

Научная статья

УДК 81'33

DOI: <https://doi.org/10.18721/JHSS.16105>

EDN: <https://elibrary/THFJON>



АВТОМАТИЗИРОВАННОЕ ИССЛЕДОВАНИЕ ИНДИКАТОРОВ ПЕРЕВОДЧЕСКОГО ЯЗЫКА НА МАТЕРИАЛЕ ПЕРЕВОДОВ НАУЧНЫХ СТАТЕЙ

О.У. Серова  

Российский государственный педагогический университет
им. А.И. Герцена, Санкт-Петербург, Российская Федерация

 serovaolesyau@gmail.com

Аннотация. В условиях роста публикационной активности ученых и международного обмена знаниями растут и требования, предъявляемые к качеству переводов научных статей. Среди факторов, влияющих на качество переводов, особое место занимает феномен переводческого языка (translationese), проявляющийся в ряде особых, свойственных переводам характеристик – в упрощении, экспликации, нормализации, интерференции, отличающих их от непереодных текстов, написанных на том же языке. Данные характеристики могут служить предикторами качества выполняемых переводов, что делает важным их учет как в самом процессе перевода, так и при принятии решения о его постредактировании. Вместе с тем анализ индикаторов этих характеристик является весьма долгим и трудоемким процессом, в связи с чем возникает необходимость в инструментах, способных автоматизировать этот процесс. В статье данная проблема решается с использованием специализированного инструмента, разработанного для сбора корпусов текстов, анализа индикаторов феномена переводческого языка (как для отдельных текстов, так и для целых корпусов текстов), а также для сопоставления показателей индикаторов по корпусам. Материалом для проверки корректности работы анализатора послужил сопоставимый корпус научных статей по социологии – корпус, включающий в себя два текстовых подкорпуса, схожих по стилю, жанру, тематике и объему, но различающихся по происхождению: первый подкорпус составили оригинальные (непереодные) тексты научных статей на русском языке, а второй – тексты англо-русских переводов статей. В результате анализа подтверждается наличие в переводах таких индикаторов переводческого языка, как снижение уровня лексической вариативности, уникальности используемой лексики, лексической повторяемости, повышение уровня эксплицитности посредством увеличения частоты использования личных, притяжательных и указательных местоимений, дискурсивных маркеров пояснения и противопоставления, а также интерференция, выражающаяся в более частом использовании прямого порядка слов, предпочтении глаголов отглагольным существительным, более частом использовании придаточных определительных вместо причастных оборотов. Дается интерпретация полученных данных и предлагаются возможные способы преодоления влияния рассматриваемого феномена на переводы научных статей с английского на русский язык.

Ключевые слова: индикаторы translationese, автоматическая обработка текста, научная статья, упрощение, нормализация, экспликация, интерференция.

Для цитирования: Серова О.У. Автоматизированное исследование индикаторов переводческого языка на материале переводов научных статей // Terra Linguistica. 2025. Т. 16. № 1. С. 61–81. DOI: 10.18721/JHSS.16105



AN AUTOMATED STUDY OF TRANSLATIONESE INDICATORS IN SCIENTIFIC PAPER TRANSLATIONS

O.U. Serova 

Herzen State Pedagogical University of Russia,
St. Petersburg, Russian Federation

 serovaolesyau@gmail.com

Abstract. The rise in global scientific publication output and international knowledge exchange has heightened the demands for the quality of scientific paper translations. Among the factors influencing their quality, the phenomenon of translationese stands out due to its characteristics such as simplification, explicitation, normalization, and interference. Not only do they differentiate translations from non-translated texts, but they might also be indicative of translation quality, suggesting the need to consider them when evaluating the quality of a translation and deciding on its further post-editing. However, calculating and analyzing the indicators of these characteristics is a lengthy and labor-intensive process, which creates the need for tools to automate it. The article offers a solution to this issue through the use of a specialized tool designed for compiling text corpora, identifying translationese indicators in individual texts and entire corpora, as well as comparing collected data across corpora. A comparable corpus of sociology-related scientific papers is used as material for validating the accuracy of the tool – this corpus includes two textual subcorpora of the same style, genre, subject matter, and size but of different origin: the first subcorpus comprised original (non-translated) Russian texts, while the second one consisted of English-to-Russian translations. The results of the analysis reveal the presence of such translationese indicators in translations as reduced lexical variety, lexical uniqueness, and lexical repetition; increased text explicitness characterized by the frequent use of personal, possessive, and demonstrative pronouns and discourse markers used for clarifying information and contrast; a more frequent use of direct word order and preference for verbs rather than their substantivized counterparts; the tendency to use relative clauses more often than participial phrases. The article discusses the implications of these findings and outlines possible strategies for reducing their impact on English-to-Russian scientific paper translations.

Keywords: translationese indicators, automatic text processing, scientific paper, simplification, normalization, explicitation, interference.

Citation: Serova O.U., An automated study of translationese indicators in scientific paper translations, *Terra Linguistica*, 16 (1) (2025) 61–81. DOI: 10.18721/JHSS.16105

Введение

Многочисленные исследования в области переводоведения демонстрируют существенные различия между переводами и непереводами текстами, изначально написанными на языке перевода (ПЯ) (см., например, [1–10]). Эти отличия лежат в основе концепции так называемого переводческого языка, обозначаемого термином *translationese* [11].

В переводных текстах феномен *translationese* эксплицируется рядом характеристик, каждая из которых представлена соответствующими ей индикаторами. Часть характеристик связана с задействованной в процессе перевода языковой парой (*интерференция*) [12, 13], другие характеристики обусловлены самим этим процессом как таковым (*упрощение, экспликация, нормализация*) [3, 14]. При этом под *интерференцией* понимается перенос в текст перевода морфологических и синтаксических особенностей, свойственных структурной организации исходного языка (ИЯ), в связи с чем предлагается концепция различных «диалектов» переводческого языка [15], что подтверждается эмпирическими данными [2]. *Упрощение* представляет собой стремление переводчика сделать текст перевода морфологически, лексически и синтаксически более простым, чтобы облегчить его понимание читателем. *Экспликация* заключается в повышении объяснительного



потенциала перевода за счет более частого использования средств обеспечения когезии и когерентности текста. *Нормализация* состоит в стремлении использовать в переводе более стандартизированные структуры ПЯ, а также избегать лексических повторов и нетипичных для ПЯ коллокаций.

Различия, обусловленные перечисленными характеристиками, иногда связывают с уровнем профессионализма переводчика – недостаток опыта приводит к «неуклюжести» и неестественности текста перевода при его восприятии носителями целевого языка [14, с. 249; 16, р. 292–293]. В недавнем исследовании М.А. Куниловской¹, основанном на анализе корпусов переводных текстов (англо-русский перевод) и оригинальных русскоязычных текстов в области СМИ, было установлено, что индикаторы *translationese* могут выступать предикторами качества перевода на уровне всего текста, при этом переводы более низкого качества демонстрируют более выраженные признаки *translationese*. Снижение качества перевода наиболее важно учитывать, когда речь идет о переводе научных текстов, в частности – научных статей, поскольку некачественный перевод способен привести к неверной интерпретации процедуры, методов исследования, выводов и иных данных, представленных в статье, а также подорвать доверие к автору и снизить влияние его работы в академическом сообществе. Особую актуальность эта проблема приобретает в условиях роста исследовательской публикационной активности в последние десятилетия² и необходимости международного обмена знаниями.

Проведение исследований индикаторов *translationese* в переводах научных статей позволило бы переводчикам более объективно оценивать качество выполняемых ими переводов, а постредакторам – выявлять потенциальные проблемные аспекты перевода, требующие корректировки. Наряду с этим следует подчеркнуть, что процесс исследования феномена *translationese* сопряжен с высокой сложностью, обусловленной двумя факторами. Первым фактором является большое разнообразие индикаторов характеристик исследуемого феномена, каждый из которых требует отдельного анализа. Второй заключается в том, что исследование феномена *translationese*, как и любое другое корпусное исследование, предполагает составление текстовых корпусов значительного объема, что, в свою очередь, обуславливает необходимость обработки больших объемов текстовых данных, «ручной» анализ которых является крайне трудоемкой и практически невыполнимой задачей, поэтому автоматизация текстового анализа в исследованиях подобного рода становится не просто желательной, а необходимой мерой [17, с. 100]. В связи с этим целью настоящей статьи является применение специально разработанного инструмента – автоматического анализатора феномена *translationese*³ – для анализа индикаторов рассматриваемого феномена на материале переводов научных статей по социологии, а также интерпретация полученных результатов и рассмотрение возможных путей смягчения влияния данного феномена на текст перевода.

