

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ  
ИМ. А. И. ГЕРЦЕНА»

На правах рукописи

**Серова Олеся Усмановна**

**«Индикаторы феномена переводческого языка в переводах научных статей с  
английского на русский язык»**

Специальность 5.9.8. Теоретическая, прикладная и сравнительно-  
сопоставительная лингвистика

Диссертация на соискание ученой степени кандидата филологических наук

Научный руководитель:  
доктор филологических наук,  
заслуженный деятель науки РФ  
профессор Беляева Лариса Николаевна

Санкт-Петербург

2026

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ .....	4
ГЛАВА 1. ПРЕДПОСЫЛКИ И ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИССЛЕДОВАНИЯ ФЕНОМЕНА ПЕРЕВОДЧЕСКОГО ЯЗЫКА.....	15
1.1. Исторические предпосылки исследований феномена переводческого языка и их место в современном переводоведении .....	15
1.2. Подходы к пониманию феномена переводческого языка.....	21
1.3. Характеристики переводных текстов как универсалии перевода и как исследовательские гипотезы .....	28
1.4. Общая характеристика научного стиля в русском и английском языках .	36
Выводы по главе 1 .....	48
ГЛАВА 2. ОТБОР ИНДИКАТОРОВ ДЛЯ ИССЛЕДОВАНИЯ ХАРАКТЕРИСТИК ФЕНОМЕНА ПЕРЕВОДЧЕСКОГО ЯЗЫКА.....	51
2.1. Характеристика упрощение и её индикаторы .....	51
2.2. Характеристика нормализация и её индикаторы.....	59
2.3. Характеристика экспликация и её индикаторы .....	66
2.4. Характеристика интерференция и её индикаторы.....	76
2.5. Индикаторы переводческого языка, отражающие особенности научного стиля русского языка .....	81
Выводы по главе 2.....	86
ГЛАВА 3. ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК ПЕРЕВОДЧЕСКОГО ЯЗЫКА НА МАТЕРИАЛЕ НАУЧНЫХ СТАТЕЙ.....	89
3.1. Формирование исследовательского корпуса.....	89
3.2. Методы анализа результатов исследования .....	96
3.3. Интерпретация результатов анализа индикаторов феномена переводческого языка .....	102
Выводы по главе 3.....	150
ЗАКЛЮЧЕНИЕ.....	157
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ.....	161
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....	161
СПИСОК ИСТОЧНИКОВ .....	185
ПРИЛОЖЕНИЕ А. Классификация дискурсивных маркеров локального и глобального уровня связности текста.....	189
ПРИЛОЖЕНИЕ Б. Список тегов, использованных в анализе частеречных последовательностей (включая специальные теги) .....	190

ПРИЛОЖЕНИЕ В. Классификация дискурсивных маркеров авторского начала и взаимодействия автора с читателем.....	192
ПРИЛОЖЕНИЕ Г. Описание интерфейса и функциональности разработанного программного комплекса для анализа индикаторов феномена переводческого языка .....	193
ПРИЛОЖЕНИЕ Д. Результаты анализа индикаторов упрощения по корпусам..	200
ПРИЛОЖЕНИЕ Е. Результаты анализа индикатора лексической повторяемости по корпусам .....	205
ПРИЛОЖЕНИЕ Ё. Результаты анализа индикаторов экспликации по корпусам	206
ПРИЛОЖЕНИЕ Ж. Результаты анализа нормализованных частот частеречных n-грамм по корпусам .....	210
ПРИЛОЖЕНИЕ З. Результаты анализа индикаторов, отражающих особенности научного стиля русского языка, по корпусам.....	217

## ВВЕДЕНИЕ

Настоящее диссертационное исследование посвящено изучению феномена переводческого языка на материале сопоставимых корпусов, включающих тексты непереводных русскоязычных научных статей и текстов англо-русских ручных и машинных переводов. При этом под сопоставимым корпусом понимается корпус, состоящий из двух моноязычных подкорпусов – подкорпуса непереводных текстов, изначально написанных на языке перевода, и подкорпуса переводов на этот язык [Baker, 1995: 223-243; Laviosa, 1997: 289-320; Snell-Hornby, 2006: 125-126; Granger, 2010: 14-21], – соответствующих единым критериям отбора (стилистическая, жанровая, тематическая принадлежность, временной период), позволяющим с большей степенью надежности интерпретировать обнаруживаемые между подкорпусами лингвистические отличия как следствие их различающегося статуса (т. е. переводного / непереводного статуса), а не как результат влияния факторов, соответствующих установленным критериям отбора.

### Актуальность

Многочисленные исследования в области переводоведения демонстрируют статистически значимые отличия переводов от непереводных текстов, изначально написанных на языке перевода, по ряду лингвистических характеристик. Эти отличия дают основания говорить об особом, свойственном переводным текстам явлении – *феномене переводческого языка*, также известном как *translationese*. Особенности переводных текстов, обусловленные действием данного феномена, нередко связывают с качеством перевода [Nida, Taber, 1969: 13, 210; Duff, 1981: 1-3; Blum-Kulka, 2000: 292-293; Newmark, 1991: 23; Baker, 1993: 249], что подтверждается эмпирическими данными [Kunilovskaya, 2023], свидетельствующими о том, что эти особенности могут служить потенциальными предикторами качества перевода на уровне всего документа (текста) и что переводы более низкого качества демонстрируют более выраженные признаки переводческого языка.

Снижение качества переводного текста представляет собой серьезную проблему в любой сфере перевода. Однако в связи с тем, что научный перевод как

особая сфера переводческой деятельности характеризуется повышенными требованиями к точности передачи информации, учет качества перевода текстов данного типа наиболее важен. Проблема качества перевода проявляется наиболее остро в случае с таким особым жанром научного текста, как научная статья. Причиной этому служит то, что снижение качества перевода научных статей способно приводить к таким серьезным последствиям, как неверная интерпретация процедуры, методов, результатов, выводов и иных данных исследования, а также способно подорвать доверие к автору и снизить влияние его работы в академическом сообществе [Арцишевская, 2025: 106-107]. Более того, в условиях стремительного роста публикационной активности ученых в последние десятилетия<sup>1</sup>, глобализации и активного расширения международного научного сотрудничества изучение влияния переводческого языка на качество выполняемых переводов научных статей приобретает особое значение.

Анализ существующей литературы в аспекте изучения особенностей переводческого языка на материале переводов научной литературы с английского языка на русский и в частности на материале переводов научных статей демонстрирует лакуарность исследований в данной области. Установление индикаторов переводческого языка, потенциально влияющих на качество переводов научных статей, позволит не только лучше понять, как именно этот феномен проявляется в данной сфере перевода, но и выявить те аспекты перевода, которые необходимо учитывать как в самом процессе перевода, так и на этапе его постредактирования, в частности при постредактировании результатов машинного перевода, что и определяет **актуальность** настоящей работы.

Кроме того, несмотря на активное использование корпусных методов в исследованиях феномена переводческого языка последних десятилетий на материале разных языковых пар [Laviosa-Braithwaite, 1997: 531-540; Laviosa, 1998: 557-570; Bernardini, Zanettin, 2004: 51-62; Puurtinen, 2004: 165-176; Baroni, Bernardini 2006: 259-274; Koppel, Ordan, 2011: 1318-1326; Ilisei, 2012; Volansky et

---

<sup>1</sup>Publication Output by Region, Country, or Economy and by Scientific Field:  
<https://nces.nsf.gov/pubs/nsb202333/publication-output-by-region-country-or-economy-and-by-scientific-field/>

al., 2015: 98-118; Kunilovskaya, Lapshinova-Koltunski, 2019: 47-56; Kunilovskaya, Corpas Pastor, 2021: 133-180; Kunilovskaya et al., 2021: 101-112 и др.], а также несмотря на их доказанную эффективность в выявлении характеристик рассматриваемого феномена, в настоящее время отсутствуют специализированные инструменты, позволяющие одновременно формировать сопоставимые корпуса текстов и проводить автоматизированный анализ индикаторов характеристик переводческого языка на материале сформированных корпусов. Разработка такого инструмента представляет особую практическую ценность, поскольку систематическое сопоставление переводных текстов с эталонными непереводаемыми подкорпусами на основе индикаторов переводческого языка может способствовать повышению качества выполняемых / редактируемых переводов. Подобный инструмент позволит переводчикам и постредакторам количественно оценивать степень проявления переводческого языка в конкретных текстах, выявляя наиболее проблемные аспекты перевода, требующие правок. Более того, в условиях растущей потребности в эффективных методах контроля качества переводов в академической сфере создание такого специализированного инструмента становится важной задачей для обеспечения более высокого качества научных переводов, что также определяет **актуальность** настоящего исследования.

### **Степень научной разработанности темы**

**Теоретико-методологическую базу** исследования составили труды отечественных и зарубежных ученых, определяющие следующие аспекты темы:

- теория феномена переводческого языка и потенциальных универсальных характеристиках переводов (Н. К. Гарбовский, Н. Г. Кантышева, М. А. Куниловская, Е. С. Краснопеева, М. Baker, М. Baroni, S. Bernardini, S. Blum-Kulka, A. Chesterman, I. Ilisei, M. Koppel, E. Lapshinova-Koltunski, S. Laviosa, A. Mauranen, N. Ordan, V. Volansky, S. Wintner);
- корпусные исследования в переводоведении (В. П. Захаров, М. Baker, S. Laviosa, M. Snell-Hornby, A. Chesterman, S. Granger, T. McEnery, A. Hardie);
- теория научного перевода и сопоставление особенностей научного стиля в русском и английском языках (И. С. Алексеева, И. В. Арнольд,

Л. Н. Беляева, Р. К. Боженкова, А. Н. Васильева, И. Р. Гальперин, В. В. Гуревич, О. Н. Камшилова, М. Н. Кожина, В. Н. Комиссаров, Т. П. Плещенко, А. Л. Пумпянский, А. В. Федоров);

- методология корпусных исследований в аспекте изучения феномена переводческого языка (Е. С. Краснопеева, М. А. Куниловская, А. Б. Кутузов, M. Baker, M. Baroni, S. Bernardini, S. Blum-Kulka, G. Corpas Pastor, I. Ijisei, M. Koppel, E. Lapshinova-Koltunski, S. Laviosa, N. Ordan, V. Volansky, S. Wintner).

**Целью** работы является установление индикаторов феномена *переводческого языка*, влияющих на качество переводов научных статей с английского языка на русский.

**Объектом** исследования является перевод научных статей с английского языка на русский.

**Предметом** исследования являются индикаторы феномена переводческого языка, характерные для переводов научных статей с английского на русский язык.

**Гипотеза исследования** состоит в том, что переводы научных статей с английского языка на русский демонстрируют статистически значимые отличия от непереводаемых русскоязычных научных статей по ряду признаков, соответствующих четырем основным характеристикам переводческого языка (*упрощению, нормализации, экспликации и интерференции*), причём часть этих отличий, связанная с отклонениями от норм русского научного стиля, может служить потенциальными предикторами качества переводного текста.

Для достижения поставленной цели и проверки выдвинутой гипотезы необходимо решить следующие **задачи**:

- 1) Изучить исторические предпосылки, обусловившие интерес к исследованиям феномена переводческого языка, а также существующие подходы к его определению;

- 2) Рассмотреть концептуальные подходы к характеристикам переводческого языка как к переводческим универсалиям и как к потенциально универсальным характеристикам переводных текстов, а также существующие подходы к исследованию этих характеристик;

3) Описать особенности научного стиля в английском и русском языках для формирования теоретической базы для интерпретации результатов корпусного анализа;

4) Отобрать формальные индикаторы потенциально универсальных характеристик переводов, а также индикаторы, отражающие нормы научного стиля русского языка, с целью их последующего анализа;

5) Разработать программный комплекс для автоматизированного анализа формальных индикаторов на материале исследовательских текстовых корпусов;

6) Сформировать два сопоставимых исследовательских корпуса: корпус научных статей социологической тематики (непереводные тексты и ручные переводы) и корпус научных статей медицинской тематики (непереводные тексты, ручные и машинные переводы), проанализировать сформированные корпуса при помощи разработанного программного комплекса и провести статистический анализ полученных результатов с применением методов оценки значимости различий между подкорпусами;

7) Основываясь на результатах анализа, определить набор индикаторов, которые могут быть использованы при оценке качества переводов научных статей с английского на русский язык, и сформулировать практические рекомендации для переводчиков и постредакторов.

