Translate.

Для анализа феномена *translationese* были выбраны упомянутые ранее универсалии (*упрощение, нормализация, интерференция, экспликация*). Эти характеристики представлены в переводе соответствующими количественными показателями – индикаторами.

Рассмотрим характеристику *упрощение (simplification)*: ее суть заключается в утверждении, что тексты перевода структурно и лексически проще оригинальных текстов [Baker, 1993, p. 244]. Характеристика оценивается следующими индикаторами.

- Коэффициент лексического разнообразия (*lexical variety* или *TTR*) представляет соотношение числа отдельных лексических единиц (*types*) к общему количеству слов в тексте (*tokens*). Предполагается, что оригинальные (исходные) тексты богаче переводных в отношении лексики.
- Средняя длина слова (*mean word length*) (1) в символах рассчитывается как результат деления общего количества буквенных символов на общее число слов; (2) в слогах – как соотношение количества последовательностей гласных звуков, отграниченных друг от друга согласными звуками или пробелом, и общего количества слов в тексте (предполагается, что в переводе используются структурно более простые слова, а именно более короткие).
- Лексическая плотность (*lexical density*) определяется как отношение знаменательных частей речи, а именно глаголов, имен существительных, имен прилагательных и наречий, к общему количеству слов в тексте. Ожидается, что переводы содержат больше служебной и меньше знаменательной лексики.
- Средняя длина предложения (*mean sentence length*) (1) в символах и (2) в словах вычисляется (1) как результат деления общего количества всех символов текста на общее количество предложений в тексте, а также (2) как результат деления общего количества слов текста на общее количество предложений. Известно, что в процессе немашинного перевода предложения зачастую упрощаются, разбиваются на более короткие и простые.
- Средний ранг слов (*mean word rank*), который предполагает, что оригинальные тексты содержат больше нечастотной лексики, нежели тексты перевода. Для расчета параметра

был использован «Новый частотный словарь русской лексики» [Ляшевская, Шаров, 2009]. По аналогии с исследованием [Volansky, Ordan, Wintner, 2013] были отобраны первые 6000 наиболее частотных слов русского языка и произведен подсчет рангов слов для каждого текста. Слова, не попавшие в частотный список, получали высший ранг в 6000. Показатель индикатора есть среднее арифметическое всех рангов слов текста.

- Индекс удобочитаемости Флэша (*the Flesch readability score*), скорректированный для русского языка [Оборнева, 2006], оценивающий уровень сложности восприятия текста (чем выше показатель, тем более легким считается текст).

Следующей характеристикой является *нормализация (normalization)*, подразумевающая, что в переводах используются более нормализованные грамматические структуры, больше устойчивых выражений, а также проявляется тенденция к избеганию лексических повторов знаменательных частей речи. Анализировался индикатор «повторяемость лексических единиц, являющихся знаменательными частями речи» (*repetition*). Показатель рассчитывается посредством деления абсолютной частоты глаголов, имен существительных, имен прилагательных и наречий, встречающихся в тексте более одного раза, на общее количество токенов в тексте.

Кроме того, была проанализирована характеристика *интерференция (interference)*, предполагающая тенденцию к переносу структурных характеристик ИЯ в текст перевода. Был выбран следующий индикатор.

- Устойчивость частеречных последовательностей (*POS-n-grams: uni-, bi-, trigrams*) предполагает перенос в текст перевода синтаксических черт ИЯ, в результате чего повышается частота использования частеречных n-граммов, свойственных ИЯ. Подсчитываются нормализованные частоты n-грамм – отношение каждой n-граммы ко всем n-граммам текста.