Материал и методы исследования

В ходе исследования на каждом его этапе использовался функционал автоматического анализатора феномена *translationese*. Сначала перечислим использованные в исследовании функции программы:

1) анализ вводимых текстов на наличие индикаторов характеристик феномена *translationese* с дальнейшим сохранением результатов анализа в выбранную базу данных, соответствующую одному из формируемых подкорпусов – непереводному или переводному подкорпусу;

¹ Kunilovskaya M. Translationese indicators for human translation quality estimation (based on English-to-Russian translation of mass-media texts): Ph.D. thesis, University of Wolverhampton, 2023. URL: <http://hdl.handle.net/2436/625250> (дата обращения: 10.07.2024).

² Publication Output by Region, Country, or Economy and by Scientific Field // Publications Output: U.S. Trends and International Comparisons. 2023. URL: <https://nces.nsf.gov/pubs/nsb202333/publication-output-by-region-country-or-economy-and-by-scientific-field/> (дата обращения: 14.09.2024).

³ Описание проекта. TRANSLATIONESE ANALYSER // translationese-analyzer 0.0.0. URL: <https://pypi.org/project/translationese-analyzer/> (дата обращения: 26.03.2025).



2) переводной подкорпус, состоящий из полнотекстовых англо-русских переводов научных статей.

Каждый подкорпус включал 30 текстов, общий объем каждого из подкорпусов составил около 166 000 словоформ.

В целях обеспечения сопоставимости материала при выборе текстов для анализа применялись строгие критерии отбора, в соответствии с которыми все тексты корпуса относятся:

- к научному функциональному стилю, к жанру «научная статья»;
- к единой предметной области – социологии, в связи с чем все анализируемые тексты (как переводы, так и непередаваемые тексты) были взяты из рецензируемых научных периодических изданий по социологии⁴. Предметная гомогенность корпуса позволила минимизировать влияние предметной области на лингвистические особенности текстов и сосредоточиться на выявлении признаков феномена *translationese*.

Все тексты обоих подкорпусов прошли предобработку: из них были автоматически удалены ссылки на литературу, вручную – заголовки, аннотации, ключевые слова, сноски, списки литературы и иллюстративные материалы.

Далее все тексты были проанализированы с помощью анализатора феномена *translationese* на наличие упомянутых ранее характеристик феномена. Так, для анализа каждой из характеристик применялся комплекс предоставляемых анализатором методов, позволяющих вычислять значения соответствующих этой характеристике индикаторов, сопоставлять полученные данные по корпусам и интерпретировать их. Выбор анализируемых программой индикаторов основывался на изучении ряда исследований по теме [1–5, 7, 14]. В частности, отбор индикаторов осуществлялся с акцентом на их значимость и доказанную эффективность в рамках практического анализа феномена.

При подсчете средних показателей индикаторов по корпусам применялись взвешенные (weighted) методы усреднения, поскольку в анализе использовались полнотекстовые версии научных статей, в связи с чем длины вошедших в корпус текстов варьировались. Так, применялся один из следующих методов усреднения:

1) для индикаторов, имеющих тенденцию к более симметричному распределению по текстам, использовалась формула взвешенного арифметического среднего:

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \quad (1)$$

где \bar{x} – взвешенное среднее арифметическое значение показателя, x_i – значение показателя для i -го текста, w_i – вес i -го текста (в токенах), n – общее количество текстов в корпусе;

2) для индикаторов с более выраженной асимметрией распределения по текстам и возможным наличием выбросов, способных исказить среднее значение, применялась формула взвешенной медианы (weighted median), минимизирующая влияние выбросов и обеспечивающая более устойчивый показатель центральной тенденции:

$$M_w = X_k, \text{ где } \sum_{i=1}^{k-1} w_i < \frac{1}{2} \sum_{i=1}^n w_i \leq \sum_{i=1}^k w_i, \quad (2)$$

где M_w – взвешенное медианное значение показателя, X_k – значение показателя, соответствующее медиане, w_i – вес i -го текста (в токенах), n – общее количество текстов в корпусе.

Перейдем к подробному рассмотрению каждой из используемых в анализе характеристик, а также приведем их формальные индикаторы.

⁴ Журнал социологии и социальной антропологии. URL: <https://www.jourssa.ru/> (дата обращения: 04.09.2024); Журнал «Социологическое обозрение». URL: <https://sociologica.hse.ru/> (дата обращения: 04.09.2024).



Индикаторы характеристики упрощение

Коэффициент лексического разнообразия (*lexical variety*) [3–5, 7]. Считается, что переводы уступают непереводам по этому показателю [3, с. 107]. Рассчитывались три разновидности коэффициента:

$$TTR = \frac{V}{N} \times 100 \text{ – базовая версия коэффициента;} \quad (3)$$

$$\log TTR = \frac{\log V}{\log N} \times 100 \text{ – версия, сглаживающая зависимость показателя от длины текста;} \quad (4)$$

$$\text{ModifiedTTR} = \frac{100 \times \log N}{1 - \frac{V_1}{V}} \text{ – версия, учитывающая слова, встречающихся в тексте один раз}$$

(так называемая *hapax legomena*), (5)

где V – общее число различных слов (объем словаря), N – общее количество словоформ (токенов) в тексте (объем текста), а V_1 – количество *hapax legomena* (слов с частотой, равной 1).

Лексическая плотность (*lexical density*) – отношение числа слов-знаменательных частей речи ко всему объему текста (в словарных токенах) [3–5, 7]. Предполагается, что в переводах показатель ниже, так как в них содержится меньше знаменательных и больше служебных частей речи.

Средняя длина слов (*mean word length*) в символах и в слогах [3, 7]. Ожидается, что в переводах используются более короткие слова.

Средняя длина предложений (*mean sentence length*) в символах и в словах [3–5]. Считается, что часто в переводе длинные и сложные предложения оригинала разбиваются на более короткие и простые [3, с. 110].

Средний ранг слов (*mean word rank*) вычисляется на основе ранжирования слов текста по частоте их встречаемости в национальном корпусе русского языка [3]. Для этого используется частотный словарь русской лексики⁵. Предполагается, что переводы содержат меньше уникальной и больше более частотной лексики. Индикатор вычисляется по формуле:

$$MWR = \frac{\sum_{i=1}^n R_i}{n}, \quad (6)$$

где R_i – ранг i -го слова в тексте, n – общее количество словарных токенов. Ранг слова равен его порядковому номеру в частотном списке слов.

Как и в [3], при подсчете используются первые 6000 слов списка. Максимальный ранг слова равен 5000. Для слов, не вошедших в этот список, применяются два подхода:

- 1) слову присваивается максимальный ранг 6000;
- 2) слово исключается из анализа и не учитывается при вычислении ранга.

Индикаторы характеристики нормализация

Повторяемость знаменательных слов (*repetition*) [3, 4, 14] вычисляется как отношение абсолютной частоты слов, встречающихся более одного раза, к общему количеству словарных токенов. Считается, что переводчики склонны к избеганию лексических повторов, причиной чему может служить представление о том, что разнообразие словарного запаса является отражением культурного уровня автора текста и что чрезмерная повторяемость используемой лексики может восприниматься читателем как показатель «бедности» языка – это заставляет переводчиков заменять

⁵ Ляшевская О.Н., Шаров С.А. Новый частотный словарь русской лексики. М.: Азбуковник, 2009. 1090 с. URL: <https://publications.hse.ru/pubs/share/folder/jrd78un2xg/69157696.pdf> (дата обращения: 13.01.2024).



повторы синонимами, словосочетаниями или вовсе убирать повторяющиеся слова из текста [19, с. 77].

Поточечная взаимная информация (Pointwise Mutual Information – PMI) [3, 20] отражает силу ассоциативной связи между двумя соседствующими словами (биграммami слов) – то, насколько чаще два слова встречаются вместе, чем можно было бы ожидать, если бы их появления были независимы друг от друга. Слова предварительно лемматизируются. PMI для биграмм вычисляется по формуле (7):

$$PMI(w_1, w_2) = \log \left(\frac{P(w_1, w_2)}{P(w_1)P(w_2)} \right), \quad (7)$$

где $P(w_1, w_2)$ – вероятность совместного появления слов, $P(w_1)$ и $P(w_2)$ и – вероятности их появления по-отдельности.

Показатель рассчитывается с учетом всех текстов корпуса, общее количество биграмм с $PMI > 0$ нормализуется, т.е. делится на общее количество всех уникальных биграмм.