**Материалом диссертационного исследования** послужили тексты научных статей, отобранные из рецензируемых научных периодических изданий, сгруппированные по тематике и способу перевода:

1) С целью сопоставительного анализа ручных переводов и непереводных текстов социологической тематики были извлечены статьи из журналов «Социологическое обозрение», «Экономическая социология», «Журнал социологии и социальной антропологии» (59 непереводных и 59 переводных научных статей);

2) С целью сопоставительного анализа ручных переводов и непереводных текстов медицинской тематики извлекались статьи из журналов «Нефрология и диализ», «Рациональная Фармакотерапия в Кардиология», «Доказательная

Гастроэнтерология», «The EYE ГЛАЗ», «Российский кардиологический журнал», «Патология кровообращения и кардиохирургия», «Остеопороз и остеопатии», «Кардиология», «Качественная клиническая практика», «Нервно-мышечные болезни», «Русский журнал детской неврологии», «Пульмонология», «Проблемы эндокринологии», «Клиническая и экспериментальная тиреоидология», «Проблемы репродукции» (70 непереводных и 70 переводных научных статей);

3) Для сопоставительного анализа машинных переводов и непереводных текстов медицинской тематики в качестве непереводных текстов использовались статьи, извлеченные из указанных выше источников (70 непереводных статей), в то время как для формирования подкорпуса машинных переводов были отобраны оригинальные англоязычные научные статьи из журналов «The Lancet», «The New England Journal of Medicine», «Biology of Reproduction», «European Journal of Pharmacology», которые затем были переведены с использованием системы машинного перевода *Google Translate* (50 машинных переводов научных статей).

Основными **методами исследования** послужили:

1) Гипотетико-дедуктивный и индуктивный методы, использованные при формулировании исходной гипотезы о наличии специфических характеристик переводческого языка в переводах научных статей на основе теоретических положений о данном феномене (дедукция), а также при последующем обобщении полученных эмпирических данных (индукция);

2) Метод лингвистического эксперимента применялся для систематического сопоставления переводных и непереводных подкорпусов текстов в контролируемых условиях (единые критерии отбора текстов по жанру, тематике, периоду создания) с целью проверки выдвинутой гипотезы;

3) Метод корпусного анализа, реализованный при помощи специально разработанного в рамках исследования программного комплекса для анализа индикаторов феномена переводческого языка на материале корпусов непереводных русскоязычных текстов и англо-русских переводов;

4) Методы статистического анализа для выявления статистически значимых различий между подкорпусами: двухвыборочный t-критерий Уэлча для

показателей, продемонстрировавших распределение данных, близкое к нормальному, а также непараметрический U-критерий Манна–Уитни для анализа показателей, чье распределение оказалось отличным от нормального. Для оценки практической значимости различий рассчитывалась величина эффекта  $r$ .

**На защиту выносятся следующие положения:**

1) переводы научных статей с английского языка на русский демонстрируют статистически значимые отличия от неперевода русских научных текстов по ряду индикаторов, соответствующих четырём основным характеристикам переводческого языка (*упрощению, нормализации, экспликации и интерференции*);

2) каждая из указанных характеристик (*упрощение, нормализация, экспликация и интерференция*) проявляется в переводах научных статей через совокупность поддающихся количественному анализу формальных индикаторов;

3) разработанный в рамках диссертационного исследования программный комплекс позволяет автоматизированно выявлять формальные индикаторы переводческого языка при сопоставлении подкорпуса переводных текстов с эталонным подкорпусом неперевода текстов на русском языке;

4) степень выраженности индикаторов переводческого языка варьируется как в зависимости от предметной области переводных научных статей (социологическая / медицинская тематика) и от способа выполнения перевода (ручной / машинный), причём машинные переводы демонстрируют более выраженные отклонения от норм русского научного стиля, чем ручные переводы;

5) установленные индикаторы переводческого языка могут быть разделены на две группы: (а) индикаторы, фиксирующие стилистические отклонения от норм русского научного стиля, потенциально снижающие качество перевода и требующие учёта при переводе и постредактировании, и (б) индикаторы, фиксирующие отличия переводных текстов от неперевода, но либо не связанные напрямую со снижением качества перевода, либо требующие дальнейших исследований для установления характера их влияния на качества перевода.

**Научная новизна исследования** заключается в том, что впервые:

1) проведён комплексный анализ феномена переводческого языка на материале переводов научных статей с английского языка на русский с использованием 46 формальных индикаторов данного феномена, охватывающих четыре потенциально универсальные характеристики переводов (*упрощение, нормализацию, экспликацию и интерференцию*), а также индикаторы соответствия нормам научного стиля русского языка;

2) разработан специализированный программный комплекс для автоматизированного сбора корпуса текстов и дальнейшего анализа сформированного корпуса с целью установления формальных индикаторов феномена переводческого языка в переводных текстах;

3) на основе сопоставительного анализа непереводных текстов с текстами ручных и машинных переводов двух предметных областей (социологической и медицинской) произведена дифференциация выявленных проявлений переводческого языка на (а) потенциально влияющие на качество перевода и (б) фиксирующие отличия переводных текстов от непереводных, но не связанные напрямую со снижением качества перевода.

**Теоретическая значимость** исследования заключается в следующем:

1) в рамках исследования получили развитие теоретические представления о проявлении потенциально универсальных характеристик перевода на конкретной языковой паре (английский → русский) и жанре (научная статья);

2) разработанная методология эмпирического исследования вносит вклад в развитие представлений о проектировании и использовании специализированных программных инструментов для корпусных исследований феномена переводческого языка.

**Практическая значимость** работы определяется:

1) возможностью применения полученных результатов в таких практических сферах, как перевод научных статей, постредактирование ручных и машинных переводов, разработка критериев оценки качества ручных и машинных переводов и подготовка переводчиков научной литературы;

2) возможностью использования разработанного программного комплекса переводчиками и постредакторами с целью выявления индикаторов переводческого языка на материале выполняемых / редактируемых переводов, а также с целью сопоставления результатов анализа этих переводов с результатами анализа непереводаемых текстов (при наличии эталонного непереводаемого подкорпуса), что может способствовать более объективной оценке качества ручных и машинных переводов и выявлению областей, требующих улучшения.

**Апробация результатов исследования.** Основные положения и результаты исследования были изложены в докладах на следующих научных и практических конференциях:

1. Международная научная конференция «Корпусная лингвистика – 2023» (Санкт-Петербург, 21–23 июня 2023 г., СПбГУ)
2. III Международная научная филологическая конференция имени Л. А. Вербицкой (Санкт-Петербург, 19–26 марта 2024 г., СПбГУ);
3. IV Международная конференция «Пиотровские Чтения – 2024» (Санкт-Петербург, 22 ноября 2024 г., РГПУ им. А. И. Герцена).
4. Всероссийская научно-практическая конференция с международным участием «Язык, текст, культура через призму цифровых технологий» (Санкт-Петербург, 16-17 декабря 2024 г., ВШЭ).
5. II Всероссийская научно-практическая конференция с международным участием «Язык, текст, культура через призму цифровых технологий» (Санкт-Петербург, 4-5 декабря 2025 г., ВШЭ).

По теме диссертации опубликовано 4 работы, из них 3 – в журналах, рекомендованных ВАК РФ, и 1 из них в журнале, индексируемом в базе данных Scopus:

1. 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.

2. Серова О. У. Феномен *translationese* в машинном переводе научных статей с английского языка на русский (количественные характеристики) // Этнопсихоллингвистика. 2025. № 4 (23). С. 72-89. DOI: 10.31249/epl/2025.04.05.

3. Серова О. У. Количественный анализ индикаторов феномена *translationese* в переводах медицинских научных статей // Известия Российского государственного педагогического университета имени А. И. Герцена. 2025. № 217. С. 312-326. DOI: 10.33910/1992-6464-2025-217-312-326.

4. Серова О. У. Исследование феномена *translationese* (на материале машинного перевода научных текстов с английского на русский язык) // Труды международной конференции «Корпусная лингвистика – 2023». – СПб.: Изд-во С.-Петербург. ун-та, 2024. С. 286-293.

**Структура и объем диссертации** определены поставленной целью и задачами.

Диссертация состоит из введения, трех глав, заключения, списка сокращений и условных обозначений, списка использованной литературы, включающего 96 наименований на русском языке, 127 наименований на английском языке, списка источников, из которых были взяты статьи для формирования исследовательских корпусов, и 9 приложений.

Во **введении** обосновывается выбор темы исследования и ее актуальность, определяются цель, задачи, объект и предмет исследования, характеризуется материал исследования и применяемые методы, формулируются теоретическая и практическая значимость с указанием возможного применения полученных результатов, а также отражена апробация работы.

В **первой главе «Предпосылки и теоретические основы исследования феномена переводческого языка»** рассматриваются исторические предпосылки изучения феномена переводческого языка, существующие подходы к его пониманию и оценке, описываются основные представления о потенциальных универсальных характеристиках переводных текстов и подходах к их исследованию, а также дается характеристика основных стилеобразующих особенностей научных текстов в английском и русском языках.

**Вторая глава «Отбор индикаторов для исследования характеристик феномена переводческого языка»** посвящена подробному рассмотрению особенностей и формальных индикаторов потенциальных универсальных характеристик перевода, которые легли в основу корпусного анализа в практической части настоящей работы.

**Третья глава «Исследование характеристик переводческого языка на материале научных статей»** посвящена описанию методологии эмпирического исследования, процедуры формирования исследовательских корпусов, разработке программного комплекса для их автоматизированного анализа, а также интерпретации полученных результатов.

Каждая глава диссертации завершается обоснованными выводами.

**В заключении** подводятся итоги работы, формулируются основные выводы исследования, определяется степень достижения поставленной цели и решения задач, а также намечаются возможные перспективы дальнейших исследований.

**Приложения** содержат классификацию дискурсивных маркеров локального и глобального уровня связности текста с примерами (ПРИЛОЖЕНИЕ А), список тегов частеречной разметки, использованных в анализе частеречных последовательностей (ПРИЛОЖЕНИЕ Б), классификацию дискурсивных маркеров авторского начала и взаимодействия автора с читателем с примерами (ПРИЛОЖЕНИЕ В), а также описание интерфейса и функциональности разработанного программного комплекса для анализа индикаторов феномена переводческого языка (ПРИЛОЖЕНИЕ Г). Помимо этого, в приложениях представлены сводные результаты статистического анализа по всем трём подкорпусам: результаты анализа индикаторов упрощения (ПРИЛОЖЕНИЕ Д), индикатора лексической повторяемости в рамках характеристики нормализация (ПРИЛОЖЕНИЕ Е), индикаторов экспликации (ПРИЛОЖЕНИЕ Ё), результаты анализа нормализованных частот частеречных n-грамм в рамках характеристики интерференция (ПРИЛОЖЕНИЕ Ж), а также результаты анализа индикаторов, отражающих особенности научного стиля русского языка (ПРИЛОЖЕНИЕ З).

## ГЛАВА 1. ПРЕДПОСЫЛКИ И ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИССЛЕДОВАНИЯ ФЕНОМЕНА ПЕРЕВОДЧЕСКОГО ЯЗЫКА

### 1.1. Исторические предпосылки исследований феномена переводческого языка и их место в современном переводоведении

Феномен переводческого языка представляет собой одну из центральных проблем современного переводоведения, все чаще привлекающую внимание исследователей на протяжении последних десятилетий в области изучения как ручных [Gellerstam, 1986; Gellerstam, 1996; Laviosa-Braithwaite, 1997; Laviosa, 1998; Baroni, Bernardini 2003; Puurtinen, 2004: 165-176; Mauranten, 2004: 65-82; Pápai, 2004: 143-164; Baroni, Bernardini, 2006; Kurokawa et al., 2009; Koppel, Ordan, 2011; Rodríguez-Castro, 2011; Popescu, 2011; Ilisei, 2012; Lapshinova-Koltunski, 2015; Volansky et al., 2015; Kunilovskaya, Kutuzov, 2015; Краснопеева, 2016; Kunilovskaya, Lapshinova-Koltunski, 2019; Bizzoni, 2020; Kunilovskaya et al., 2021; Pylypenko et al., 2021; Kunilovskaya, Corpas Pastor, 2021; François, Lefer, 2022; Kunilovskaya, 2023 и др.], так и машинных переводов [Lapshinova-Koltunski, 2015; Zhang, Toral, 2019; Bizzoni, 2020; Graham et al., 2019; Riley et al., 2020; Graham et al., 2020; Jiang, Niu, 2022; Zhou, 2024; Kotait, 2024 и др.]. Накопленные в результате исследований данные демонстрируют, что переводные тексты образуют особый подкласс языковых произведений ПЯ, статистически значительно отличающихся от непереводаемых текстов на этом языке по целому ряду характеристик. Совокупность таких характеристик как раз и получило в научной литературе название *переводческий язык*.