Следующей рассматриваемой характеристикой стала *экспликация (explication)*, предполагающая тенденцию к эксплицитному выражению в переводе элементов, имплицитно подразумевающихся в тексте оригинала [Volansky, Ordan, Wintner, 2013]. Результатом ее действия является наличие большего по сравнению с текстами оригиналов количества средств обеспечения логической последовательности элементов и семантической целостности текста. Следует отметить, что для научного текста в целом характерно частотное использование специальных служебных слов и слово-

сочетаний, обеспечивающих логические связи между элементами высказываний, что делает текст более структурированным и точным [Пумпянский, 1965]. Это обусловлено необходимостью установления причинных, временных, пространственных и других отношений между предложениями [Комиссаров, 1990]. Данная черта свойственна научным текстам на английском языке и их русскоязычным переводам, однако в англоязычных текстах логические связи между отдельными высказываниями чаще обнаруживаются в самом их содержании, в то время как для русскоязычных научных текстов, напротив, в большей мере характерно эксплицитное выражение логических связей [там же]. С.В. Когут предлагает применять к таким связующим элементам определение «дискурсивные маркеры» (ДМ) – языковые единицы, выполняющие в научном дискурсе вспомогательную, коммуникативно- и речеоорганизующую функцию [Когут, 2016]. В связи с этим был выбран индикатор «дискурсивные маркеры» (*discourse markers*). Предполагается, что для переводов характерно более частое использование ДМ в сравнении с оригинальными текстами. Для анализа ДМ была взята классификация ДМ научного текста [там же]:

- маркеры связи текста на глобальном уровне, обеспечивающие связность разных частей текста и его цельность (ДМ введения в тему, порядка следования информации, расположения иллюстративного материала, выводов);
- маркеры связи текста на локальном уровне, обеспечивающие связь предложений и их частей (ДМ введения новой информации, пояснения, противопоставления, введения примеров).

Для анализа *экспликаций* были выбраны следующие индикаторы [Volansky, Ordan, Wintner, 2013].

- Эксплицитное название (*explicit naming*) – отношение личных и притяжательных местоимений к именованным сущностям. Предполагается, что переводам свойственно более частое использование местоимений в целях упрощения идентификации именованных сущностей.
- Единичное именование (*single naming*) – отношение именованных сущностей из первого токена ко всем именованным сущностям. Считается, что этот показатель ниже в переводах, так как переводчики, стремясь к точности, предпочитают использовать полные версии имен, тогда как в оригинальных текстах чаще используются их более лаконичные варианты.

- Средняя длина именованных сущностей (*mean named entities length*) – среднее арифметическое длин именованных сущностей в токенах.

Помимо представленных выше индикаторов были отобраны индикаторы, не вошедшие ни в одну из перечисленных характеристик:

- местоимения (*pronouns*), при анализе которых подсчитываются их нормализованные частоты относительно всех словарных токенов. Анализируется использование личных, притяжательных, относительных, неопределенных, отрицательных, указательных, определительных и возвратного местоимений;
- пунктуационные знаки (*punctuation*), анализ которых подразумевает подсчет нормализованных частот знаков препинания (, . ? ! : ; – () «» “”) относительно всех знаков пунктуации.

По каждому из перечисленных индикаторов сначала рассчитывались нормализованные частоты показателей для каждого текста, это обеспечило сопоставимость показателей для текстов разной длины. Далее производился подсчет средних показателей индикаторов по каждому подкорпусу, для этого использовалась формула взвешенного арифметического среднего (*weighted arithmetic mean*), которая учитывает объем текстов, присваивая каждому из них вес, пропорциональный количеству токенов. Подсчет всех показателей выполнялся с использованием специально написанных для этой цели Python-скриптов, что позволило минимизировать погрешности в результатах. Лемматизация слов и тегирование частей речи, где это требовалось, проводились с использованием библиотеки *rumorphy*¹, распознавание именованных сущностей – с помощью библиотеки *Natasha*².

Результаты исследования

Перечисленные индикаторы позволили провести анализ корпуса на наличие специфических признаков *translationese* в МП.

Первой анализируемой характеристикой стало *упрощение* (табл. 1). Существенное отличие продемонстрировал индикатор

¹ URL: <https://pymorphy2.readthedocs.io/en/stable/> – морфологический анализатор для русского языка.

² URL: <https://pypi.org/project/natasha/> – библиотека для обработки естественного русского языка.

среднего ранга слов. Большой разрыв по этому показателю в пользу оригинальных текстов указывает на то, что уникальность используемой лексики в текстах МП значительно ниже, что является признаком *лексического упрощения*, в пользу которого также могут свидетельствовать результаты анализа *лексической вариативности* и *лексической плотности*. Показатели *средней длины слов / предложений* не подтвердили прогнозов. Удлинение предложений в МП может быть результатом действия характеристики *экспликация*. Анализ *удобочитаемости текста*, вопреки ожиданиям, показал, что тексты МП значительно более сложны для восприятия, нежели оригинальные тексты.