Индикаторы характеристики экспликация

Эксплицитное называние (explicit naming) – отношение числа личных / притяжательных местоимений к числу именованных сущностей [3]. Считается, что в переводах используется больше местоимений для облегчения идентификации именованных сущностей, о которых идет речь [14, с. 243].

Единичное именование (single naming) – соотношение числа именованных сущностей, состоящих из одного токена, ко всем именованным сущностям в тексте [3]. Предполагается, что в переводах предпочтение отдается полным версиям имен (например, *Йоганн Вольфганг фон Гёте*), тогда как в непереводах чаще используется вариант из одной лексемы (например, *Гёте*).

Средняя длина именованных сущностей (mean named entities length) – среднее количество токенов, составляющих именованные сущности [3]. Данный показатель связан с предыдущим: ожидается, что в переводах чаще используются более длинные варианты имен собственных.

Характеристика *экспликации* была бы неполной без учета *дискурсивных маркеров (discourse markers)* (ДМ), способствующих повышению когезии и когерентности текста и, как следствие, повышающих его эксплицитность [2–5]. Так, в анализе подсчитываются нормализованные частоты двух групп дискурсивных маркеров русского языка [21, с. 160]:

1) ДМ связи текста на локальном уровне (когезия предложений и их частей) – маркеры введения новой информации, пояснения, противопоставления, введения примеров;

2) ДМ связи текста на глобальном уровне (когерентность разных фрагментов текста друг другу, связь текста как единого целого – маркеры введения в тему, порядка следования информации, расположения иллюстративного материала, внутритекстовой референции).

Для каждого маркера подсчитывается нормализованная частота, т.е. абсолютная частота, деленная на общее количество токенов в тексте, что позволяет нивелировать влияние объема текста и обеспечивает корректное сравнение между текстами разной длины по данному показателю. Если ДМ состоит из нескольких токенов (например, *в то же время*), в анализе он считается за один токен.

Индикаторы характеристики интерференция

Анализ индикаторов данной характеристики нацелен, в отличие от анализа предыдущих трех характеристик, на определение влияния исходного языка на текст перевода. В анализ вошли следующие индикаторы:

Частеречные n-граммы (POS uni-, bi-, trigrams) представляют собой частеречные последовательности из одного, двух или трех токенов, анализируемые с целью выявления характерных для



текста перевода частеречных паттернов [1, 3, 4, 7]. Теги, используемые в анализе для обозначения частей речи, взяты из библиотеки *rumorphy2*⁶. Дополнительно вводятся специальные обозначения *S_START* (начало предложения), *S_END* (конец предложения) и *КОММА* (запятая) для анализа *n*-грамм, стоящих на первой, последней позициях в предложении и в начале придаточных клауз. Рассчитываются нормализованные частоты *n*-грамм, т.е. отношение абсолютной частоты определенной *n*-граммы к общему количеству *n*-грамм в тексте.

Символьные n-граммы (character uni-, bi-, trigrams) – последовательности из одного, двух или трех буквенных символов, нормализованные по частоте их встречаемости относительно общего числа буквенных *n*-грамм в тексте [3]. Индикатор используется для выявления морфологических особенностей слов, используемых в переводных текстах. Дополнительно вводятся специальные символы < и > для обозначения начала и конца слов соответственно, они позволяют учитывать границы слов и идентифицировать морфологические структуры, используемые в начале и конце слов (префиксы, суффиксы и окончания).

Остальные индикаторы переводческого языка

Кроме того, производился подсчет индикаторов, не вошедших ни в одну из предложенных характеристик [3], но позволяющих расширить спектр анализа и выявить те закономерности перевода, которые не охватываются рассмотренными выше индикаторами:

Использование местоимений разных разрядов (pronouns) – нормализованные частоты личных, притяжательных, относительных, неопределенных, отрицательных, указательных и определенных местоимений относительно всех токенов.

Использование различных знаков препинания (punctuation) – нормализованные частоты знаков препинания относительно всех токенов и всех знаков препинания в тексте.

Соотношение глаголов в страдательном залоге (passive verbs ratio) ко всем глаголам в тексте.

Индекс удобочитаемости Флэша (the Flesch readability score) для русского языка⁷ отражает уровень сложности текста с точки зрения его восприятия (чем ниже показатель, тем более сложным считается текст):

$$K = 206,836 - 60,1 \times W - 1,3 \times S, \quad (8)$$

где *K* – сложность текста, *W* – средняя длина слов (в слогах), *S* – средняя длина предложений (в словах).

Там, где в целях анализа индикаторов требовались лемматизация, частеречное тегирование и распознавание именованных сущностей, в анализаторе для этих задач использовались библиотеки для обработки естественного языка – *rumorphy2*⁸ и *Natasha*⁹.

Результаты исследования и их обсуждение

Рассмотрим полученные в ходе исследования результаты по каждой из проанализированных характеристик, сопоставляя и интерпретируя полученные данные по корпусам.

Результаты анализа индикаторов характеристики *упрощение* показали более низкие значения для всех индикаторов в переводных текстах (табл. 1). Для оценки стабильности средних значений индикаторов *упрощения* вычислялся коэффициент вариации (the coefficient of variation – CV)¹⁰, отражающий степень колеблемости средних значений индикаторов между текстами в каждом из

⁶ Обозначения для грамем (русский язык) // Морфологический анализатор *rumorphy2*. URL: <https://rumorphy2.readthedocs.io/en/stable/user/grammemes.html> (дата обращения: 22.08.24).

⁷ Оборнева И.В. Автоматизированная оценка сложности учебных текстов на основе статистических параметров: дисс. ... канд. пед. наук; 13.00.02. М., 2006. С. 47.

⁸ Морфологический анализатор *rumorphy2*. URL: <https://rumorphy2.readthedocs.io/en/stable/> (дата обращения: 26.03.2025).

⁹ Описание проекта // *natasha* 1.6.0. URL: <https://pypi.org/project/natasha/> (дата обращения: 26.03.2025).

¹⁰ Коэффициент вариации (Variation coefficient) // Вики. URL: <https://wiki.loginom.ru/articles/variation-coefficient.html> (дата обращения: 15.03.25).



подкорпусов. Коэффициент вариации рассчитывался как отношение стандартного квадратического отклонения к взвешенному среднему арифметическому. Принято считать, что если коэффициент вариации не превышает 33%, то выборочная совокупность является однородной и величина исследуемого признака может считаться надежной характеристикой этой совокупности. Как видно из табл. 1, для всех индикаторов значения колеблемости не превышают порога в 33%, что свидетельствует о достаточно стабильном распределении показателей по корпусам.

Таблица 1. Средние показатели индикаторов упрощения по корпусам
Table 1. Average indicators of simplification indicators by buildings

Показатель	Непереводные тексты	Переводы	Разница	CV (Непереводные тексты)	CV (Переводы)
Lexical density	64%	61%	3%	3,5%	4,5%
TTR	28,18%	26,94%	1,24%	15%	18%
Log TTR	85,25%	84,76%	0,49%	1,8%	1,9%
Modified TTR	1995,64	1921,57	74,07	8,6%	12,7%
Mean word length					
(1) characters	6,9	6,7	0,2	3%	4,4%
(2) syllables	2,95	2,89	0,06	3,3%	4%
Mean sentence length					
(1) tokens	30,1	25,6	4,5	28%	11%
(2) characters	176	148	28	28%	13%
Mean word rank 1	1831,53	1689,3	142,23	8%	13%
Mean word rank 2	886,29	847,72	38,57	6%	9%

Результаты анализа подтверждают действие упрощения на морфологическом, лексическом и синтаксическом уровнях. Примечательно, что для корпуса непереводных текстов коэффициент вариации средней длины предложений (в токенах и в символах) приближается к пороговому значению и значительно превосходит аналогичный показатель у переводных текстов. Вероятно, это свидетельствует о большей вариативности длины предложений в непереводных текстах, что может быть связано с большим разнообразием используемых синтаксических конструкций. В то время как в переводных текстах длина предложений часто более стандартизирована, что, вероятно, вызвано стремлением избежать чрезмерно сложных или стилистически разнообразных синтаксических структур.