Возникновение исследовательского интереса к изучению характеристик переводных текстов принято связывать с зарождением в последней четверти XX века дескриптивного направления переводческих исследований, или дескриптивного переводоведения (*Descriptive Translation Studies*) [Прунч, 2015: 252-253]. В основе данного направления, как это следует из названия, лежит дескриптивный подход, принципиально противопоставляемый прескриптивному: если прескриптивное переводоведение ориентировано на установление норм и правил – на предписание того, каким должен быть перевод для достижения

переводческой эквивалентности, – то дескриптивное переводоведение изучает перевод «как он есть», систематически описывая явления, реально наблюдаемые в переводных текстах, независимо от того, отвечают ли они установленным нормам.

Идеи о важности дескриптивного подхода в переводоведении неоднократно высказывались отечественными учёными задолго до того, как дескриптивное переводоведение оформилось в самостоятельную отрасль научных исследований в рамках западноевропейской дисциплины *Translation Studies* [Прунч, 2015: 252-253]. Так, ещё А. В. Федоров подчеркивал [Федоров, 2002: 20] особую важность для теории перевода объективного (по сути дескриптивного) подхода, суть которого состоит в изучении явлений, действительно существующих в непосредственно наблюдаемых переводных текстах. Использование дескриптивного подхода, как он считал, позволяет, в противоположность прескриптивным переводческим концепциям, фиксировать, обобщать и систематизировать реально существующие в переводе факты, делая на их основе более аргументированные и убедительные умозаключения о феномене перевода. Аналогичную позицию занимали и другие отечественные исследователи. И. И. Ревзин и В. Ю. Розенцвейг [Ревзин, Розенцвейг, 1963: 19-25] критиковали прескриптивный подход за то, что тот не способен обеспечить должный уровень теоретического осмысления переводческой деятельности, так как формулируемые в его рамках нормативные положения зачастую основываются на уже существующих положениях других лингвистических дисциплин (лексикологии, грамматики, стилистики и литературоведения), в то время как изучение того, что действительно существует в переводе, остается без должного внимания. Подходя к исследованию перевода с позиций структурной лингвистики, они указывали на важность использования дескриптивного подхода для описания реально наблюдаемых в переводах явлений с целью последующего установления и формулирования действующих в переводе законов и правил. Л. С. Бархударов [Бархударов, 1975: 39-40], в свою очередь, утверждал, что теория перевода по самой своей природе является преимущественно дескриптивной дисциплиной, главная задача которой состоит в анализе уже существующих переводов и

систематическом описании наблюдаемых в них закономерностей, в описании того, как на самом деле протекает процесс перевода, – и что именно этот анализ должен предшествовать любым нормативным предписаниям, выступая необходимым условием их научной обоснованности.

Вместе с тем методологические предпосылки для систематического изучения переводных текстов, сложившиеся в рамках дескриптивного направления, смогли в полной мере реализоваться лишь в условиях принципиального переосмысления самого статуса переводов, произошедшего в последней четверти XX – начале XXI века: если ранее переводы воспринимались как нечто вторичное и производное от текстов оригинала, как вынужденная необходимость, во всём уступающая исходным текстам, то теперь они стали рассматриваться как самостоятельное явление, обладающее своими собственными уникальными характеристиками, отличающими переводы от непереводаемых текстов, и достойное отдельного изучения само по себе [Even-Zohar, 1987: 107-115; Baker, 1993: 233-234; Olohan, 2002: 419-429; Toury, 2012: 18-20; Краснопеева, 2015: 66; Куниловская, 2015: 13-19; Багдасарян, 2018: 295-298; Кантышева, 2021: 137]. Такое переосмысление стало возможным прежде всего благодаря возросшему пониманию важности выполняемой переводными текстами роли в контексте культуры-реципиента. Само появление новых переводов не есть явление стихийное и бессистемное. Напротив, оно подчиняется определенным закономерностям: из всего многообразия существующей в мире литературы выбор всегда падает на перевод текстов определенных авторов, тем и даже целых предметных областей. Иными словами, выбор того, что переводить, вовсе не случаен, но обязательно отражает потребности и запросы принимающей культуры, её стремление к обновлению и восполнению существующих пробелов в той или иной области знания [Even-Zohar 1987: 109; Toury, 1985: 26; Латышев, Семенов, 2003: 5-6, 69-70; Toury, 2012: 6], а сам перевод в этом контексте можно рассматривать как особую, телеологическую деятельность, направленную на достижение конкретной цели в принимающей культуре.

Так, возросло понимание того, что переводная литература играет далеко не последнюю роль в формировании мировоззрения и опыта носителей принимающей культуры, так как с её помощью политические, художественные, научные и иные идеи получают широкое распространение, становясь доступными в мировом масштабе, на что в разное время обращали внимание как отечественные, так и зарубежные ученые [Ревзин, Розенцвейг, 1963: 5, 12-13; Even-Zohar, 1987: 107-115; Швейцер, 1988: 20-21, 45; Baker, 1993: 233-234; Латышев, Семенов, 2003: 69-70; Алексеева, 2004: 179; Леонтович, 2008: 20; Toury, 2012: 18-22, Прунч, 2015: 412-415; Гарбовский, Костикова, 2018: 17-18; Сдобников, 2019: 309-310; Сдобников и др., 2019: 26-27; Кушнина и др., 2019: 48-56; Волкова, 2021: 62–69; Михайлова, 2023: 370; Козлова, 2024: 271; Костикова, 2024: 179-187 и др.]. Перевод играет роль важного и незаменимого посредника, обеспечивающего трансфер идей между различными культурами. И, что ещё более важно, очень часто он является главным и даже единственным источником информации, с которым сталкивается читатель, не имея возможности читать текст оригинала.

Важнейшим теоретическим основанием для переосмысления статуса переводов стала теория литературной полисистемы И. Эвен-Зохара [Even-Zohar, 1979: 287-310; Even-Zohar 1987: 107-115], во многом опирающаяся на наследие русского формализма – прежде всего на работы Ю. Н. Тынянова и Р. О. Якобсона. Ю. Н. Тынянов, изучая литературную эволюцию, отмечал, что как отдельное произведение (текст) по своей сути является системой взаимосвязанных элементов, так и литература в целом представляет собой не простой набор отдельных произведений (текстов), а сложную систему, состоящую из самостоятельных литературных подсистем, постоянно взаимодействующих друг с другом и влияющих друг на друга. При этом изучение текста в изоляции от той литературной подсистемы, элементом которой он является, не позволяет адекватно судить как о самом произведении, так и о составляющих элементах этого произведения и выполняемой этими элементами функции, поскольку игнорирует соотнесённость данного текста с другими произведениями той же жанровой традиции, того же временного периода, автора и т. д. [Тынянов, 1977: 272-274, 281].

Применительно к переводам полисистемная теория аналогичным образом предлагает рассматривать их не как совокупность разрозненных текстов, а как отдельную литературную подсистему в составе обширной литературной полисистемы принимающей культуры [Even-Zohar, 1979: 287-310; Even-Zohar 1987: 107-109; Прунч, 2015: 260-263; Шелестюк, Гриценко, 2016: 205; Багдасарян, 2018: 295-298]: как только перевод становится частью этой культуры, он перестаёт восприниматься как отдельный, изолированный текст и интегрируется в литературное пространство этой культуры, приобретая статус элемента её литературной полисистемы и начиная функционировать в ней наравне с непереводаемыми произведениями [Toury, 1985: 18-19]. Иными словами, переводы не существуют обособленно, а становятся органичной частью литературной системы языка перевода, взаимодействуя с другими её элементами. При этом сама переводная литература в разное время способна занимать как центральное, так и периферийное положение в этой полисистеме. Занимая центральное положение, она активно влияет на литературные нормы принимающей культуры, видоизменяя и обновляя их. Напротив, находясь на периферии, она выполняет консервативную функцию, сохраняя и закрепляя уже существующие [Even-Zohar, 1987: 107-115]. По сути решение о переводе какого-либо текста запускает процесс его реконтекстуализации: возникая в виде переводного произведения, он перестаёт принадлежать системе исходной культуры и приобретает статус элемента литературной полисистемы принимающей культуры, входя в какую-то конкретную подсистему текстов этой полисистемы и сталкиваясь с её ограничениями и требованиями. А дальнейшая оценка текста перевода осуществляется по правилам этой конкретной подсистемы – в координатах её функциональных, стилистических и жанровых конвенций [Even-Zohar 1987: 107-115; Toury, 1985: 23].

В русле идей, созвучных полисистемной теории, на особый статус переводной литературы в рамках принимающей культуры обращает внимание Н. К. Гарбовский [Гарбовский, 2011: 3-19; Гарбовский, 2012: 130-136]. Он вводит понятие '*переводного дискурса*', под которым понимается некоторая совокупность речевых произведений, возникающих в результате переводческой деятельности и

формирующих особый массив текстов в дискурсивном пространстве русскоязычной речевой культуры. Занимая особое место в дискурсивном пространстве принимающей культуры, переводной дискурс со временем способен оказывать на это пространство существенное воздействие, изменяя дискурсивные нормы русского языка. Вся совокупность составляющих переводной дискурс речевых произведений определяется через три ключевые характеристики: онтологическую вторичность, телеологическую двойственность и необходимость преодоления межъязыковой и межкультурной асимметрии в процессе перевода. Вторичность переводов является их основополагающей характеристикой [Гарбовский, 2012: 130-136]: переводные произведения всегда ограничены рамками исходного «предтекста», а условия самого их порождения принципиально отличаются от условий порождения оригинальных произведений, так как последние создаются на других языках, в иных культурных и социально-исторических условиях. Такая вторичность обуславливает как двойственность в оценке переводных произведений, так и телеологическую двойственность применяемых переводчиками стратегий. Так, переводные тексты оцениваются, с одной стороны, по степени их близости к тексту оригинала и, с другой стороны, по степени соответствия дискурсивным традициям и нормам принимающей культуры – всё это определяет известное противоречие в стремлениях переводчика: сохранить верность перевода тексту оригинала и одновременно с этим сделать его естественным для восприятия носителями переводящего языка. При этом, как утверждает Н. К. Гарбовский, одновременное достижение обеих этих целей представляется крайне затруднительным по причине языковой и культурной асимметрии, в результате которой каждый акт перевода представляет собой процесс преодоления создаваемых этой асимметрией противоречий [Гарбовский, 2012: 130-136]. По завершении акта перевода принимающая культура получает особое речевое произведение, в котором неизбежно находит свое отражение противоречие между идеальной целью перевода – одновременно соблюсти верность оригиналу и адаптировать текст к нормам принимающей культуры – и реальными возможностями её достижения, проистекающее из несоответствия

языковой и культурной действительностей, к которым принадлежат исходное и переводное произведения [Гарбовский, 2012: 130-136].

Специфические условия, в которых возникают переводы, позволяют выдвинуть гипотезу о том, что на практике их онтологическая специфика может обнаруживать себя в совокупности особых характеристик переводных произведений, проявляющихся на семантическом, синтаксическом и стилистическом уровнях организации текста [Гарбовский, 2012: 130-136]. Так, Н. К. Гарбовский подчеркивает важность проведения исследований, направленных на выяснение того, существуют ли в реальности специфические характеристики переводных текстов, которые, с одной стороны, объединяют их в единое дискурсивное образование, а с другой – отличают от оригинальных произведений на языке перевода. Как утвердительный, так и отрицательный ответы на данный вопрос, по мнению Н. К. Гарбовского, во многом остаются на уровне гипотез из-за недостаточного объема эмпирических исследований, а решение данной проблемы способно существенно обогатить переводоведение в аспектах его теоретико-методологических основ, критического анализа переводов, а также методики обучения переводу [Гарбовский, 2012: 130-136].