Таблица 1

Средние значения индикаторов характеристики simplification

Индикатор	Показатели		
	Оригинальные тексты	Машинный перевод	Разница
Lexical variety (TTR)	30,61%	28,95%	1,66%
Lexical density	61,97%	61,9%	0,07%
Mean word length	6,67 символа 2,85 слога	6,9 символа 2,98 слога	0,23 символа 0,13 слога
Mean sentence length	152,6 символа 26,5 слов	169 символов 28,5 слова	16,4 символа 2 слова
Mean word rank	2333	1843	498
Readability score	8,39	-2,3	6,09

В результате анализа индикатора *повторяемости* характеристики *нормализация* не было выявлено значительных различий: для оригинальных текстов она составила 40,6%, для МП – 40,4%, что говорит об отсутствии действия рассматриваемой характеристики в МП по этому параметру.

Результаты анализа *устойчивости частеречных последовательностей* характеристики *интерференция* представлены в табл. 2. В анализ были добавлены специальные обозначения для запятой (COMMA), конечного знака препинания (S_END) и начала предложения (S_START).

На основании полученных данных можно сделать предварительные выводы по некоторым n-граммам, продемонстрировавшим наибольшее различие. Вероятно, более частое использование запятых, союзов и местоименных существительных (и комбинаций

с ними) связано с *экспликацией* и обусловлено «стремлением» системы МП сделать перевод более связным и понятным, в том числе посредством частого использования придаточных предложений, вводимых союзами / союзными словами, и уточнения имен существительных-антецедентов соответствующими им местоимениями. Возможно также, что увеличение использования запятых в переводе связано с *нормализацией* текста: в английском языке пунктуация менее регламентирована, особенно если речь идет о придаточных предложениях, в то время как при переводе на русский язык, где грамматические правила требуют более четкого структурирования и выделения клауз с помощью пунктуации, система МП может прибегать к *нормализации* текста. Рост использования местоимений, вероятно, частично обусловлен и *интерференцией* с английского языка, в котором в связи со слабо представленной системой глагольных окончаний обязательно присутствие в предложении подлежащего, даже если оно выражено местоимением. Помимо этого, определенный артикль *the* иногда переводится на русский язык указательным местоимением.

Значительное увеличение комбинаций личных форм глаголов с инфинитивом, возможно, также есть результат интерференции: в английском языке, по сравнению с русским языком, конструкция *VERB+INFN* используется довольно часто для выражения цели, намерения или действия, в том числе и в научных работах. Например, *We aim to develop a new project* переводится системой МП как «Мы *стремимся разработать* новый проект», в то время как для русского языка научного стиля более характерно использование в подобных случаях имен существительных (в том числе отглагольных): «Наша *цель* заключается в *разработке* нового проекта».

Снижение частоты употребления в переводе длинных именных групп и увеличение сочетаний типа *NOUN+PREP+NOUN* может говорить о «предпочтении» системой МП в некоторых случаях замены первых на предложные конструкции. Так, *risk management strategy* может переводиться как «стратегия по управлению рисками» вместо «стратегия управления рисками».

Показатель *эксплицитного* называния продемонстрировал значительное различие между оригинальными текстами и текстами МП – 103% и 158% соответственно. Это подтверждает гипотезу о том, что в переводе чаще используются личные и притяжательные местоимения для уточнения имен собственных. В то же время индикаторы *единичного* именования и *средней длины* именованных

сущностей не подтвердили прогноз: первый оказался заметно выше для МП (67% против 77%), а второй – несколько выше для оригинальных текстов (1,47 против 1,41 токена).