Среди индикаторов особенно выделяются *модифицированный коэффициент лексической вариативности* и показатель *среднего ранга слов*. Показатель *лексической вариативности* оказался существенно ниже в переводах, что указывает на стремление переводчиков избегать использования лексических единиц, встречающихся в тексте лишь единожды (*hapax legomena*). Значительный разрыв между корпусами по *показателю среднего ранга слов*, в свою очередь, свидетельствует о снижении уникальности используемой в переводах лексики и о предпочтении переводчиками более частотных вариантов лексем (например, выбор лексемы *сложный* с рангом 754 вместо менее частотной лексемы *трудоемкий* с рангом 6000). Следует отметить, что такое снижение количества *hapax legomena* и среднего ранга слов может стать причиной сглаживания стилистических особенностей текста, заложенных автором оригинальной статьи, и потери его оригинальности.

Сопоставление полученных показателей индикаторов нормализации продемонстрировало снижение *лексической повторяемости* в переводах (40,13% в непереводных текстах против 38,4% в переводах), что свидетельствует в пользу *нормализации*. Потенциальным следствием такого снижения является ослабление смысловых связей между частями текста, для обеспечения которых в оригинальном тексте употреблялись опущенные в переводе лексемы. Еще одним



следствием может стать утрата смысловых акцентов, намеренно расставленных автором оригинальной статьи посредством повторного употребления той или иной лексемы.

Нормализованная частота биграмм с PMI > 0, вопреки прогнозам, была почти одинаковой по корпусам: 97,3% для непереводных текстов против 97% для переводов.

Перейдем к рассмотрению результатов анализа *экспликации*. В отличие от индикаторов *упрощения*, относящихся к более общим, глобальным характеристикам текста и имеющим более равномерное распределение по корпусу, индикаторы *экспликации* подразумевают учет языковых единиц, выбор и частота использования которых в значительной мере зависят от таких факторов, как контекст и тематическая соотнесенность текста (в случае с именованными сущностями), сознательное намерение автора оригинала использовать определенные языковые единицы с целью выразить собственное отношение к обсуждаемому предмету, усилить определенные логические связи, привлечь внимание читателя к чему-либо и т.д. (в случае с дискурсивными маркерами). Все это может стать причиной менее равномерного распределения анализируемых языковых единиц между текстами в корпусе и, как следствие, увеличить вероятность возникновения выбросов – значений, значительно отклоняющихся от общей тенденции. В связи с этим при расчете средних показателей индикаторов было решено использовать взвешенную медиану, а коэффициент вариации рассчитывался как отношение взвешенного медианного абсолютного отклонения к взвешенному медианному значению. Результаты анализа характеристики *экспликация* приведены в табл. 2.

Таблица 2. Средние (медианные) показатели индикаторов экспликации по корпусам
Table 2. Average (median) indicators of explication by corpus

Показатель	Непереводные тексты	Переводы	Разница	CV (Непереводные тексты)	CV (Переводы)
Explicit naming	168,5%	248%	49,5%	47,2	51
Single naming	61%	71%	10%	23	20
Mean named entity length	1,4 токена	1,3 токена	0,1 токена	13	11
ДМ пояснения	0,37%	0,5%	0,13%	25	21
ДМ противопоставления	1,04%	1,15%	0,11%	21	10
ДМ порядка следования информации	0,2%	0,186%	0,014%	42	17
ДМ выводов	0,4%	0,32%	0,08%	18	32

Согласно показателям *эксплицитного называния*, переводам свойственно более частое использование личных и притяжательных местоимений для идентификации имен собственных. При этом коэффициент вариации показателей индикатора указывает на то, что оба подкорпуса демонстрируют по нему высокую и вместе с тем схожую степень вариативности. Как упоминалось ранее, такая вариативность может быть обусловлена тем, что частота использования именованных сущностей, используемых при подсчете показателей *эксплицитного называния*, во многом зависит от контекста, тематики статьи и авторских предпочтений. Результаты анализа *единичного именованного* и *средней длины именованных сущностей* не подтвердили прогнозов – в переводах чаще используются именованные сущности, состоящие из одного токена, а средняя длина именованных сущностей по корпусам почти не различается.

Что касается дискурсивных маркеров, обеспечивающих когерентность частей текста друг другу и связь текста как единого целого, то, вопреки ожиданиям, результаты анализа показали снижение частоты использования в переводах *дискурсивных маркеров порядка следования информации* (маркеры типа *прежде всего, во-первых, с одной стороны*) и выводов (маркеры типа *таким*



образом, следовательно, *итак*). Такое снижение может стать причиной ослабления логических связей между частями текста и затруднения восприятия приводимых в статье аргументов и результатов исследования. Следует отметить, что коэффициент вариации *дискурсивных маркеров порядка следования информации* в непереводном корпусе значительно превысил пороговое значение (33%), что говорит о неравномерном распределении маркеров этого типа между непереводными текстами. Коэффициент вариации *дискурсивных маркеров выводов* в переводах составил приближающееся к пороговому значение, которое, к тому же, заметно превышает то же значение в непереводных текстах, что свидетельствует о большей вариативности в использовании этих маркеров в переводе. В этом контексте перспективным направлением дальнейших исследований представляется расширение корпуса и привлечение дополнительных текстов, что позволит более точно установить закономерности использования маркеров этих типов.

В то же время результаты анализа подтвердили наличие в переводах характеристики *экспликация* по показателям *дискурсивных маркеров пояснения* (маркеры типа *потому что, иными словами*) и *противопоставления* (маркеры типа *хотя, тем не менее*), обеспечивающим когезию между предложениями / их частями, что может свидетельствовать о повышении объяснительного потенциала текста.

Перейдем к рассмотрению результатов анализа индикаторов *интерференции* по корпусам. Данные об использовании *частеречных n-грамм*, продемонстрировавших наибольшие различия, приведены в табл. 3.

Таблица 3. Нормализованные частоты частеречных n-грамм по корпусам
Table 3. Normalized frequencies of partial n-grams by case

<i>n</i> -граммы	Непереводные тексты (%)	Переводы (%)	Разница (%)
NOUN	31,4	29,6	1,8
ADJF	15,4	14,6	0,8
CONJ	7,8	8,6	0,8
NPRO	2,2	2,7	0,5
VERB	5,9	6,8	0,9
NOUN, NOUN	6,96	6,23	0,73
NOUN, ADJF	3,9	3,1	0,8
NOUN, PREP	3,7	3,3	0,4
COMMA, PRTF	0,8	0,6	0,2
COMMA, CONJ	2,58	2,96	0,38
NPRO, VERB	0,5	0,9	0,4
S_START, CONJ	0,55	0,75	0,2
CONJ, NPRO	0,31	0,46	0,15
S_START, NPRO	0,13	0,27	0,14
NOUN, ADJF, NOUN	2,96	2,3	0,66
NOUN, NOUN, PREP	0,9	0,67	0,23
NOUN, NOUN, NOUN	1,47	1,25	0,22
NOUN, COMMA, PRTF	0,71	0,56	0,15
VERB, ADJF, NOUN	0,72	0,84	0,12
CONJ, NPRO, VERB	0,1	0,2	0,1
CONJ, NOUN, VERB	0,19	0,27	0,08
VERB, NOUN, PREP	0,24	0,28	0,04
VERB, NOUN, NOUN	0,4	0,42	0,02

Из таблицы видно, что текстам переводов свойственно более частое использование таких *n*-грамм, как (*CONJ*), (*COMMA, CONJ*) и (*NPRO*), что может быть следствием *экспликации*. *n*-



граммы (*CONJ*) и (*COMMA, CONJ*), вероятно, говорят о более частом использовании в переводах придаточных предложений, в то время как *n*-грамма (*NPRO*) – о более частом употреблении местоименных существительных, часто используемых для указания на существительные-антецеденты в тексте перевода.

В целях иллюстрации других выявленных различий рассмотрим некоторые примеры *частеречных триграмм*, взятые из проанализированных подкорпусов, и приведем возможные причины этих различий (табл. 4).