Таким образом, описанные выше теоретические сдвиги – методологический поворот к дескриптивному исследованию переводов, переосмысление их статуса как самостоятельного объекта научного изучения, а также признание переводного дискурса особым элементом дискурсивного пространства принимающей культуры – в совокупности создали необходимые предпосылки для постановки вопроса о природе феномена переводческого языка. Далее сосредоточимся на существующих подходах к определению данного феномена.

## **1.2. Подходы к пониманию феномена переводческого языка**

Феномен переводческого языка привлекал внимание многих переводчиков и теоретиков перевода задолго до того, как стал предметом пристального научного изучения. В самом общем виде *переводческий язык* можно определить как совокупность лингвистических – лексических и грамматических – особенностей,

характерных для переводных текстов и отличающих их от непереводаемых текстов на языке перевода. Вместе с тем существуют принципиально разные оценки данного феномена относительно природы и причин его возникновения, его влияния на качество переводных текстов, а также относительно его универсальности (т.е. его наличия во всех переводных текстах без исключения). Рассмотрим основные подходы к его осмыслению, сложившиеся в отечественной и зарубежной научной традиции.

Ряд отечественных [Чуковский, 1941: 103-108; Кашкин, 1977: 499; Савко, 2010: 6; Умерова, 2010: 139; Лекомцева, 2016: 99-104; Бойко, 2023: 9-10; Латышев, Северова, 2025: 81-83] и зарубежных исследователей [Nida, Taber, 1969: 13, 210; Duff, 1981: 1-3; Newmark, 1991: 21-23; Gellerstam, 2005; Rayson et al., 2008] квалифицируют феномен переводческого языка как нежелательное явление – результат отрицательного переноса в текст перевода особенностей исходного текста, вследствие которого у носителей ПЯ создается впечатление некоторой неестественности и чужеродности текста перевода в сравнении с непереводаемыми произведениями на переводящем языке. Согласно данному подходу, на уровне отдельных форм и структур – слов, словосочетаний, синтаксических конструкций, пунктуационного оформления, фразеологии – переводной текст может производить впечатление соответствия нормам ПЯ, вместе с тем при более пристальном рассмотрении обнаруживается, что характер их употребления несёт на себе следы структурных и стилистических особенностей языка и текста оригинала. Эти следы, не всегда заметные при первом приближении, создают ощущение стилистической тяжеловесности и неуклюжести переводного текста. Общим для сторонников данного подхода является признание того, что феномен переводческого языка способен приводить к ряду негативных последствий: «размыванию» или, в более серьёзных случаях, искажению смысла исходного текста, к нарушениям норм и узуса ПЯ. Причинами этому принято считать недостаточное знание переводчиком специфики переводящего языка и принимающей культуры и / или чрезмерную ориентацию на формальные структуры исходного текста в ущерб естественности текста перевода [Чуковский,

1941: 103-108; Nida, Taber, 1969: 13, 210; Кашкин, 1977: 499; Duff, 1981: 1-3; Newmark, 1991: 21-23; Rayson et al., 2008; Савко, 2010: 6; Умерова, 2010: 139; Лекомцева, 2016: 99-104; Лекомцева и др., 2022: 158-167; Бойко, 2023: 9-10; Латышев, Северова, 2025: 81-83 и др.]. При этом сам факт наличия феномена переводческого языка в тексте перевода, согласно этому подходу, не всегда зависит от таких, казалось бы, определяющих факторов, как профессиональный уровень подготовки переводчика и то, является ли он носителем ПЯ: феномен проявляется в переводе в случае, если ПЯ не является родным для переводчика и он им, соответственно, не владеет в совершенстве, а также и в случае, когда ПЯ является родным, но переводчик недооценивает всю глубину межъязыковых различий и степень потенциально вызываемой этими различиями отрицательной интерференции [Duff, 1981: 1-3; Newmark, 1991: 23]. Для обозначения переводческого языка как негативного явления в англоязычной научной традиции закрепились термины *translationese* [Nida, Taber, 1969; Newmark, 1991] и *the third language* [Duff, 1981: 1-3].

Вместе с тем не все исследователи разделяют столь однозначную критическую оценку переводческого языка. Так, М. Геллерштам [Gellerstam, 2005: 201-213], не оспаривая сам факт лингвистического своеобразия переводов, предлагает рассматривать их отличительные особенности не как переводческую оплошность, а как особые «отпечатки» (*fingerprints*) исходного языка и текста – закономерное и в определённом смысле естественное явление, систематически проявляющееся в переводах вследствие интерференции языка оригинала и расхождений в культурных и жанрово-стилистических традициях задействованной языковой пары. По мнению исследователя [Gellerstam, 2005: 201-213], изучение подобных «отпечатков» требует объективного научного анализа, свободного от их априорно негативной оценки.

При другом подходе, также рассматривающем феномен переводческого языка с нейтральных позиций, данный феномен осмысливается через концепцию уникального «третьего кода» (*third code*), получающегося в результате сложного взаимодействия двух других кодов – системы исходного и системы переводящего

языка, но не сводящегося ни к одной из них [Frawley, 2000: 250-263; Granger, 2018: 185; Shuttleworth, Cowie, 2014: 173; Шакирова, 2016: 82-87; Бойко, Харитонов, 2024: 36-37]. При этом для обозначения исходного языка вводится понятие ‘матричного кода’ (*matrix code*) – системы, предоставляющей информацию для перекодирования на ПЯ, а для обозначения языка перевода – понятие ‘целевого кода’ (*target code*), т. е. системы, задающей спектр возможных средств и ограничений, доступных для перекодирования заложенной в тексте оригинала информации. Переводческий язык в этом контексте предстаёт как результат динамичного двустороннего взаимодействия между этими кодами, их взаимного приспособления друг к другу: переводчик вынужден постоянно балансировать между ними, соотнося возможности и ограничения обеих систем и принимая решения о том, что можно сохранить, а что необходимо изменить или опустить в процессе перевода [Frawley, 2000: 250-263]. Целевой код не является набором готовых элементов, штампов, которые автоматически соответствуют элементам исходного текста. Вместо этого он устанавливает рамки и задает параметры (правила грамматики, стилистические требования, культурные нормы), в пределах которых информация из матричного кода может быть перекодирована. Переводчик не просто находит в целевом коде элементы, идентичные элементам оригинала, но адаптирует передаваемую информацию к этим рамкам и параметрам. Иными словами, информация матричного кода приспособляется к параметрам и ограничениям целевого кода в той же степени, в какой параметры целевого кода приспособляются к информации матричного кода. А получающийся в итоге «третий код» никогда не подчиняется полностью ни одному из задействованных в переводе кодов и, хотя и зависит от них, является самостоятельной системой, со своими закономерностями и логикой.

Наконец, Г. Тури [Toury, 1979: 223-231; Гарбовский, 2012: 133-134] интерпретирует переводческий язык через призму концепции промежуточного языка, или языка-посредника (*interlanguage*) – феномена, характеризующегося промежуточным положением между ИЯ и ПЯ в том смысле, что переводной текст одновременно испытывает на себе перенос черт, свойственных обеим языковым

системам. Примечательно, что сам термин *interlanguage* был позаимствован для обозначения переводческого языка из исследований, посвященных изучению процесса овладения иностранным языком (*Second Language Acquisition*), где он применяется для обозначения особой промежуточной языковой системы, которую обучающиеся используют в своей речи в процессе языковой практики [Selinker, 1972: 209-230; Вагнер, 1997; Пилипчук, 2003: 122-126; Залевская, 2013; Громова, 2015: 221-224; Власенко, Толмачёва, 2016: 96-101; Шаклеин, Цяньминь, 2017: 170-175 и др.]. Эта система испытывает на себе одновременное влияние интерференции обеих языковых систем – системы родного (*L1*) и системы иностранного (*L2*) языка. А получающийся в процессе речевой практики промежуточный язык рассматривается как особая «третья система», для которой свойственны свои собственные правила функционирования грамматики и лексики, отличные от правил, принятых в родном и изучаемом языках. Применяя термин *interlanguage* для обозначения языка переводов, Г. Тури подчеркивает, что данное явление способно проявляться не только в условиях изучения иностранного языка, но и в любой другой ситуации, где два языка вступают в непосредственный контакт друг с другом [Toury, 1979: 223-224]. Перевод, в свою очередь, представляет собой одну из наиболее распространенных и типичных ситуаций языкового контакта, поскольку неизбежно ставит переводчика в положение активного посредника между двумя языковыми системами и культурами [Toury, 1979: 223-224]. При этом направление перевода и степень владения языками не имеют принципиального значения, поскольку в той или иной степени *interlanguage* возникает всегда в силу самой природы перевода как процесса, в котором два языка и их культурные контексты сталкиваются, неизбежно порождая уникальные языковые формы [Toury, 1979: 223-224].

Так, возникновение промежуточного языка принято объяснять неустранимым противоречием между т. н. постулатом адекватности и постулатом приемлемости [Toury, 1979: 227; Умерова, 2006: 468-474]. Оба постулата связаны с известной двойственностью при оценивании текста перевода [Комиссаров, 1999: 126-127; Chesterman, 2010: 39; Костикова, 2010: 43; Гарбовский, 2012: 131-132;

Багдасарян, 2018: 296]: с одной стороны, перевод всегда оценивается через призму его верности тексту оригинала, где основным критерием становится то, насколько точно, полно и достоверно переводной текст отражает содержание, форму и иные особенности оригинального текста, с другой – через призму его соответствия дискурсивным, стилистическим нормам и практикам принимающей культуры, то есть по сути сравнивается с соответствующими непереводаемыми речевыми произведениями данного типа на языке перевода. В реальности из-за языковой и культурной асимметрии переводчику редко удается соблюсти оба постулата одновременно [Гарбовский, 2012: 131-132]. При стремлении сделать текст перевода как можно более верным тексту оригинала получающийся в результате перевод, максимально отражая особенности исходного текста, нередко воспринимается носителями языка перевода как не совсем естественный, как нарушающий языковую норму и как очевидно переводной. В то же время в случае стремления как можно лучше адаптировать перевод к дискурсивным нормам и практикам переводящего языка переводной текст, становясь более гладким, естественным и идиоматичным [Бойко, 2023: 9-10], утрачивает некоторые черты текста оригинала, все более отдаляясь от него. Иными словами, перевод всегда находится на стыке между двумя этими крайностями и никогда в полной мере не совпадает ни с одной из них, поскольку полностью адекватным (в указанном выше смысле) может быть только сам исходный текст, а полностью приемлемым – только текст, изначально созданный на языке перевода (т. е. непереводаемой) [Toury, 1979: 227; Комиссаров, 1999: 126-127; Багдасарян, 2018: 296].

Интересно также отметить, что Г. Тури [Toury, 1979: 228] предлагает рассматривать получающийся в результате перевода *interlanguage* с двух позиций: переводческого идиолекта и переводческого диалекта. В случае идиолекта получающийся в результате перевода вариант промежуточного языка представляет собой языковую систему, характерную для конкретного переводчика, или, в более узком смысле, систему, характерную для конкретного переводчика в процессе работы над конкретным текстом – в этом случае переводчик является единственным носителем данного конкретного варианта языка. В случае диалекта

исходным пунктом является положение о том, что переводческая деятельность, подобно любой форме социальной деятельности, в значительной степени регулируется интерсубъективными нормами: переводчики следуют определённым правилам и конвенциям, общепринятым в профессиональном сообществе [Захарова, 2011: 138-143]. Регулирующее воздействие этих норм приводит к формированию совокупности сходных лингвистических черт переводов и переводческих решений в рамках конкретных языковых пар. Данное обстоятельство позволяет рассматривать переводческий язык как самый настоящий диалект, что, в свою очередь, подтверждается эмпирическими исследованиями (См. [Koppel, Ordan, 2011: 1318-1326; Nisioi et al., 2022: 1-14]).