Таблица 2

Нормализованные частоты частеречных n-грамм

Parts of Speech	Оригинальные тексты	Машинный перевод	Разница
1-grams			
NOUN	35%	32%	3%
COMMA	6,6%	9%	2,4%
CONJ	5,9%	7,2%	1,3%
VERB	5,3%	6,1%	0,08%
INFN	0,9%	2%	1,1%
NPRO	1,2%	1,5%	0,3%
2-grams			
NOUN+NOUN	9,7%	7,9%	1,8%
NOUN+ADJF	4,4%	3%	1,4%
NOUN+PREP	5,37%	5,62%	0,25%
COMMA+CONJ	1,9%	2,4%	0,5%
CONJ+NOUN	1,7%	2,5%	0,8%
VERB+INFN	0,4%	0,75%	0,35%
S_TART+CONJ	0,35%	0,5%	0,15%
INFN+NOUN	0,25%	0,7%	0,45%
NPRO+VERB	0,25%	0,45%	0,2%
S_TART+NPRO	0,08%	0,11%	0,03%
VERB+NPRO	0,06%	0,1%	0,04%
3-grams			
NOUN+ADJF+NOUN			
ADJF+NOUN+ NOUN	3,4%	2,4%	1%
NOUN+NOUN+NOUN	2,6%	1,76%	0,84%
NOUN+NOUN+ADJF	2,45%	1,7%	0,75%
	1%	0,5%	0,5%
NOUN+ PREP+ NOUN			
NOUN+ COMMA+ NOUN	3%	3,7%	0,7%
	1%	1,5%	0,5%
NOUN + CONJ+NOUN	1%	1,5%	0,5%
NOUN+ COMMA+ CONJ	1%	1,2%	0,2%
	0,4%	0,6%	0,2%
COMMA+ CONJ+NOUN	0,09%	0,2%	0,11%
VERB+INFN+NOUN			

Таблица 3

Нормализованные частоты ДМ связи научного текста

Тип маркеров	Оригинальные тексты	Машинный перевод	Разница (%)
Локальный уровень			
ДМ противопоставления	0,9%	0,7%	0,2%
ДМ пояснения	0,2%	0,52%	0,32%
ДМ введения новой / доп. информации	0,5%	0,7%	0,2%
ДМ введения примеров	0,16%	0,4%	0,24%
Глобальный уровень			
ДМ порядка следования информации	0,9%	0,15%	0,75%
ДМ выводов	0,4%	0,3%	0,1%

Результаты анализа ДМ (табл. 3) демонстрируют, что в МП маркеры связи текста на локальном уровне используются гораздо чаще (кроме ДМ противопоставления), что подтверждает действие характеристики *экспликация*. При этом наблюдается снижение частоты некоторых ДМ связи текста на глобальном уровне. Вероятно, это связано с тем, что системы МП не всегда способны «уловить» и передать макроструктурные связи текста, что приводит к общему снижению когерентности текста.

Анализ знаков препинания (табл. 4) показал значительный разрыв в использовании запятых в пользу МП, что согласуется с полученными ранее результатами. Снижение количества точек, вероятно, связано с удлинением предложений в МП. Интересно, что в МП используется больше кавычек-ёлочек, свойственных для русского языка, в то время как английские двойные кавычки в переводе не представлены.

Анализ частот местоимений (табл. 5) показал, что для МП в целом характерно увеличение использования всех местоименных групп, это свидетельствует о «склонности» МП к *экспликации*. Более частое употребление некоторых местоимений, вероятно, объясняется действием *интерференции*. Так, в английском языке возвратные местоимения более распространены и используются в случаях, где русский язык обходится либо без них, либо постфиксом «-ся». Увеличение частоты местоимений *другой, любой, некоторый* и *несколько*, скорее всего, обусловлено частым использованием в английском языке местоимений *another, any* и *some*.

Таблица 4

Нормализованные частоты знаков препинания

Знак	Оригинальные тексты	Машинный перевод	Разница (%)
.	25,15%	23,6%	1,55%
)	11%	7,5%	3,5%
(11%	7,2%	3,8%
—	4%	0,5%	3,5%
:	1,9%	1,3%	0,6%
“	1,53%	0%	1,53%
”	1,53%	0%	1,53%
,	41,5%	54%	12,5%
;	1,5%	2%	0,5%
«	0,5%	1,8%	1,3%
»	0,5%	1,8%	1,3%
?	0,03%	0,2%	0,17%