Таблица 4. Примеры триграмм и причины различий в их частотах
Table 4. Examples of trigrams and the reasons for differences in their frequencies

<i>n</i> -граммы	Примеры	Частота	Причина
CONJ, NPRO, VERB	<i>поскольку она играет</i>	Выше в переводе из-за интерференции	В АЯ прямой порядок слов в утверждениях, почти полное отсутствие глагольных окончаний, вследствие чего наличие подлежащего обязательно
CONJ, NOUN, VERB	<i>поскольку ценность приобретается</i>		
VERB, NOUN, NOUN	<i>устанавливается новая связь (вместо установление новой связи)</i>	Выше в переводе из-за интерференции	Академическому АЯ свойственен более активный стиль изложения, из-за чего при описании чаще используются глаголы, нежели имена существительные
VERB, NOUN, PREP	<i>проявляется склонность к (вместо проявление склонности к)</i>		
NOUN, NOUN, NOUN	<i>описание данных наблюдения (в переводе чаще используются конструкции типа описываются данные наблюдения)</i>	Ниже в переводе из-за интерференции	См. причину выше
NOUN, NOUN, PREP	<i>адаптация условий для (в переводе чаще используются конструкции типа адаптировать условия для)</i>		
VERB, ADJF, NOUN	<i>проявлялись культурные образцы (вместо проявление культурных образцов)</i>	Выше в переводе из-за интерференции	См. причину выше
NOUN, ADJF, NOUN	<i>появление взаимного доверия (в переводе чаще используются конструкции типа появляется взаимное доверие)</i>	Ниже в переводе из-за интерференции	См. причину выше
NOUN, COMMA, PRPF	<i>функция, обеспечивающая (в переводе чаще используются конструкции типа функция, которая обеспечивает)</i>	Ниже в переводе из-за интерференции	В АЯ конструкции причастиями настоящего времени (<i>participial phrases</i>) используются реже, чаще предпочтение отдается придаточным определительным / распространительным

Рассмотрим некоторые примеры буквенных *n*-грамм, наиболее информативными являются триграммы (табл. 5), так как они позволяют захватывать морфемы слов целиком.

Анализ демонстрирует снижение частотности окончаний *-ой*, *-го* и *-ого* в переводах, что свидетельствует о менее частом употреблении форм родительного падежа прилагательных. Это явление может быть объяснено интерференцией английского языка, являющегося, как показано в табл. 4, более динамичным, в результате чего фразы с отглагольными существительными типа *появление взаимного доверия* с большей вероятностью будут передаваться в переводе с использованием глагола (ср. с *появляется взаимное доверие*), устраняющего необходимость употребления



родительного падежа. О снижении использования в переводе отглагольных существительных также говорит и уменьшение частоты триграмм типа *-ция / -ции* и *-ния / -ние*, часто используемых в качестве суффиксов в отглагольных существительных.

Таблица 5. Нормализованные частоты буквенных *n*-грамм
Table 5. Normalized letter *n*-gram frequencies

3-граммы	Непереводные тексты (%)	Переводы (%)	Разница (%)
о, й, >	0,367	0,269	0,098
г, о, >	0,4	0,33	0,07
о, г, о	0,31	0,265	0,045
н, и, я	0,275	0,259	0,016
ц, и, и	0,131	0,116	0,015
ц, и, я	0,053	0,048	0,005
н, и, е	0,205	0,195	0,01

Рассмотрим некоторые результаты анализа *триграмм со служебными словами* (табл. 6), в которых два токена заменены служебными словами, а один токен — соответствующим ему тегом.

Таблица 6. Нормализованные частоты 3-грамм со служебными словами
Table 6. Normalized 3-gram frequencies with service words

3-граммы	Непереводные тексты (%)	Переводы (%)	Разница (%)
в, том, СОММА	0,05	0,09	0,04
того, СОММА, чтобы	0,01	0,05	0,04
VERB, в, том	0,023	0,06	0,037
для, того, СОММА	0,007	0,034	0,027
то, СОММА, что	0,06	0,084	0,024
том, СОММА, что	0,08	0,1	0,02

Представленные данные свидетельствуют о более частом использовании в переводе элементов, вводящих придаточные предложения, которые служат для повышения *экспликации* какого-либо содержания текста. Эти данные согласуются с представленными ранее результатами анализа *дискурсивных маркеров пояснения*, также преобладающих в переводных текстах и используемых с такой же целью.

Перейдем к анализу *позиционной частоты токенов* в предложениях по корпусам. В табл. 7 представлены данные о токенах на первой, второй, третьей с конца и предпоследней позициях в предложении — приводятся токены, абсолютная взвешенная частота которых больше или равна единице. Последняя позиция из анализа исключается, так как используется исключительно для определения конца предложений. Токены распределены по своей частеречной принадлежности, в скобках указаны их нормализованные частоты.

Значительное увеличение в переводах частоты использования союзов на первой позиции, вероятно, свидетельствует в пользу *экспликации*, проявляющейся в стремлении переводчиков сделать логические и причинно-следственные связи между частями текста более явными. Более высокая частота противительных союзов, в свою очередь, может быть признаком *интерференции*, вызванной частым использованием в английском языке на первой позиции таких союзов, как *however, although, nevertheless*.

Более частое употребление указательных, личных и притяжательных местоимений на первой, второй и третьей позициях может быть следствием *экспликации*, которая выражается в стремлении переводчиков сделать отсылку к ранее упомянутому в тексте существительным. Увеличение



Таблица 7. Нормализованные частоты токенов на разных позициях (по убыванию частот)
Table 7. Normalized token frequencies at different positions (in descending order of frequencies)

Непереводные тексты (%)			
Первая	Вторая	Третья с конца	Предпоследняя
Предлоги (15,57)	Пунктуация (9,2)	Союзы (6,3)	Пунктуация (5,5)
В (9,4), По (1,2), При (1,14), С (1), Для (0,93), На (0,9), К (0,54), Кроме (0,46)	, (9,2)	и (5,7), или (0,6)) (5,5)
Союзы (7,74)	Предлоги (4,9)	Предлоги (6,15)	Им. сущ. (3,47)
Однако (2,1), Но (1,6), И (1,3), Если (1,24), Поэтому (1), А (0,5)	в (2,2), из (0,9), на (0,8), с (0,6), к (0,4)	в (2,6), на (1,05), к (1), с (0,54), для (0,48), от (0,48)	времени (0,54), труда (0,53), общества (0,5), производства (0,43), жизни (0,41), мира (0,4), действия (0,36), власти (0,3)
Местоимения (9,39)	Местоимения (5,57)	Пунктуация (1,74)	
Это (1,7), Таким (1,45), Он (0,8), Эти (0,67), Этот (0,65), Эта (0,64), Все (0,63), Иными (0,6) Тем (0,57) Они (0,5), Мы (0,44), Она (0,4), Что (0,34)	этом (1,7), это (1,2), того (0,9), то (0,45), этой (0,4), этого (0,4), ее (0,4), его (0,3)	, (1,34),) (0,4)	
Наречия (4,57)	Частицы (2)	Частицы (1,34)	
Как (1,7), Так (1,25), Именно (0,65), Здесь (0,47), Более (0,5)	не (1,2), же (0,8)	не (1,34)	
Вводные слова (1,26)	Им. сущ. (2,67)	Местоимения (1,84)	
Во-первых (0,4), Во-вторых (0,4), Например (0,46)	образом (1,5), словами (0,67), результате (0,5)	его (1,04), их (0,8)	
Предикатив (0,43)	Союзы (1,4)	Им. прил. (1,74)	
Можно (0,43)	и (1,4)	социальной (0,76), социального (0,52), социальных (0,46)	
Частицы (0,4)	Наречия (0,4)		
Не (0,4)	как (0,4)		
Переводы (%)			
Первая	Вторая	Третья с конца	Предпоследняя
Предлоги (15,65)	Пунктуация (8,8)	Союзы (7,88)	Пунктуация (2,5)
В (8,9), С (1,3), На (1), Для (0,94), По (0,88), Кроме (0,6), При (0,5), К (0,5), Несмотря (0,35), Согласно (0,34), Во (0,34)	, (8,8)	и (7,13), или (0,75)) (2,5)
Союзы (11,95)	Предлоги (4,11)	Предлоги (6,26)	Им. сущ. (2,55)
Но (3,2), Однако (2,4), Если (1,85), И (1,37), Хотя (0,78), А (0,53), Поэтому (0,5), Поскольку (0,46), Чтобы (0,46), Когда (0,4)	в (1,9), из (1,03), на (0,78), с (0,4)	в (2,65), к (1,05), на (1), с (0,96), для (0,6)	жизни (0,6), общества (0,5), сообщества (0,4), политики (0,38), деятельности (0,37), организации (0,3)
Местоимения (13,24)	Местоимения (7,04)	Пунктуация (0,7)	
Это (3,05), Они (1,7), Он (1,14), Таким (1,13), Эти (1,1), Мы (0,85), Я (0,82), Эта (0,7), Этот (0,6), Все (0,6), Тем (0,6), Что (0,55), Она (0,4)	это (1,3), того (1), этом (0,9), мы (0,75), то (0,6), этого (0,56), этой (0,5), эти (0,4), они (0,37), другой (0,34), что (0,32)	, (0,7)	