Наконец, наиболее известным взглядом на феномен переводческого языка в современной научной литературе стала концепция *переводческих универсалий* (*translation universals*), выдвинутая в основополагающей статье М. Бейкер «Corpus Linguistics and Translation Studies: Implications and Applications» [Baker, 1993]. М. Бейкер предложила рассматривать феномен переводческого языка как особое, универсальное для всех переводных текстов явление, проистекающее из самого процесса перевода как такового, не зависящее от задействованной в нем языковой пары и отличающее переводы от непереводных текстов, изначально созданных на ПЯ [Baker, 1993]. Для обозначения составляющих это явление характеристик исследователь ввела термин *переводческие универсалии* [Baker, 1993: 242-245], к которым отнесла следующее:

- повышение уровня эксплицитной выраженности имплицитно заложенной в исходном тексте информации;
- тенденцию к устранению двусмысленности, к упрощению текста перевода на уровне лексики, морфологии и синтаксиса;
- использование в переводе более традиционных, конвенциональных грамматических структур ПЯ;
- тенденцию к избеганию лексических повторов либо через опущение повторяющихся в оригинальном тексте лексем, либо посредством их перефразирования / замены на другие лексем;

- преувеличение черт ПЯ в переводе, связанное со стремлением компенсировать тот факт, что изначально переводимый текст не предназначался для функционирования в принимающей культурной и языковой среде;
- специфическое распределение в переводных текстах некоторых языковых элементов в сравнении с непереводаемыми текстами на том же языке, в частности повышенная частота встречаемости в переводах единиц или структур, в большей степени свойственных системе языка оригинала, нежели системе ПЯ.

Именно эта работа [Baker, 1993] стала отправной точкой для масштабных эмпирических исследований феномена переводческого языка и во многом определила теоретико-методологическую повестку корпусного переводоведения на последующие десятилетия [Chesterman, 2010: 40]. Вместе с тем она подняла принципиальный вопрос, вокруг которого сосредоточилась дальнейшая научная дискуссия: являются ли отличительные особенности переводных текстов подлинно универсальными характеристиками переводов (т. е. *универсалиями перевода*), или же они представляют собой лишь вероятностные характеристики, наличие которых в переводе зависит от множества переменных? В силу значимости данного вопроса в контексте исследований феномена переводческого языка сосредоточимся на нем более подробно в следующем разделе.

### **1.3. Характеристики переводных текстов как универсалии перевода и как исследовательские гипотезы**

Трактовка характеристик переводных текстов как *переводческих универсалий* породила обширную научную дискуссию. Ряд исследователей [Olohan, 2002: 419-429; Chesterman, 2004a: 1-13; Chesterman, 2004b: 33-49; Toury, 2004: 15-32] поставил под сомнение как правомерность самого термина *универсалия* применительно к характеристикам переводных текстов, так и методологические основания для подобных обобщений.

Г. Тури [Toury, 2004: 15-32] утверждает, что, поскольку характеристики перевода всегда зависят от целого ряда факторов (языковой пары, принимающей культуры, стиля, жанра текста и т. д.), было бы некорректным считать эти

характеристики универсальными, приписывая их абсолютно всем переводным текстам без исключения лишь на том основании, что они являются переводными. Он предложил формулировать утверждения о свойственности переводам тех или иных характеристик как вероятностные утверждения, зависящие от множества переменных [Toury, 2004: 15-32; Toury, 2012: 9-10]. Такие утверждения не должны формулироваться в жёстко детерминированной форме типа: «*Если в переводе наблюдается А, то отсюда непременно следует В*». Скорее, это должны быть условные, вероятностные утверждения типа: «*Если в переводе наблюдается А, В и С, то с высокой вероятностью можно утверждать, что D*». При этом даже накопление значительного массива эмпирических фактов в пользу характеристики D не будет давать оснований для признания её универсальной. В данном случае Г. Тури предлагает [Toury, 2004: 15-32] рассматривать такие утверждения как своего рода законы (*laws*), которые, в отличие от универсалий (универсальных по определению и, следовательно, применимых ко всем случаям перевода), могут уточняться на основе данных, получаемых в процессе новых дескриптивных исследований. Более того, законы допускают наличие исключений, которые можно было бы объяснить, сформулировав в свою очередь другие законы. Иными словами, законы обладают большей гибкостью и объяснительной силой, что особенно важно в такой (в первую очередь эмпирической) дисциплине, как переводоведение.

Э. Честерман [Chesterman, 2004a: 1-13; Chesterman, 2004b: 33-49] занимает близкую позицию, настаивая на том, что предполагаемые в качестве универсальных характеристики переводов следует рассматривать не как установленные факты, в равной мере присущие всем переводным текстам, а как гипотетические характеристики, лишь потенциально свойственные переводам в той или иной степени и требующие строгой эмпирической верификации. Их гипотетичность обусловлена тем, что выявляемые в результате анализа переводов характеристики совсем необязательно несут абсолютный, свойственный всем переводам характер и могут зависеть от разнообразных факторов.

Проблема универсальности характеристик переводных текстов становится особенно очевидной, если сравнить понятие ‘*переводческие универсалии*’ с понятием ‘*языковые универсалии*’ (ср. Успенский 1970: 5-30; Гринберг и др. 1970: 31-44). Под языковой универсалией понимается свойство, присутствующее во всех без исключения или практически во всех естественных языках мира. При этом любая гипотеза о языковой универсалии принципиально верифицируема на значительной, однако конечной выборке существующих в мире языков. В то время как ситуация с переводческими универсалиями принципиально иная: совокупность всех существующих в мире переводных произведений представляет собой массив данных несопоставимо большего объёма, что делает их полноценную верификацию практически неосуществимой. Данное различие ставит под сомнение правомерность использования термина *универсалия* применительно к выявляемым в переводах характеристикам и требует более осторожного подхода к формулированию теоретических обобщений в данной области. В связи с этим более корректным представляется называть их *потенциально универсальными характеристиками переводов*.

Для систематического изучения таких характеристик Э. Честерман [Chesterman, 2004a: 1-13] предлагает использовать четыре типа исследовательских гипотез: объяснительные, дескриптивные, предсказательные и интерпретативные (См. Табл. 1).

Таблица 1 – Типы гипотез о переводческих универсалиях (по Э. Честерману)

<b>Тип гипотезы</b>	<b>Функция</b>	<b>Вопрос, на который стремится ответить</b>
Объяснительная гипотеза ( <i>Explanatory Hypothesis</i> )	Объясняет причины возникновения универсальных характеристик в переводах.	Почему переводные тексты обладают общими характеристиками?
Описательная гипотеза ( <i>Descriptive Hypothesis</i> )	Фиксирует наблюдаемые закономерности в переводах, позволяет идентифицировать тексты как переводы, отличая их от непереводных текстов; позволяет классифицировать характеристики переводов по категориям.	Какие общие характеристики можно потенциально наблюдать во всех переводных текстах?

Продолжение таблицы 1

Тип гипотезы	Функция	Вопрос, на который стремится ответить
Прогностическая гипотеза ( <i>Predictive Hypothesis</i> )	Формулирует прогнозы на основе объяснительных и описательных гипотез, позволяя проверять их на новых данных.	Будут ли эти характеристики проявляться в новых, еще не проанализированных переводах?
Интерпретативная гипотеза ( <i>Interpretive Hypothesis</i> )	Дает возможность концептуально осмыслить изучаемое явление, предлагая новые подходы к его интерпретации.	Как можно интерпретировать изучаемое явление?

Объяснительные гипотезы представляют собой утверждения, устанавливающие причинно-следственные связи между наблюдаемыми в переводе характеристиками и влияющими на их появление факторами.

Дескриптивные гипотезы описывают, какими характеристиками обладают переводные тексты, и на основе этих характеристик (предположительно) позволяют идентифицировать тот или иной текст как переводной и отличить его от непереводных текстов, а также позволяют выделить наблюдаемые характеристики переводов в отдельные категории. Иными словами, дескриптивные гипотезы обладают дифференцирующей и классифицирующей функциями.

Предсказательные гипотезы необходимы для проверки объяснительных и дескриптивных гипотез. На основе объяснительных гипотез формулируются предсказания о том, что при определенных условиях  $A$ ,  $B$  и  $C$  изучаемое явление  $D$  будет наблюдаться на новом исследовательском материале. В то время как на основе дескриптивных гипотез строятся предсказания о наличии описываемых характеристик в новых случаях, подобных уже изученному. Иными словами, формулирование дескриптивной гипотезы служит способом описания характерных признаков изучаемого явления, а ее проверка помогает установить, применимо ли данное описание ко всем похожим случаям. Например, если дескриптивная гипотеза утверждает, что нечто (в данном случае – перевод) обладает признаком  $F$ , то предсказательная гипотеза предсказывает, что каждый новый случай этого явления (перевода) предположительно будет обладать признаком  $F$ .

Наконец, интерпретативные гипотезы предлагают способы осмысления и интерпретации изучаемого явления. Э. Честерман считает, что исследование универсальных характеристик переводов следует начинать с предварительного формулирования интерпретативной гипотезы о том, что следует считать такими характеристиками. Он отмечает [Chesterman, 2004a: 8], что при исследовании потенциально универсальных характеристик переводов следует особенно сосредоточиться на выявлении сходств и различий между переводами и непереводами текстами, поскольку их сходства указывали бы на естественность тех или иных особенностей переводных текстов в контексте языка перевода, в то время как различия могли бы свидетельствовать об обратном.

Э. Честерман приводит пример механизма, согласно которому проверяются гипотезы о наличии универсальных характеристик переводов на примере *интерференции* исходного языка [Chesterman, 2004a: 8] Процедура включает пять последовательных шагов:

(1) исследователь фиксирует различия между переводными текстами и непереводами текстами на ПЯ, по причине которых переводы кажутся не совсем естественными;

(2) исследователь пытается выяснить причину такой неестественности, сравнивая языковые средства, используемые в переводах, со средствами выражения, характерными для языка оригинала, либо конкретные тексты перевода с их оригиналами, в результате чего обнаруживается ряд формальных сходств;

(3) исследователь выдвигает предположение, что неестественные черты переводов объясняются интерференцией исходного языка / текста, который как бы «навязывает» переводчикам использование выявленных на предыдущем этапе форм выражения, в большей мере свойственных языку оригинала, нежели ПЯ;

(4) для проверки универсальности исследуемой характеристики исследователь расширяет спектр (корпус) анализируемых переводов, включая в анализ тексты других типов;

(5) в случае подтверждения наличия исследуемой характеристики на новом материале вероятность её универсальности возрастает.

Также выделяются два принципиально разных подхода к изучению потенциальных универсальных характеристик переводов [Chesterman, 2004b: 40-41]:

1. Подход «сверху вниз» (формулирование гипотетического утверждения, претендующего на универсальность → проверка на большом корпусе переводных текстов → ограничение / уточнение изначального утверждения). При использовании данного подхода исследователь выдвигает универсальную гипотезу о наличии в переводных текстах определенных универсальных характеристик, изначально подразумевая применимость данной гипотезы ко всем переводам вообще. После чего исследователь анализирует переводные тексты, чтобы проверить, справедливо ли для них это утверждение, а затем по необходимости принимает решение о том, нужно ли скорректировать данную гипотезу (например, сузить её применимость до какого-то конкретного спектра переводных текстов).

Недостатком данного подхода является высокая степень риска чрезмерного обобщения и необходимость проверки гипотез на максимально репрезентативном корпусе данных, который отражал бы ситуацию по всем существующим переводам и типам переводов вообще (имеются в виду разные языки, стили, жанры и т. д.), что на практике представляется крайне затруднительным, если вообще возможным.

2. Подход «снизу вверх» (формулирование частного утверждения для конкретной группы текстов, ограниченной рядом выборочных критериев → проверка на конкретном корпусе текстов → постепенное расширение области применения сформулированного утверждения и увеличение спектра анализируемых текстов). При использовании данного подхода исследователь ограничивает спектр анализа конкретным подмножеством текстов, отобранных на основе ряда критериев. Среди таких критериев могут быть:

- Вовлеченная в процесс перевода языковая пара и направление перевода;
- Временной период и культура, в условиях которых выполняется перевод;

- Тип переводимого текста (текст определенного стиля и жанра, определенного назначения);
- Переводчик или группа переводчиков, объединенных общими признаками (уровень подготовки, является ли язык перевода родным / неродным языком);
- Редакционные нормы или требования определенного издательства, в котором публикуется перевод.

Данный подход позволяет установить типичность или нетипичность определенных характеристик для некоторого подмножества переводных текстов, то есть подтвердить или опровергнуть выдвинутую на начальном этапе гипотезу относительно проанализированной группы текстов. Затем можно расширить спектр исследования, исследуя другие подмножества текстов, соответствующих иным критериям.