Таблица 5

Нормализованные частоты местоимений

Местоимение	Оригинальные тексты	Машинный перевод	Разница (%)
Личные местоимения			
он	0,21%	0,15%	0,06%
они	0,5%	0,7%	0,2%
вы	0,003%	0,02%	0,017%
Возвратное местоимение			
себя	0,08%	0,2%	0,12%
Притяжательные местоимения			
наш	0,07%	0,23%	0,16%
свой	0,12%	0,16%	0,04%
мой	0,001%	0,01%	0,009%
ваш	0%	0,004%	0,004%
Указательные местоимения			
так	0,26%	0,12%	0,14%
указанный	0,014%	0,003%	0,011%
это	0,4%	0,62%	0,22%
этот	0,3%	0,52%	0,22%
такой	0,24%	0,35%	0,11%
тот	0,21%	0,5%	0,29%
то	0,06%	0,11%	0,05%

Местоимение	Оригинальные тексты	Машинный перевод	Разница (%)
Определительные местоимения			
самый	0,04%	0,03%	0,01%
иной	0,021%	0,003%	0,018%
весь	0,22%	0,2%	0,02%
другой	0,17%	0,32%	0,15%
каждый	0,1%	0,14%	0,04%
сам	0,03%	0,06%	0,03%
любой	0,007%	0,03%	0,023%
Относительные местоимения			
сколько	0,006%	0,001%	0,005%
что	0,7%	1,1%	0,4%
какой	0,014%	0,05%	0,036%
чей	0,001%	0,01%	0,009%
кто	0,005%	0,007%	0,002%
Неопределенные местоимения			
некоторый	0,07%	0,14%	0,07%
несколько	0,06%	0,1%	0,04%

Выводы

В ходе исследования было установлено, что тексты МП демонстрируют характерные признаки лексического упрощения, экспликации и интерференции. На основе индикаторов, отражающих эти признаки, можно предложить возможные стратегии постредактирования для улучшения качества МП научных статей.

1. *Повышение лексического разнообразия* подразумевает анализ перевода на наличие частотной знаменательной лексики с использованием частотных списков слов (например, [Ляшевская, Шаров, 2009]). Это позволит понять, какие лексемы могут быть заменены более уникальными синонимичными вариантами. Например, слово *рост* с невысоким рангом уникальности (559), в зависимости от контекста, можно заменить на более уникальные варианты: *прирост* (6327), *прибавка* (13924), *возрастание* (14014) и т.п. С целью нахождения более уникальных и интересных вариантов рекомендуется пользоваться как словарями синонимов, так и тезаурусами. Например, RuWordNet¹ предлагает не только синонимичные

¹ URL: <https://www.ruwordnet.ru/ru> – тезаурус русского языка.

слова, но и слова, состоящие с заменяемым словом в отношениях антонимии, гипо-, гиперонимии, меро- и холонимии.

2. *Улучшение удобочитаемости текста* предполагает разбиение слишком длинных предложений МП на более короткие и перефразирование сложных для восприятия конструкций.

3. *Устранение избыточной экспликации* заключается в тщательной проверке МП на предмет избыточного использования местоимений и их дальнейшем удалении / замене на соответствующие им имена-антецеденты, а также, если это необходимо, в упрощении синтаксической структуры сложных предложений более лаконичными конструкциями.

4. *Контроль интерференции со стороны английского языка* предполагает нахождение и устранение / замену на более естественные варианты таких несвойственных русскому языку явлений как чрезмерное употребление составных глагольных сказуемых, указательных местоимений, местоимений *себя, другой, любой, некоторый и несколько*, а также замену конструкций типа NOUN+PREP+NOUN более длинными номинативными цепочками при необходимости.

5. *Повышение когерентности текста* посредством тщательного анализа логических связей между предложениями и абзацами текста и коррекции / обеспечения этих связей при необходимости. При этом рекомендуется пользоваться списком ДМ научного текста для связи на глобальном уровне.

Предложенные стратегии могут поспособствовать повышению качества МП научных статей, что определяет практическую значимость исследования. Теоретическая значимость состоит в углублении понимания проявления *translationese* в МП научных статей на русский язык. В дальнейшем целесообразным представляется расширение корпуса анализируемых текстов МП, сравнение результатов анализа МП с результатами ручных переводов научных статей, а также расширение спектра анализируемых индикаторов.