Окончание таблицы 7

Наречия (3,73)	Частицы (1,95)	Частицы (1)	
Как (1,4), Так (0,75), Именно (0,6), Здесь (0,55), Более (0,43)	не (1,25), же (0,7)	не (1)	жизни (0,6), общества (0,5), сообщества (0,4), политики (0,38), деятельности (0,37), организации (0,3)
Вводные слова (0,82)	Им. сущ. (1,56)	Местоимения (2,5)	
Например (0,56), Итак (0,26)	образом (1,2), словами (0,36)	их (1,22), его (0,72), ее (0,56)	
Частицы (0,45)	Союзы (2,3)	Им. прил. (1,73)	
Не (0,45)	и (1,4), также (0,5), если (0,4)	социальных (0,76), социальной (0,5), социального (0,47)	

частоты использования местоимений на первой позиции может также быть связано с *интерференцией* английского языка с его обязательным формальным наличием подлежащего в предложении – это приводит к тому, что местоимения, которые могли бы быть опущены в русскоязычном тексте, в переводе сохраняются. Дополнительным фактором здесь служит тенденция современной англоязычной научной литературы к использованию активного стиля изложения¹¹, особенно если речь идет о таких разделах научной статьи как «Введение», «Обсуждение» и «Результаты» – в них чаще предпочтение отдается активным формам глаголов, так как это позволяет читателю легко идентифицировать деятеля, что снижает риск двусмысленности и повышает уровень ответственности авторов за излагаемые в статье положения и результаты. Также в английской научной литературе нормой является использование местоимений *we* или *I* для указания автора исследования в качестве деятеля, в то время как русскоязычной литературе такая практика менее распространена, а порой и вовсе не приветствуется.

Кроме того, из табл. 7 видно, что в непереводных текстах закрывающие скобки на предпоследней и третьей с конца позициях встречаются гораздо чаще, чем в переводах, что, вероятно, говорит о более частом использовании в русскоязычной научной литературе скобок для указания каких-либо уточнений / дополнительной информации, при этом без нарушения структуры основного текста. В переводах же более частое употребление союзов *and* / *or* или *and* перед последним элементом перечисления, в то время как русскоязычный текст в данном случае нередко обходится запятой, что также подтверждается данными из таблицы.

Наконец, перейдем к рассмотрению индикаторов, не вошедших ни в одну из характеристик. Из табл. 8 видно, что переводам свойственно более частое использование некоторых личных, притяжательных и указательных местоимений, что еще раз подтверждает действие характеристики *экспликации*.

Более частое использование местоимения *on* в непереводных текстах связано с тем, что в английском языке, в отличие от русского, у существительных отсутствует грамматическая категория рода, и что в случае, если в оригинале используются местоимения *he* или *she*, в переводе они однозначно будут переданы как *он* или *она*, в то время как местоимение *it* используется для обозначения абстрактных понятий, неодушевленных объектов или животных, а соответствующие им существительные в русском языке могут быть любого рода, в том числе и мужского, из-за чего нередко *it* переводится как *он*. Что касается местоимения *so*, оно не имеет прямого аналога в английском языке, что делает его использование более характерным для непереводных текстов.

Рост использования относительных местоимений, выполняющих функцию союзных слов, которые вводят придаточные предложения, свидетельствует об *экспликации*.

¹¹ Plotnick. J. How to Use Active Voice in the Sciences // The Founding College of the University of Toronto. URL: <https://www.uc.utoronto.ca/active-voice-in-science> (дата обращения: 22.12.24).



Таблица 8. Нормализованные частоты местоимений по разрядам
Table 8. Normalized frequencies of pronouns by digits

Местоимение	Непереводные тексты (%)	Переводы (%)	Разница (%)
Личные местоимения			
они	0,627	0,962	0,335
мы	0,225	0,404	0,179
я	0,061	0,173	0,112
она	0,349	0,375	0,026
вы	0,002	0,016	0,014
он	0,686	0,574	0,112
Притяжательные местоимения			
свой	0,434	0,47	0,036
наш	0,087	0,118	0,031
мой	0,008	0,033	0,025
Указательные местоимения			
это	0,638	0,795	0,157
этот	0,491	0,63	0,139
тот	0,577	0,693	0,116
такой	0,374	0,419	0,045
то	0,313	0,358	0,045
столько	0,02	0,01	0,01
Определительные местоимения			
другой	0,28	0,4	0,12
сам	0,26	0,16	0,10
иной	0,12	0,06	0,06
Относительные местоимения			
который	0,84	1,05	0,21
что	0,9	1,04	0,14
чем	0,1	0,15	0,05
какой	0,07	0,1	0,03
Неопределенные местоимения			
некоторый	0,05	0,13	0,08
несколько	0,04	0,06	0,02

Более высокая частота местоимений *другой*, *некоторый* и *несколько* в переводах может быть объяснена *интерференцией* английского языка, которому свойственно очень частое использование лексем *other / another / the other* и *some*.

Сравнение частот некоторых *пунктуационных знаков*, нормализованных относительно всех знаков препинания, приведено в табл. 9.

Более частое употребление запятых, нередко использующихся для введения придаточных предложений, вероятно, является показателем *экспликации*, а рост частоты употребления точек подтверждает наличие синтаксического *упрощения*. Что касается увеличения частоты использования точек с запятой в переводах, это может быть связано с *интерференцией* – в английских текстах точку с запятой зачастую используют для соединения двух и более предложений, сохраняя их относительную автономность, но при этом подчеркивая тесную смысловую связь¹². Для примера приведем предложение, взятое из оригинала одного из проанализированных переводов: *If the*

¹² Semicolons, colons, and dashes // The Writing Center. The University of North Carolina at Chapel Hill. URL: <https://writingcenter.unc.edu/tips-and-tools/semi-colons-colons-and-dashes/> (дата обращения: 22.12.24).



argument about the current postsocial transition is right, such extensions will be needed in several respects; to make them is perhaps the major challenge confronting social theory today.

Увеличение использования в переводах тире, вероятно, также является следствием *интерференции* и связано с особенностями их употребления в оригинальных англоязычных текстах. В английском языке чаще, чем в русском, тире используется для акцентирования внимания на важных аспектах предложения или для пояснения какого-то элемента в нем. Для примера приведем следующее предложение: *I now need to sketch out what an expanded conception of sociality that includes (but is not limited to) material objects – which I will call an object-centered sociality – is about.* В данном варианте тире используется для разъяснения термина, в то время как в русском языке аналогичная по смыслу структура часто оформляется с помощью запятых или же вовсе практикуется разбиение предложения на несколько более простых.

Таблица 9. Нормализованные частоты знаков препинания
Table 9. Normalized punctuation mark frequencies

Знак	Непереводные тексты (%)	Переводы (%)	Разница (%)
,	50,86	55,21	4,35
.	20,66	23,97	3,31
–	0,51	1,11	0,60
;	0,71	0,97	0,26

Что касается использования *глаголов в форме страдательного залога*, в переводах их процент несколько ниже (12,6% против 15%). Одной из причин этого, как уже указано, может служить тенденция англоязычных научных текстов к активному стилю изложения. Так, например, предложение *We propose a new method for analyzing...* в русскоязычных непереводных статьях с большей вероятностью будет выглядеть как *В статье предлагается / предложен новый метод анализа...* Однако вопрос о том, следует ли в процессе постредактирования переводов намеренно заменять активные конструкции оригинала на пассивные, остается открытым, поскольку это может нарушить намерения автора оригинала, особенно если он осознанно стремился к персонализированному стилю.

Анализ *индекса удобочитаемости Флеша* показал, что переводные тексты сложнее непереводных для восприятия – для непереводных текстов индекс составил 5,8, а для переводных – 1,97. Это, вероятно, вызвано тем, что, хотя средняя длина слов и предложений в переводах становится несколько меньше, возрастающая тенденция к *экспликациям*, а именно к более частому использованию местоимений, союзов, а также придаточных предложений, может усложнять восприятие текста и влиять на значение данного показателя.