Преимуществами подхода «снизу вверх» являются высокая степень эмпирической достоверности и осторожность выводов, позволяющие избежать преждевременных утверждений об универсальности определенных характеристик для всех переводов вообще. С другой стороны, недостаток состоит в том, что движение к более широким обобщениям происходит медленнее и полученные на начальных этапах выводы имеют ограниченную сферу применимости.

Наиболее распространенной и методологически разработанной классификацией потенциально универсальных характеристик переводов в современной исследовательской практике является следующая их классификация [Laviosa, 2089: 306-310; Куниловская и др., 2017: 17-27; Кантышева, 2021: 139]:

- *Упрощение* – характеристика переводных текстов быть морфологически, лексически и синтаксически проще непереводаемых текстов, изначально созданных на ПЯ;
- *Нормализация* – характеристика, выражающаяся в использовании в переводных текстах более стандартизированных структур ПЯ, в избегании лексических повторов и нетипичных для ПЯ коллокаций;

- *Экспликация* – характеристика, подразумевающая повышение объяснительного потенциала текстов перевода за счет более частого использования средств обеспечения когезии и когерентности текста;
- *Интерференция* – характеристика переводов, проявляющаяся в переносе в текст перевода особенностей, свойственных структурной организации исходного языка / текста.

Данная классификация обладает рядом существенных преимуществ: она опирается на обширный эмпирический материал, накопленный в рамках корпусных исследований за последние десятилетия, а каждая из четырёх характеристик операционализируется через формальные индикаторы, поддающиеся количественному измерению. Последнее обстоятельство определяет и её совместимость с корпусным подходом, принятым в настоящем исследовании: наличие каждой характеристики верифицируется путём систематического сопоставления её индикаторов на материале подкорпусов переводных и непереводных текстов на ПЯ, что позволяет выявлять статистически значимые различия между подкорпусами и делать выводы о том, какие из обнаруженных особенностей являются специфичными именно для переводов. С учётом изложенного данная классификация принята в качестве методологической основы настоящего исследования.

Поскольку настоящее исследование сосредоточено на переводе научных текстов, адекватная интерпретация проявлений феномена переводческого языка требует учёта норм научного стиля, характерных для задействованной языковой пары. Это обусловлено тремя обстоятельствами. Во-первых, логично предположить, что феномен переводческого языка проявляется по-разному в текстах той или иной стилистической принадлежности, а оценка проявлений его характеристик должна проводиться с учетом норм соответствующего функционального стиля. Во-вторых, английский язык выступает в качестве языка общемировой науки, из-за чего возникает риск проникновения некоторых особенностей англоязычного научного стиля, нарушающих нормы ПЯ, в русскоязычный научный дискурс через переводы. В-третьих, для адекватной

идентификации индикаторов переводческого языка необходимо четко разграничивать, какие явления в переводах обусловлены влиянием ИЯ и представляют собой отклонения от норм русскоязычного научного стиля, а какие отражают естественную вариативность в рамках научного стиля русского языка. В связи с этим следующий раздел посвящён рассмотрению основных характеристик научного стиля в русском и английском языках.

#### **1.4. Общая характеристика научного стиля в русском и английском языках**

Исследование характеристик феномена переводческого языка не может осуществляться без учета функционально-стилистической специфики анализируемого материала: различия между англоязычной и русскоязычной традициями в реализации научного стиля, степени его формализации, предпочитаемых лексических и грамматических средствах выражения способны существенно влиять на характер и интенсивность проявления изучаемого феномена. Следовательно, адекватная интерпретация и оценка выявляемых в ходе корпусного анализа характеристик требует рассмотрения ключевых особенностей в реализации научного стиля обоих языков. Кроме того, как будет показано в Главе 2, многие из рассмотренных далее особенностей, касающихся научного стиля русского языка, лягут в основу дополнительных индикаторов феномена переводческого языка, специально отобранных для оценки степени соответствия переводных текстов нормам русскоязычного научного стиля.

Итак, начнем с рассмотрения особенностей научного стиля, свойственных обоим языковым системам. Научный стиль – функциональный стиль книжной речи, используемый в научной сфере общения и речевой деятельности и обслуживающий различные отрасли науки, техники и производства [Кожина и др., 2008: 128-129]. Научный текст, являющийся воплощением данного стиля, отражает когнитивную и коммуникативно-речевую деятельность субъекта, целью которой является сообщение нового знания о реальности и доказательство его достоверности [Чернявская, 2007: 22]. В рамках научного стиля выделяется собственно научный подстиль [Плещенко и др., 1999: 36; Алексеева 2001: 165;

Владимирова, 2010: 7; Иваницкая, 2016: 67; Минакова, Краснова, 2017: 143-144], к жанрам которого относится научная статья – основной объект анализа в настоящем исследовании.

К стилиобразующим чертам научного стиля, определяющим отбор и организацию языковых средств, относятся:

- *логичность изложения*, проявляющаяся в последовательном развёртывании мысли со строгим соблюдением причинно-следственных связей [Пумпянский, 1981: 11-12, 18-19; Комиссаров, 1990: 132-134; Паршин, 2000: 63-62; Кожина и др., 2008: 289-311; Ван, 2011: 110];

- *точность и ясность изложения*, достигаемые посредством выверенного использования лексики [Васильева, 1976: 110; Пумпянский, 1981: 11-12; Плещенко и др., 1999: 37-38; Кожина и др., 2008: 289-311; Yakhontova, 2020: 1-4];

- *отвлечённость и обобщённость*, выражающиеся в широком употреблении абстрактной лексики для обозначения общих понятий, абстрактных предметов и явлений [Кожина и др., 2008: 289-311; Владимирова, 2010: 6; Иваницкая, 2016: 69];

- *объективность и некатегоричность*, проявляющиеся во взвешенности даваемых автором оценок относительно степени изученности обсуждаемой темы, эффективности той или иной теории, путей решения проблемы, степени завершенности результатов исследования и т. п., а также в отношении позиций других ученых и своей личной [Пумпянский, 1981: 26-27; Комиссаров, 1990: 110; Паршин, 2000: 63-62; Плещенко и др., 1999: 36; Кожина и др., 2008: 289-311; Владимирова, 2010: 6; Yakhontova, 2020: 1-4; Гуревич, 2024: 19-20];

- *информационная насыщенность*, проявляющаяся в высокой смысловой плотности текста, достигаемая посредством частого использования терминов и специализированной лексики, логически выверенных конструкций при минимуме избыточных языковых средств выражения [Комиссаров, 1990: 110; Плещенко и др., 1999: 36];

- *диалогичность*, т. е. направленность речи на адресата, например, с целью привлечения его внимания к особо важным фрагментам текста, приглашения к совместному рассуждению и т. д. [Кожина и др., 2008: 289-311];

- *безобразность изложения*, выражающаяся в избегании таких средств создания образности, как метафора, метонимия, гипербола и т. п., поскольку данные средства во многом вносят в текст элемент субъективных ассоциаций, несовместимый с требованиями объективной подачи научной информации [Гальперин, 1958: 424; Комиссаров, 1990: 130; Плещенко и др., 1999: 36].

- *отсутствие ярко выраженного проявления авторской индивидуальности и подчеркнутая авторская скромность*, обусловленные тем, что любое научное знание является результатом коллективного труда, в связи с чем автор выступает не от себя лично, а представитель научного сообщества, опираясь на достижения своих предшественников [Алексеева, 2001: 166].

Перечисленные характеристики требуют высокой степени формализации используемых языковых средств, вследствие чего научные тексты, создаваемые разными авторами, демонстрируют значительное единообразие в использовании лексических, морфологических и синтаксических структур [Алексеева, 2001: 167-168; Ван, 2011: 107-116].

При переводе научного текста определяющее значение при выборе лексических и грамматических средств выражения имеет соблюдение стилистических норм ПЯ [Федоров, 2002: 248]. Под нормой при этом понимается типичная реализация возможностей языковой системы в конкретной коммуникативной сфере, признаваемая носителями языка корректной и допустимой для данной ситуации общения [Арнольд, 2010: 318; Латышев, Северова, 2025: 59-60]. Тем не менее успешная реализация стилистических норм ПЯ невозможна без понимания сходств и различий научного стиля обоих языков, задействованных в процессе перевода. В связи с этим обратимся к рассмотрению некоторых особенностей научного стиля в русском и английском языках.

Начнем рассмотрение особенностей с *лексического уровня*. Одной из ключевых характеристик научного текста на лексическом уровне является обилие

терминов – слов и словосочетаний, выполняющих строго номинативную функцию, обладающих однозначностью в рамках конкретной терминологической системы и лишенных эмоциональной окраски [Комиссаров, 1990: 110-111; Плещенко и др., 1999: 37-38; Паршин, 2000: 64; Алексеева, 2001: 167; Федоров, 2002: 233; Владимирова, 2010: 8; Ван, 2011: 110; Беляева, 2013: 209-219; Беляева, 2015; Беляева, Камшилова, 2020: 108-116] (ср. определения Терминологического стандарта<sup>2</sup>). Передача терминологии при переводе всегда должна осуществляться с использованием однозначных соответствий-эквивалентов [Алексеева, 2001: 167], их замена синонимичными вариантами недопустима даже в целях облегчения восприятия представленной в тексте информации, поскольку приводит к нарушению терминологической консистентности текста и утрате строгости научного изложения [Алексеева 2001: 167; Кожина и др., 2008: 297]. К тому же, собственно научные тексты, к которым относится и научная статья [Пумпянский, 1981:13], предназначаются для узкого круга квалифицированных читателей, в полной мере обладающих необходимой для их понимания профессиональной компетенцией и хорошо знакомых со специализированной терминологией своей предметной области [Федоров, 2002: 247, 250]. Использование иных, схожих по значению вариантов перевода может затруднять восприятие текста и вводить читателя в заблуждение относительно сути описываемого явления. Помимо терминов, значительную долю лексики научного стиля составляет общенаучный слой лексики, или лексика общенаучного описания – совокупность стилистически нейтральных слов и словосочетаний, используемых преимущественно в научной сфере общения для описания материала исследования, толкования понятий и установления связей между ними [Комиссаров, 1990: 112-113; Плещенко и др., 1999: 38; Алексеева, 2001: 168; Котюрова, 2011: 258-260]. Поскольку исследование способов передачи терминологии и общенаучной лексики в переводе выходит за рамки настоящей работы, перейдем к рассмотрению других особенностей научного стиля.

---

<sup>2</sup> Определения Терминологического стандарта: <https://files.stroyinf.ru/Data/586/58694.pdf>

Следует отметить, что для лексического уровня научных текстов характерно преобладание абстрактной лексики (*мышление, свойство, влияние, процесс, закономерность; process, impact, dependence, degradation, radiation*) над конкретной, что обусловлено ориентацией научного текста на обобщение, установление закономерностей и формулирование общих теорий и принципов, а не на описание конкретных явлений или предметов [Плещенко и др., 1999: 37-38; Кожина и др., 2008: 291-292; Ван, 2011: 111].

Перейдём к рассмотрению характеристик научного стиля, проявляющихся на *грамматическом уровне*. В обоих языках научный стиль характеризуется частым употреблением имен существительных, что обусловлено номинативной природой научного текста [Васильева, 1976: 113; Biber, 2006: 13-14; Боженкова и др., 2019: 441-445; Ван, 2011: 109-110; Hyland, Jiang, 2021: 255-282; Kazemi, 2025: 29-55]. Номинативный характер проявляется через расширение сферы использования имени существительного за счет сужения сферы употребления глагола вследствие номинализации (субстантивации) – процесса, при котором глаголы, выражающие разнообразные процессы или действия, преобразуются в отглагольные существительные [Алексеева 2001: 167; Кожина и др., 2008: 294; Балыгина, Круковская, 2018: 3-10; Боженкова и др., 2019: 441-442].

Номинализация выполняет в научном тексте несколько важных функций: создаёт восприятие описываемого явления как твердо установленного, объективно существующего факта, что в целом способствует восприятию отстаиваемой автором позиции как более объективной и достоверной [Kazemian et al., 2013: 151-152; Bello, 2016: 1-23; Kazemi, 2025: 29-55]; позволяет сократить длинные синтаксические конструкции, как бы инкапсулируя несколько лексических единиц в одну клаузу и тем самым делая текст более компактным, информационно насыщенным, а также увеличивая его лексическую плотность [Ван, 2011: 110-111; Kazemian et al., 2013: 151-152]. Ср.:

(1) «*When we **analyzed** the data carefully, we discovered that the temperature **had increased** significantly, which had caused the reaction to **accelerate** rapidly.*» →

«Careful data **analysis** revealed significant temperature **increase**, causing rapid reaction **acceleration**.» (пример мой – О. У.).