Список литературы

- Беляева Л.Н. Машинный перевод в современной технологии процесса перевода // Известия Российского государственного педагогического университета им. А.И. Герцена. – 2022. – № 203. – С.22–30.
- Беляева Л.Н., Камшилова О.Н., Шубина Н.Л. Научная статья в технологическом пространстве машинного перевода: правила и процедуры редактирования. – Санкт-Петербург : Книжный дом, 2023. – 90 с.

- Берендяев М.В., Светова С.Ю., Коканова Е.С. Автоматическая генерация текстов в профессиональном переводе // *Didactica Translatoria*. – 2024. – № 2. – С. 5–10.
- Козут С.В. Дискурсивные маркеры в письменном научном дискурсе // *Сибирский филологический журнал*. – 2016. – № 2. – С. 157–163.
- Комиссаров В.Н. Теория перевода (лингвистические аспекты). – Москва : Высшая школа, 1990. – 253 с.
- Краснопеева Е.С. Лексические особенности русскоязычного переводного дискурса: корпусное сравнительно-сопоставительное исследование на материале современной художественной прозы: дис. ... канд. филол. наук. – Челябинск, 2015. – 211 с.
- Ляшевская О.Н., Шаров С.А. Новый частотный словарь русской лексики – Москва: Азбуковник, 2009. – URL: dict.ruslang.ru/freq.php
- Оборнева И.В. Автоматизированная оценка сложности учебных текстов на основе статистических параметров : дис. ... канд. пед. наук. – Москва, 2006. – 165 с.
- Пумпянский А.Л. Введение в практику перевода научной и технической литературы на английский язык. – Москва : Наука, 1965. – 302 с.
- Baker M. Corpus linguistics and translation studies: Implications and applications // *Text and Technology*. – 1993. – Vol. 64. – P. 233–252.
- Frawley W. Prolegomenon to a theory of translation, Translation // *Translation: Literary, Linguistic and Philosophical Perspectives* / ed. by W. Frawley. – Newark ; London : University of Delaware Press; Associated University Presses, 1984. – P. 159–175.
- Gellerstam M. Translationese in Swedish novels translated from English // *Translation Studies in Scandinavia* / ed. by L. Wollin, H. Lindquist. – Lund : CWK Gleerup, 1986. – P. 88–95.
- Koppel M., Ordan N. Translationese and its dialects // *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. – Portland : Association for Computational Linguistics, 2011. – P. 1318–1326.
- Kotait R. Richness lost in machine translationese: lexical richness in human translation versus neural machine translation from Arabic into English // *Egyptian Journal of Language Engineering*. – 2024. – Vol. 11, N 1. – P. 66–85.
- Kunilovskaya M., Corpas Pastor G. Translationese and register variation in English-to-Russian professional translation // *New Perspectives on Corpus Translation Studies* / ed. by V.X. Wang, D. Li, L. Lim. – Singapore : Springer Nature Singapore, 2021. – P. 133–180.
- Kunilovskaya M., Kutuzov A. A quantitative study of translational Russian (based on a translational learner corpus) // *Corpus Linguistics 2015: Proceedings of 7th International Conference*. – Saint-Petersburg : Saint-Petersburg State University, 2015. – P. 33–40.
- Rabinovich E., Wintner S. Unsupervised identification of translationese // *Transactions of the Association for Computational Linguistics*. – 2015. – Vol. 3. – P. 419–432.
- Toury G. Interlanguage and its manifestations in translation // *Meta: Translators' Journal*. – 1979. – Vol. 24, N 2. – P. 223–231.
- Volansky V., Ordan N., Wintner Sh. On the features of translationese // *Digital Scholarship in the Humanities*. – 2013. – Vol. 30, N 1. – P. 98–118.
- Zhou F. The comparison of translationese in machine translation and human translation in terms of translation relations // *arXiv preprint*. – 2024. – URL: <https://arxiv.org/abs/2404.08661>