Заключение

Таким образом, результаты проведенного анализа подтверждают, что переводам научных статей с английского на русский язык свойственно наличие всех ключевых характеристик феномена *translationese*. Приведем выявленные характеристики феномена и их индикаторы, а также рассмотрим возможные пути преодоления их влияния на перевод:

Упрощение. Все индикаторы данной характеристики подтвердили ее влияние на переводы. Индикаторами, продемонстрировавшими наиболее существенные различия, оказались *модифицированный коэффициент лексического разнообразия* и *средний ранг слов*. Так, оба они указывают на лексическое упрощение текста и снижение уникальности используемой в переводах лексики, что может приводить к нивелированию стилистических особенностей, заложенных в тексте оригинала. Преодолеть это влияние, повысив уровень уникальности и разнообразие используемой в переводе лексики, можно несколькими способами:



1) обращаясь к словарям синонимов / антонимов, а также тезаурусам (к примеру, RuWordNet¹³), предлагающим широкий спектр лексических единиц, находящихся с интересующим переводчика / постредактора словом не только в отношениях анто- и синонимии, но также и гипо-, гипер-, меро- и холонимии;

2) используя частотный словарь русской лексики для выбора менее частотных вариантов слов (слов с бóльшим порядковым номером в списке).

Нормализация. Влияние данной характеристики подтвердилось по индикатору *лексической повторяемости знаменательных слов* – было выявлено снижение по этому показателю. Учет данного явления в переводе видится важным, поскольку, как говорилось выше, оно может повлечь ослабление логических связей между фрагментами текста и привести к снижению их понимания читателем, а также снизить важность идей, которые автор оригинального текста намеревался подчеркнуть, повторно используя опущенную в переводе лексику.

Экспликация. Было обнаружено, что данная характеристика проявляется в переводах через увеличение *показателя эксплицитного названия*, а именно частоты использования личных, притяжательных, указательных местоимений относительно имен собственных (рост частоты местоимений также подтверждается отдельным анализом местоимений и анализом *n*-грамм), а также через увеличение частоты *дискурсивных маркеров пояснения* и *противопоставления*, что свидетельствует о повышении связности и объяснительного потенциала текста. Вероятно, увеличение частоты этих маркеров необязательно ухудшает качество перевода, однако необходимы дальнейшие исследования, чтобы оценить влияние данного явления на восприятие качества перевода носителями русского языка.

В то же время было зафиксировано снижение частоты *дискурсивных маркеров*, организующих *порядок следования информации* и *выводы* статьи, что может привести к ослаблению понимания русскоязычным читателем логики аргументации и к тому, что выводы статьи станут менее очевидными.

Интерференция. Данная характеристика проявляется в переводе по всем проанализированным индикаторам. Так, анализ *частеречных n-грамм* продемонстрировал снижение триграмм (*NOUN, NOUN, NOUN*), (*NOUN, NOUN, PREP*) и (*NOUN, ADJF, NOUN*) и соответствующее этому снижению увеличение триграмм (*VERB, NOUN, NOUN*), (*VERB, NOUN, PREP*) и (*VERB, ADJF, NOUN*), связанное с более динамичной природой английского языка, где употребление глаголов более предпочтительно, чем использование их субстантивированных версий, что также подтверждается *анализом буквенных триграмм*. Также было выявлено, что переводам в меньшей мере свойственно использование триграммы (*NOUN, COMMA, PRTF*), обусловленное менее частым употреблением в английском языке фраз с причастиями настоящего времени (*participial phrases*), что приводит к снижению частоты причастных оборотов в русскоязычных переводах. Так, эти явления также необходимо учитывать, поскольку научный стиль русского языка отличается от английского научного стиля тем, что в первом гораздо чаще, чем во втором, предпочтение отдается отглагольным существительным вместо самих глаголов, а также чаще используются причастные обороты. Если при переводе не учитывать эти различия, вероятно, слишком частое использование глаголов (вместо существительных) и придаточных определительных (вместо причастий) может повлиять на восприятие текста русскоязычным читателем.

По результатам анализа *триграмм со служебными словами* было выявлено более частое использование в переводе придаточных конструкций, указывающее на повышение уровня эксплицитности.

Анализ индикатора *позиционной частоты токенов* показал увеличение использования союзов в начале предложений, вероятно, свидетельствующее о стремлении переводчиков более ясно выражать причинно-следственные связи в тексте (*экспликация*). Более частое использование

¹³ Тезаурус русского языка RuWordNet. URL: <https://www.ruwordnet.ru/ru> (дата обращения: 26.03.2025).



противительных союзов на первой позиции может быть результатом *интерференции*, так как в английском нередко союзы типа *however, although* стоят в начале предложения. Повышенная частота указательных, личных и притяжательных местоимений на первых двух и третьей с конца позициях, вероятно, также свидетельствует в пользу *экспликации*. А что касается увеличения частоты употребления личных местоимений на первой позиции, оно может быть вызвано интерференцией английского языка, выражающейся: 1) в необходимости (даже формального) присутствия подлежащего на первой позиции в утверждениях; 2) в тенденции к активному стилю изложения, т.е. к использованию глаголов в активном залоге, нередко сопровождающемуся употреблением местоимений *I* и *we* (если речь идет об авторе как деятеле) — данная тенденция отражает особенности современного английского научного дискурса, однако она в какой-то степени вступает в противоречие с нормами русскоязычной академической традиции и, вероятно, требует учета в процессе перевода / постредактирования.

Другие индикаторы. Анализ частот местоимений показал, что местоимения всех разрядов чаще встречаются в переводных текстах. Так, например, увеличение количества личных, притяжательных, указательных и относительных местоимений (последние используются как союзные слова для ввода придаточных предложений) подтверждает тенденцию к *экспликации*, на что стоит обращать внимание при переводе / постредактировании, поскольку избыточное употребление местоимений может приводить к усложнению текста, о котором также свидетельствует более низкий показатель *индекса удобочитаемости Флеша* для переводов. На *интерференцию* указывает значительно более частое использование в переводах местоимений *другой, некоторый* и *несколько*, аналогов очень частотных в английском языке слов *other, another, the other* и *some*. Вероятно, при переводе стоит контролировать частоту их использования, опуская их либо заменяя на другие лексемы (например, *другой* на *иной, еще один*).

Анализ *пунктуации* показал рост употребления точек с запятой и тире, что, вероятно, отражает *интерференцию* английской пунктуации, где данные знаки активно используются для соединения близких по смыслу предложений (точка с запятой) или для акцентирования / пояснения важных элементов текста (тире), что в меньшей мере характерно для норм русского языка.

Наконец, анализ *глаголов в форме страдательного залога* продемонстрировал снижение их процентного соотношения в тексте, что с большой долей вероятности является следствием упомянутой ранее тенденции более активного стиля изложения в англоязычном письменном дискурсе и что также необходимо учитывать при переводе / постредактировании, так как для русского научного стиля изложения характерно более частое использование страдательного залога.

Для более глубокого понимания выявленных характеристик необходимо проведение дальнейших исследований. Такие исследования позволят с большей точностью определить, какие из характеристик оказывают наиболее существенное влияние на качество перевода и должны быть учтены на этапе перевода и постредактирования, а какие менее значимы или вовсе не влияют на качество перевода. И в этом процессе использованный в представленном исследовании анализатор феномена *translationese* может сыграть ключевую роль, став инструментом систематического анализа и сопоставления корпусов текстов.

Перспективы дальнейшего исследования, помимо всего прочего, включают анализ (и включение в программу) других индикаторов феномена *translationese*, что обеспечит его более глубокий и детальный анализ. Кроме того, важным представляется увеличение объема исследовательского корпуса научных статей, что позволит повысить достоверность получаемых данных.