(2) «После того, как мы **тщательно проанализировали** полученные данные, мы обнаружили, что температура значительно **повысилась**, что привело к тому, что реакция быстро **ускорилаь**.» → «Тщательный **анализ** полученных данных выявил значительное **повышение** температуры, вызвавшее стремительное **ускорение** реакции.» (пример мой – О. У.).

В русском языке номинализация глаголов происходит путём использования суффиксов *-ье, -ни(е), -нь(е)* (*изучение, охлаждение*), *-стви(е)* (*следствие*), *-ти(е), -ть(е)* (*развитие*), *-к(а)* (*сушка, плавка, нарезка, регулировка*), *-ци(я)* (*дезинфекция, инспекция, эмиграция*) [Васильева, 1976: 87-107; Земская, 2011: 258]; в английском – посредством герундиальных форм (*-ing*), конверсии и суффиксов *-tion/-sion, -ment, -ence/-ance, -age* и др. [Viber et al., 2021: 320-325]. Важно отметить, что тенденция к субстантивации глаголов английскому языку свойственна в меньшей мере, чем русскому, что в ряде случаев обуславливает необходимость замены глаголов текста оригинала на их субстантивированные версии в тексте перевода [Комиссаров, 1990: 132; Бреус, 2000: 34-39; Балыгина, Круковская, 2018: 3-10; Yakhontova, 2020: 1-4].

Если говорить о грамматических характеристиках имен существительных, то в русскоязычных научных текстах преобладают существительные в форме единственного числа, что обусловлено (1) высокой долей отглагольных существительных, как правило, употребляющихся в форме единственного числа (*изучение, рассмотрение* и т. д.), а также (2) употреблением в форме единственного числа существительных, обозначающих вне научного стиля конкретные предметы, для обозначения целого класса объектов (Например: *Берёза хорошо переносит мороз*, где слово *береза* выступает как представитель класса *берез* вообще, а не как конкретное дерево) [Плещенко и др., 1999: 40-41; Кожина и др., 2008: 293]. При этом в обоих языках абстрактные и вещественные существительные, являющиеся вне научного стиля неисчисляемыми, часто употребляются в форме множественного числа, приобретая значение конкретности (*шумы в сердце / heart*

*murmurs*) или обозначая разновидности вещества или материала (*fats / жиры, oils / масла, greases / смазки, gasolines / бензины.*) [Арнольд, 2010: 342; Владимирова, 2010: 9; Ван, 2011: 110; Боженкова и др., 2019: 442]. При этом в научных текстах на английском языке существительные в форме множественного числа в целом употребляются достаточно часто [Комиссаров, 1990: 117].

В научных текстах на русском языке, которому свойственна грамматическая категория рода, достаточно часто употребляются имена существительные среднего рода, поскольку им чаще, чем существительным мужского и женского рода, свойственно отвлеченно-обобщенное значение (например, свойство, качество, явление, состояние, действие, условие) [Кожина и др., 2008: 293].

Другой активно употребляемой частью речи в научном тексте является имя прилагательное, что связано с необходимостью обозначения видового признака описываемых предметов и явлений, с логико-понятийной конкретизацией имен существительных и также свидетельствует об именованном характере научного стиля в обоих языках [Viber, 2006: 13-14].

В русскоязычном научном тексте при образовании сравнительной и превосходной степеней сравнения преобладают аналитические формы (*более интересный, наиболее значимый*), тогда как синтетические формы на *-ее (сильнее)* и на *наи-* (*наибольший*) встречаются значительно реже, а формы на *-ей (сильней)* и *-ейш-, -айш-* (*красивейший, высочайший*) в силу присущей им эмоционально-экспрессивной окрашенности для научной речи нетипичны [Плещенко и др., 1999: 41; Боженкова и др., 2019: 443]. В английском языке в текстах научного стиля чаще используются многосложные прилагательные (типа *efficient, sophisticated, significant* и т. д.), что приводит к более частому употреблению аналитических форм степеней сравнения по сравнению с другими стилями [Viber et al., 2021: 521-523]. При этом использование в научном тексте превосходной степени в целом ограничено требованиями объективности научного изложения.

В русском языке, где одновременно существуют полная и краткая формы прилагательных, в научном тексте для обозначения постоянного признака предметов или явлений чаще употребляются краткие формы (Ср.: *Сталь упруга при*

нагрузке) [Плещенко и др., 1999: 41; Паршин, 2000: 80; Кожина и др., 2008: 294; Владимирова, 2010: 9].

Что касается глаголов, то их доля в текстах научного стиля снижена за счет глагольной номинализации (субстантивации), вследствие которой значение процессуальности передается имени существительному [Комиссаров, 1990: 115]. Глаголы в научной речи, как правило, не описывают конкретное действие, а либо фиксируют факт протекания процесса (*наблюдается, происходит; occurs, is observed*), либо задают логические отношения (*приводит к; causes, leads to, results in*), либо устанавливают внутритекстовые связи (*свидетельствует о; proves, indicates*) [Halliday, Martin, 2003: 70-72]. Глаголы, обычно употребляющиеся в конкретном значении, десемантизируются (Ср.: *Она следовала за мной* и *Отсюда следует вывод, что...*; *The guide led the group to...* и *This led to a significant reduction in...*) [Васильева, 1976: 116-117; Плещенко и др., 1999: 43; Кожина и др., 2008: 291-292]. Глаголы, обычно употребляющиеся в конкретном значении, зачастую десемантизируются (Ср. *Она следовала за мной* и *Отсюда следует вывод, что...*; *The guide led the group to...* и *This led to a significant reduction in error rates*) [Васильева, 1976: 116-117; Плещенко и др., 1999: 43; Кожина и др., 2008: 291-292].

Подавляющее число глаголов в научных текстах обоих языков употребляется в форме настоящего времени во вневременном значении [Комиссаров, 1990: 116; Кожина и др., 2008: 291-292; Владимирова, 2010: 8-9; Боженкова и др., 2019: 442]. В английском языке это, как правило, форма настоящего простого времени (*The Present Simple Tense*) [Поспелова, 2013: 279; Камышова, Колмакова, 2015: 197]. Форма настоящего времени позволяет обозначать свойства предметов и явлений, делая акцент на том, каким предмет / явление является *вообще* или каким образом он(о) ведет себя *всегда*, а не в данный конкретный момент времени (Ср. *Под воздействием тепла тела расширяются; Sound travels faster in water than in air*). Во вневременном значении нередко используется и форма будущего (Ср. *Скорость света всегда будет оставаться постоянной величиной; The oil will float on water, creating a rainbow sheen on the surface*), а иногда и прошедшего времени (Ср.,

*Звёзды образовывались из газопылевых облаков; The stars were formed in molecular clouds*) [Кожина и др., 2008: 291-292; Quirk et al., 1985: 228; Leech, 2013: 106-107].

В русском языке с его глагольной категорией вида глаголы несовершенного вида в научном тексте используются значительно чаще глаголов совершенного вида (около 80% всех глагольных форм), поскольку обладают более отвлеченно-обобщенным значением [Плещенко и др., 1999: 42].

Формы страдательного (пассивного) залога являются одним из ключевых средств реализации объективно-нейтральной манеры изложения в научном тексте в обоих языках [Пумпянский, 1981: 11-12, 18; Комиссаров, 1990: 116; Арнольд, 2010: 341; Гуревич, 2024: 19-20]. Страдательные конструкции позволяют (1) сместить фокус с личности исследователя на объект и результаты исследования, на описываемые в тексте факты [Ding, 1998: 117-135; Halliday, Martin, 2003: 212-213], (2) являются отражением авторской скромности, поскольку любые результаты научных исследований и научные достижения есть следствие труда не одного только автора текста, а целого коллектива ученых, в том числе и ученых-предшественников [Пумпянский, 1981: 295-296], а также (3) позволяют подчеркнуть универсальность и воспроизводимость научных наблюдений и выводов, что соответствует требованиям к объективному изложению [Пумпянский, 1981: 295-296]. Кроме того, страдательный залог служит удобным инструментом для описания действий, производитель которых либо неизвестен, либо очевиден из контекста, либо же не важен для заданной ситуации общения [Viber et al., 2021: 166-167].

В русскоязычном научном тексте активно употребляются глаголы с постфиксом *-ся* в страдательном значении (*проблема рассматривается*), а также глаголы в форме кратких (*вопрос будет рассмотрен*) или полных (*описанный эксперимент*) причастий [Боженкова и др., 2019: 444-445]. Вместе с тем нельзя не отметить растущую в современном английском языке тенденцию к все более частому употреблению глаголов в форме активного залога и закономерному снижению доли глаголов в форме пассивного залога [Leong, 2021: 77-102; Banks, 2017: 1-17; Арцишевская, 2025: 104-113]. Такой сдвиг объясняется стремлением

сделать акцент на личном вкладе автора в процесс и результат исследования и обозначить его ответственность за делаемые в рамках исследования заявления и выводы. Упомянутая тенденция нередко находит отражение в рекомендациях по стилю<sup>3</sup> и в редакционной политике англоязычных научных журналов.

Если говорить о местоимениях, их употребление в научной речи во многом подчинено требованиям *объективности* и *логической связности* текста. Требование *объективности* обуславливает правила употребления личных и притяжательных местоимений. Так, важной особенностью в контексте научной письменной коммуникации является особый статус автора и адресата. Несмотря на то что автор научного текста всегда известен, в русскоязычном тексте, как правило, отсутствуют явные проявления авторской индивидуальности в виде личных местоимений 1-го лица единственного числа. Изредка автор текста обнаруживает себя в виде авторского *мы*, что позволяет сохранить объективность при необходимости выражения авторской позиции [Алексеева 2001: 166-167; Владимирова, 2010: 10]. Форма 1-го лица множественного числа личного и притяжательного местоимения (*как нам кажется, с нашей точки зрения, мы предполагаем*) также может выражать позицию коллектива авторов, подчёркивать коллективный характер научной деятельности и / или вовлекать читателя в совместный с автором процесс рассуждения и доказательства [Арнольд, 2010: 340]. В противоположность русскоязычному научному стилю, в англоязычных научных текстах часто можно встретить местоимения первого лица единственного и множественного числа (*I* и *we*) из-за упоминавшейся выше тенденции к употреблению глаголов в форме активного залога [Щемелева, 2013: 143-153]. Требование *логической связности* обеспечивается употреблением указательных, а также личных и притяжательных местоимений 3-го лица [Боженкова и др., 2019: 443-444]. Личные и притяжательные местоимения 3-го лица используются для отсылки к ранее упомянутым личностям, объектам, явлениям, исследованиям, позициям других ученых (*они доказали, что...; в их модели главным элементом*

---

<sup>3</sup> How to Use Active Voice in the Sciences: <https://www.uc.utoronto.ca/active-voice-in-science>, <https://advice.writing.utoronto.ca/types-of-writing/active-voice-in-science/>

являлось...; *in their model, the key parameter is...; ...its surface area increased twofold*).

Указательные местоимения обеспечивают связность текста, помогая эксплицитировать логические связи между высказываниями, фрагментами высказываний (*такой подход позволяет...; тот же эффект наблюдается при...; this finding suggests...; this raises the question of...*).

Требование логичности, последовательности, аргументированности и доказательности научного изложения наиболее ярко проявляется на синтаксическом уровне [Кожина и др., 2008: 295; Ван, 2011: 110]. В первую очередь всем этим целям служит широкий спектр языковых средств, обеспечивающих синтаксическую связность текста через установление причинных, временных, пространственных и др. отношений между описываемыми в тексте объектами и явлениями реального мира, этапами рассуждения и аргументации, а также между фрагментами самого текста (абзацами, предложениями и частями предложений) [Комиссаров, 1990: 117, 132-134; Владимирова, 2010: 10]. К таким средствам относятся союзы / союзные слова, некоторые местоимения и наречия, а также целые словосочетания (*в дополнение к, в результате чего, as a result, in contrast to, in addition to, owing to* и др.) [Арнольд, 2010: 338-339; Ван, 2011: 110]. Для научного текста в целом весьма характерно употребление т. н. вводных метатекстовых компонентов [Вежбицкая, 1978: 402-421] – слов или словосочетаний, структурирующих текст и создающих пошаговую логику научного рассуждения или доказательства (*во-первых, во-вторых, итак, таким образом; first, to begin with, next, then, finally*). Считается, что без подобных связующих элементов научный текст теряет свою плавность, становясь отрывистым и скачкообразным, что, в свою очередь, делает его менее понятным и убедительным.