References

- Beljaeva, L.N., Kamshilova, O.N., Shubina, N.L. (2023). *Nauchnaja stat'ja v tehnologicheskom prostranstve mashinnogo perevoda: pravila i protsedury redaktirovanija*. Saint-Petersburg: Knizhnyj dom.
- Beljaeva, L.N. (2022). Mashinnyj perevod v sovremennoj tehnologii protsessa perevoda. *Izvestija Rossijskogo gosudarstvennogo pedagogicheskogo universiteta im. A.I. Gertsena*, 203, 22–30.
- Berendjaev, M.V., Svetova, S.Ju., Kokanova, E.S. (2024). Avtomaticheskaja generatsija tekstov v professional'nom perevode. *Didactica Translatorica*, 2, 5–10.
- Kogut, S.V. (2016). Diskursivnye markery v pis'mennom nauchnom diskurse. *Sibirskij filologicheskij zhurnal*, 2, 157–163.
- Komissarov, V.N. (1990). *Teorija perevoda (lingvisticheskie aspekty)*. Moscow: Vysshaja shkola.
- Krasnopeeveva, E.S. (2015). *Leksicheskie osobennosti russkojazychnogo perevodnogo diskursa: korpusnoe sravnitel'no-sopostavitel'noe issledovanie na materiale sovremennoj hudozhestvennoj prozy*. (Unpublished doctoral dissertation). Cheljabinsk.
- Ljashevskaja, O.N., Sharov, S.A. (2009). *Novyj chastotnyj slovar' russkoj leksiki*. Moscow: Azbukovnik. Retrieved from: <http://dict.ruslang.ru/freq.php>
- Oborneva, I.V. (2006). *Avtomatizirovannaja otsenka slozhnosti uchebnyh tekstov na osnove statisticheskikh parametrov*. (Unpublished doctoral dissertation). Moscow.
- Pumpjanskij, A.L. (1965). *Vvedenie v praktiku perevoda nauchnoj i tehnicheckoj literatury na anglijskij jazyk*. Moscow: Nauka.
- Baker, M. (1993). Corpus linguistics and translation studies: implications and applications. *Text and technology*, 64, 234–247.
- Frawley, W. (1984). Prolegomenon to a theory of translation. In *Translation. Literary, Linguistic and Philosophical Perspectives* (pp. 159–175). Newark: University of Delaware Press; London: Associated University Presses.
- Gellerstam, M. (1986). Translationese in Swedish novels translated from English. In *Translation Studies in Scandinavia* (pp. 88–95). Lund: CWK Gleerup.
- Koppel, M., Ordan, N. (2011). Translationese and its dialects. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 1318–1326). Portland: Association for Computational Linguistics.
- Kotait, R. (2024). Richness Lost in Machine Translationese: Lexical Richness in Human Translation versus Neural Machine Translation from Arabic into English. *Egyptian Journal of Language Engineering*, 11(1), 66–85.
- Kunilovskaya, M., Corpas Pastor, G. (2021). Translationese and register variation in English-to-Russian professional translation. In *New Perspectives on Corpus Translation Studies* (pp. 133–180). Singapore: Springer Nature Singapore.
- Kunilovskaya, M., Kutuzov, A. (2015). A quantitative study of translational Russian (based on a translational learner corpus). In *Corpus Linguistics 2015: Proceedings of 7th International Conference* (pp. 33–40). Saint-Petersburg: Saint Petersburg State University.
- Rabinovich, E., Wintner, S. (2015). Unsupervised identification of translationese. *Transactions of the Association for Computational Linguistics*, 3, 419–432.

- Toury, G. (1979). Interlanguage and its manifestations in translation. *Meta: Translators' Journal*, 24(2), 223–231.
- Volansky, V., Ordan, N., Wintner, Sh. (2013). On the features of translationese. *Digital Scholarship in the Humanities*, 30(1), 98–118.
- Zhou, F. (2024). The comparison of translationese in machine translation and Human translation in terms of translation relations. arXiv preprint arXiv:2404.08661. Retrieved from: <https://arxiv.org/abs/2404.08661>
-

Сведения об авторе

Серова Олеся Усмановна – аспирант кафедры образовательных технологий в филологии, Российский государственный педагогический университет им. А.И. Герцена, Россия, Санкт-Петербург, serovaolesyau@gmail.com

About the author

Serova Olesya Usmanovna – Postgraduate student of the Department of Educational Technologies in Philology, Herzen State Pedagogical University of Russia, Russia, Saint-Petersburg, serovaolesyau@gmail.com