СПИСОК ИСТОЧНИКОВ

1. **Baroni M., Bernardini S.** A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text // *Literary and Linguistic Computing*. 2005. Vol. 21, Iss. 3. P. 259–274. DOI: 10.1093/llc/fqi039
2. **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.
3. **Volansky V., Ordan N., Wintner Sh.** On the features of translationese // *Digital Scholarship in the Humanities*. 2013. Vol. 30, Iss. 1. P. 98–118. DOI: 10.1093/llc/fqt031
4. **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, 2021. P. 133–180. DOI: 10.1007/978-981-16-4918-9_6
5. **Kunilovskaya M., Lapshinova-Koltunski E., Mitkov R.** Translationese in Russian Literary Texts // *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Punta Cana: Association for Computational Linguistics, 2021. P. 101–112. DOI: 10.18653/v1/2021.latechclfl-1.12
6. **Bizzoni Y., Lapshinova-Koltunski E.** Measuring Translationese across Levels of Expertise: Are Professionals more Surprising than Students? // *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*. Reykjavik; Linköping University Electronic Press, 2021. P. 53–63.
7. **Pylypenko D., Amponsah-Kaakyire K., Dutta Chowdhury K., van Genabith J., España-Bonet C.** Comparing Feature-Engineering and Feature-Learning Approaches for Multilingual Translationese Classification // *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Punta Cana: Association for Computational Linguistics, 2021. P. 8596–8611. DOI: 10.18653/v1/2021.emnlp-main.676
8. **Kunilovskaya M., Lapshinova-Koltunski E.** Lexicogrammatic Translationese across Two Targets and Competence Levels // *Proceedings of the 12th Conference on Language Resources and Evaluation*. Marseille: European Language Resources Association, 2020. P. 4102–4112.
9. **Ni J., Jin Z., Freitag M., Sachan M., Schölkopf B.** Original or Translated? A Causal Analysis of the Impact of Translationese on Machine Translation Performance // *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle: Association for Computational Linguistics, 2022. P. 5303–5320. DOI: 10.18653/v1/2022.naacl-main.389
10. **Lee Y.C.** Function Words as Markers of Translationese: A Corpus-based Approach to Mental Translation in Second Language Writing // *Korean Journal of English Language and Linguistics*. 2021. Vol. 21. P. 261–281. DOI: 10.15738/kjell.21..202103.261
11. **Gellerstam M.** Translationese in Swedish novels translated from English // *Translation Studies in Scandinavia* / ed. by L. Wollin, H. Lindquist. Lund: CWK Gleerup Publ., 1986. P. 88–95.
12. **Toury G.** Interlanguage and its Manifestations in Translation // *Meta*. 1979. Vol. 24, No. 2. P. 223–231. DOI: 10.7202/004502ar
13. **Frawley W.** Prolegomenon to a theory of translation // *Translation: Literary, Linguistic and Philosophical Perspectives* / ed. by W. Frawley. Newark: University of Delaware Press, 1984. P. 159–175.
14. **Baker M.** Corpus Linguistics and Translation Studies – Implications and Applications // *Text and Technology: In honour of John Sinclair* / ed. by G. Francis, E. Tognini-Bonelli. Amsterdam: John Benjamins Publ., 1993. P. 233–252. DOI: 10.1075/z.64.15bak
15. **Toury G.** *In Search of a Theory of Translation*. Tel Aviv: The Porter Institute for Poetics and Semiotics, 1980. 159 p.
16. **Blum-Kulka Sh.** Shifts of Cohesion and Coherence in Translation // *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, 1986. P. 17–35.
17. **Бабина О.И., Зиновьева А.Ю., Неручева Е.Д.** Влияние предварительной обработки набора данных на концептуальную разметку текстовых токенов на основе двунаправленной LSTM // *Terra Linguistica*. 2024. Т. 15, № 3. С. 109–123. DOI: 10.18721/JHSS.15310
18. **Laviosa S.** How Comparable Can Comparable Corpora Be? // *Target-international Journal of Translation Studies*. 1997. Vol. 9, No. 2. P. 289–319. DOI: 10.1075/target.9.2.05lav



19. **Ben-Ari N.** The Ambivalent Case of Repetitions in Literary Translation. Avoiding Repetitions: a “Universal” of Translation? // *Meta*. 1998. Vol. 43, No. 1. P. 68–78. DOI: 10.7202/002054ar
20. **Church K., Hanks P.** Word association norms, mutual information, and lexicography // *Proceedings of the 27th Annual Meeting on Association for Computational Linguistics*. Vancouver: Association for Computational Linguistics, 1989. P. 76–83. DOI: 10.3115/981623.981633
21. **Корут С.В.** Дискурсивные маркеры в письменном научном дискурсе // *Сибирский филологический журнал*. 2016. № 2. С. 157–163. DOI: 10.17223/18137083/55/17

REFERENCES

- [1] **Baroni M., Bernardini S.**, A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text, *Literary and Linguistic Computing*, 21 (3) (2005) 259–274. DOI: 10.1093/lc/fqi039
- [2] **Koppel M., Ordan N.**, Translationese and Its Dialects, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, 2011, pp. 1318–1326.
- [3] **Volansky V., Ordan N., Wintner Sh.**, On the features of translationese, *Digital Scholarship in the Humanities*, 30 (1) (2013) 98–118. DOI: 10.1093/lc/fqt031
- [4] **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., Springer, Singapore, 2021, pp. 133–180. DOI: 10.1007/978-981-16-4918-9_6
- [5] **Kunilovskaya M., Lapshinova-Koltunski E., Mitkov R.**, Translationese in Russian Literary Texts, *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Association for Computational Linguistics, Punta Cana, 2021, pp. 101–112. DOI: 10.18653/v1/2021.latechclfl-1.12
- [6] **Bizzoni Y., Lapshinova-Koltunski E.**, Measuring Translationese across Levels of Expertise: Are Professionals more Surprising than Students?, *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, Reykjavik, Linköping University Electronic Press, 2021, pp. 53–63.
- [7] **Pylypenko D., Amponsah-Kaakyire K., Dutta Chowdhury K., van Genabith J., España-Bonet C.**, Comparing Feature-Engineering and Feature-Learning Approaches for Multilingual Translationese Classification, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Punta Cana, 2021, pp. 8596–8611. DOI: 10.18653/v1/2021.emnlp-main.676
- [8] **Kunilovskaya M., Lapshinova-Koltunski E.**, Lexicogrammatic Translationese across Two Targets and Competence Levels, *Proceedings of the 12th Conference on Language Resources and Evaluation*, European Language Resources Association, Marseille, 2020, pp. 4102–4112.
- [9] **Ni J., Jin Z., Freitag M., Sachan M., Schölkopf B.**, Original or Translated? A Causal Analysis of the Impact of Translationese on Machine Translation Performance, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, Seattle, 2022, pp. 5303–5320. DOI: 10.18653/v1/2022.naacl-main.389
- [10] **Lee Y.C.**, Function Words as Markers of Translationese: A Corpus-based Approach to Mental Translation in Second Language Writing, *Korean Journal of English Language and Linguistics*, 21 (2021) 261–281. DOI: 10.15738/kjell.21..202103.261
- [11] **Gellerstam M.**, Translationese in Swedish novels translated from English, *Translation Studies in Scandinavia*, ed. by L. Wollin, H. Lindquist, CWK Gleerup Publ., Lund, 1986. pp. 88–95.
- [12] **Toury G.**, Interlanguage and its Manifestations in Translation, *Meta*, 24 (2) (1979) 223–231. DOI: 10.7202/004502ar
- [13] **Frawley W.**, Prolegomenon to a theory of translation, *Translation: Literary, Linguistic and Philosophical Perspectives*, ed. by W. Frawley, University of Delaware Press, Newark, 1984. pp. 159–175.
- [14] **Baker M.**, *Corpus Linguistics and Translation Studies – Implications and Applications*, Text and Technology: In honour of John Sinclair, ed. by G. Francis, E. Tognini-Bonelli, John Benjamins Publ., Amsterdam, 1993, pp. 233–252. DOI: 10.1075/z.64.15bak



[15] **Toury G.**, In Search of a Theory of Translation, The Porter Institute for Poetics and Semiotics, Tel Aviv, 1980.

[16] **Blum-Kulka Sh.**, Shifts of Cohesion and Coherence in Translation, Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies, Gunter Narr, Tübingen, 1986, pp. 17–35.

[17] **Babina O.I., Zinoveva A.Yu., Nerucheva E.D.**, Dataset preprocessing effects on Bi-LSTM based concept tagging of text tokens, Terra Linguistica, 15 (3) (2024) 109–123. DOI: 10.18721/JHSS.15310

[18] **Laviosa S.**, How Comparable Can Comparable Corpora Be?, Target-international Journal of Translation Studies, 9 (2) (1997) 289–319. DOI: 10.1075/target.9.2.05lav

[19] **Ben-Ari N.**, The Ambivalent Case of Repetitions in Literary Translation. Avoiding Repetitions: a “Universal” of Translation?, Meta, 43 (1) (1998) 68–78. DOI: 10.7202/002054ar

[20] **Church K., Hanks P.**, Word association norms, mutual information, and lexicography, Proceedings of the 27th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, Vancouver, 1989, pp. 76–83. DOI: 10.3115/981623.981633

[21] **Kogut S.V.**, Discourse markers in written scientific discourse, Siberian Journal of Philology, 2 (2016) 157–163. DOI: 10.17223/18137083/55/17

СВЕДЕНИЯ ОБ АВТОРЕ / INFORMATION ABOUT AUTHOR

Серова Олеся Усмановна

Olesya U. Serova

E-mail: serovaolesyau@gmail.com

ORCID: <https://orcid.org/0009-0009-4902-1738>

Поступила: 14.01.2025; Одобрена: 23.03.2025; Принята: 28.03.2025.

Submitted: 14.01.2025; Approved: 23.03.2025; Accepted: 28.03.2025.