По насыщенности такими связующими элементами научный стиль занимает первое место среди всех функциональных стилей [Ван, 2011: 110]. Однако, как указывает В. Н. Комиссаров [Комиссаров, 1990: 133], англоязычные научные тексты чаще опираются на имплицитное выражение логических отношений, которые могут быть выведены из содержания самих высказываний, в то время как

в русскоязычной научной речи предпочтение чаще отдается эксплицитным средствам маркировки логических связей. Ср:

*«The data were obtained exclusively through online self-report questionnaires, combining Likert-scale items with brief open-ended prompts. The findings should be interpreted cautiously.»* → «Сбор данных проводился только через онлайн-опросники для самоотчета, включавшие варианты ответов по шкале Ликерта и краткие открытые вопросы. **В связи с этим** полученные результаты следует интерпретировать с осторожностью.» (примеры мои – О. У.).

Отличительной чертой русскоязычного научного стиля является широкое использование т. н. *номинативных рамочных конструкций* с нетипичным для английского языка порядком слов [Паршин, 2000: 78]. Они представляют собой структуры, в которых имя существительное имеет в препозиции развернутое определение, нередко включающее в свой состав предложную конструкцию. Такие конструкции строятся по схеме: [*развернутое определение с причастием / именем прилагательным + (предложная конструкция)*] + *имя существительное*. Так, в русскоязычном научном тексте широко употребительны варианты типа:

- *характерные для данного класса веществ свойства;*
- *полученные в лабораторных условиях образцы;*
- *определённые на предыдущем этапе методологические ограничения;*

В английском подобные конструкции в препозиции невозможны: определяющая группа размещается строго в постпозиции к существительному. Ср.:

- *\* typical of this class of substances properties* → *properties typical of this class of substances;*
- *\* obtained in laboratory conditions samples* → *samples obtained in laboratory conditions;*
- *\* defined at the previous steps methodological limitations* → *methodological limitations defined at the previous steps;*

Учитывая, что номинативные рамочные конструкции являются частью стилистической нормы русского научного стиля и не свойственны стилю

английского языка, можно ожидать их недостаточную представленность в переводах научных текстов с английского языка на русский.

Итак, рассмотрев некоторые особенности научного стиля в английском и русском языках, в следующей главе сосредоточимся на подробном описании особенностей потенциально универсальных характеристик переводов, а также их формальных индикаторов, уделив особое внимание способам расчета соответствующих этим индикаторам метрик.

### **Выводы по главе 1**

В настоящей главе были рассмотрены исторические и теоретические основания, на которых строится настоящее исследование: предпосылки исследований феномена переводческого языка в переводоведении, подходы к его определению, концепция потенциально универсальных характеристик переводов и подходы к исследованию этих характеристик, а также функционально-стилистические особенности научного стиля в английском и русском языках. Рассмотренные положения позволяют сделать следующие выводы:

1. Возникновение научного интереса к феномену переводческого языка было обусловлено теоретико-методологическими сдвигами в переводоведении последней четверти XX века: методологический поворот к дескриптивному исследованию переводов, переосмысление их статуса как самостоятельного объекта научного изучения, а также признание переводного дискурса особым элементом дискурсивного пространства принимающей культуры в совокупности создали необходимые предпосылки для постановки вопроса о природе и характере специфических особенностей переводных текстов.

2. В современном переводоведении отсутствует единая трактовка феномена переводческого языка. Подходы к его осмыслению варьируются от оценки как нежелательного явления – результата отрицательной интерференции исходного языка и текста – до нейтральных концепций языковых «отпечатков» исходного языка (М. Геллерштам), «третьего кода», возникающего на стыке двух языковых систем (У. Фроули), и промежуточного языка (*interlanguage*) (Г. Тури).

Несмотря на разнообразие трактовок, все они признают объективное существование специфических характеристик, систематически проявляющихся в переводах и отличающих их от непереводаемых текстов на ПЯ.

3. Концепция переводческих универсалий М. Бейкер, определившая теоретическую повестку корпусного переводоведения на несколько десятилетий, поставила принципиальный вопрос о подлинной универсальности характеристик переводных текстов. Критический анализ данной концепции, предпринятый Г. Тури и Э. Честерманом, показал, что утверждать абсолютную универсальность выявляемых в переводах характеристик некорректно ввиду зависимости проявления последних от разнообразных переменных (языковой пары, жанрово-стилистической специфики текста, культурного контекста и т. д.). В связи с этим в настоящей работе принят подход к характеристикам феномена переводческого языка как к *гипотетически* универсальным особенностям переводов. В качестве методологической основы исследования принята классификация, включающая четыре характеристики переводных текстов: *упрощение, нормализацию, экспликацию и интерференцию*. Данная классификация отвечает требованиям корпусного подхода, поскольку каждая из характеристик операционализируется через формальные индикаторы, поддающиеся количественному измерению на текстовом материале. В качестве подхода к исследованию в настоящей работе принят описанный в разделе 1.3. подход «снизу вверх», позволяющий избежать преждевременных обобщений об универсальности изучаемого феномена и обеспечивающий более высокую степень достоверности делаемых выводов.

4. Описанные в разделе 1.4. особенности научного стиля в английском и русском языках послужат теоретической базой для адекватной интерпретации результатов корпусного анализа проявлений феномена переводческого языка на материале переводов научных статей, а также лягут в основу дополнительных индикаторов изучаемого феномена, учитывающих нормы научного стиля ПЯ.

Таким образом, синтез рассмотренных теоретических положений позволяет перейти к детальному рассмотрению конкретных характеристик переводческого

языка и их индикаторов, которые будут использоваться для корпусного анализа в последующих главах настоящего исследования.

## ГЛАВА 2. ОТБОР ИНДИКАТОРОВ ДЛЯ ИССЛЕДОВАНИЯ ХАРАКТЕРИСТИК ФЕНОМЕНА ПЕРЕВОДЧЕСКОГО ЯЗЫКА

### 2.1. Характеристика упрощение и её индикаторы

Начнем рассмотрение характеристик феномена переводческого языка с характеристики *упрощение* (*simplification*). Данная характеристика основывается на предположении о том, что переводчики склонны делать текст перевода лексически и грамматически более простым вследствие сознательного или бессознательного стремления избежать потенциальной неверной интерпретации текста читателем, которая может быть вызвана употреблением семантически неоднозначных или низкочастотных (редких) лексических единиц ПЯ, сложных синтаксических структур и пунктуационных паттернов [Baker, 1993: 244; Baker, 1996: 181-182].

Рассмотрим, как характеристика *упрощение* проявляется в переводе на различных языковых уровнях – лексическом, морфологическом и синтаксическом, – а затем перейдем к описанию её конкретных индикаторов, которые в дальнейшем будут использованы в практической части настоящего исследования для анализа данной характеристики на материале англо-русских переводов научных статей.

Считается, что на лексическом уровне *упрощение* выражается в снижении разнообразия используемой в текстах перевода лексики в сравнении с текстами непереводаемых произведений на том же языке [Lapshinova-Koltunski, 2015: 96; Volansky et al., 2015: 98-118]. Одновременно с этим наблюдается снижение доли так называемых *hapax legomena* – лексем, встречающихся в тексте единожды [Volansky et al., 2015: 98-118].

Другим проявлением *упрощения* на лексическом уровне является то, что переводы содержат меньше знаменательной и больше функциональной (служебной) лексики. К функциональной лексике в данном случае принято относить не только предлоги, союзы и частицы, но также и местоимения, поскольку они не обладают конкретным лексическим значением и выполняют в тексте референциальную функцию [Baker, 1996: 181-183; Laviosa- Braithwäite, 1997: 531-540; Laviosa, 1998: 557-570; Lapshinova-Koltunski, 2015: 97].

Ещё одним проявлением *упрощения* является более частое использование в переводах высокочастотной и общеупотребительной лексики языка перевода [Blum, 1978: 407-408; Laviosa-Braithwaite, 1997: 531-540; Laviosa, 1998: 557-570; Volansky et al., 2015: 98-118; Lapshinova-Koltunski, 2015: 96]: переводе лексем оригинального текста переводчики могут выбирать более частотные единицы ПЯ, предпочитая их менее распространенным (частотным) синонимичным вариантам.

В частности, лексическое *упрощение* в данном случае может проявляться через замену исходной лексемы с более узким значением (гипонима) на единицу с более общим значением (гипероним) в переводе [Blum-Kulka, Levenston, 1983: 119-139]. Например, английский глагол *gaze* может быть переведён словом с более общим значением – глаголом *смотреть*, в то время как возможным вариантом является глагол *вглядываться*, обладающий более узким значением и более точно передающий значение исходной единицы. При этом если обратиться к частотному словарю русской лексики [Ляшевская, Шаров, 2009: <http://dict.ruslang.ru/freq.php>], содержащему данные о частотности употребления приблизительно 50 000 слов русского языка, упорядоченных от наиболее частотных к наименее частотным, то обнаружится, что лексема *смотреть* будет находиться на 136-ой позиции в списке, в то время как лексема *вглядываться* – на 5394-ой.

Другим примером является замена стилистически маркированной лексики исходного текста на стилистически нейтральные (немаркированные), общеупотребительные единицы языка перевода, несмотря на существование в нём стилистически маркированных синонимов, более точно соответствующих единицам исходного текста [Blum-Kulka, Levenston, 1983: 119-139]. К примеру, лексема *commence*, принадлежащая к формальному регистру английского языка, в переводе может быть заменена общеупотребительным глаголом *начинать* (344-ая позиция в частотном словаре), тогда как более подходящим стилистическим эквивалентом в русском языке является глагол *приступить* (к) (7716-ая позиция) [Ляшевская, Шаров, 2009: <http://dict.ruslang.ru/freq.php>].

Подобные замены способны приводить к ряду нежелательных последствий в переводе: замена слов с более узким значением на слова с более широким

значением в ряде случаев может вести к утрате смысловой точности текста, в то время как замена стилистически маркированных единиц на нейтральные и общеупотребительные варианты – к нивелированию регистровых, жанрово-стилистических различий [Blum-Kulka, Levenston, 1983: 119-139].

Что касается грамматического уровня, здесь *упрощение* проявляется, как в морфологической структуре слов, так и в синтаксической организации предложений. На морфологическом уровне упрощение выражается в сокращении средней длины используемых в переводе лексем: переводчики зачастую отдают предпочтение структурно (морфологически) более простым единицам ПЯ [Volansky et al., 2015: 98-118; Pylypenko et al., 2021: 8596-8611]. Так, например, структурно более сложному варианту *продемонстрировать* может быть предпочтен морфологически более простой вариант *показать*.

На синтаксическом уровне сокращается средняя длина предложений: длинные и сложные предложения оригинального текста нередко разбиваются на более короткие и менее сложные в переводе [Baker, 1996: 181-183; Laviosa-Braithwaite, 1997: 531-540; Volansky et al., 2015: 98-118]. К примеру:

*«Therefore, the impact of the new dosing method on the anticancer effect of the drugs is difficult to discern, but the doses from the tables fall within the range of doses derived using the existing infant dosing methods»* [Balis et al. Dosing anticancer drugs in infants: current approach and recommendations from the Children's Oncology Group's Chemotherapy Standardization Task Force // Pediatric blood & cancer. – 2017. – Volume 64, № 11. – DOI: 10.1002/pbc.26636] →

*«Следовательно, сложно выявить влияние нового метода дозирования на противоопухолевый эффект препарата. При этом дозы, назначенные в соответствии с разработанной таблицей, попадают в диапазон доз, полученных с использованием существующих методов дозирования для детей первого года жизни»* [Балис и др. Дозировка противоопухолевых препаратов у детей первого года жизни: современный подход и рекомендации рабочей группы по стандартизации химиотерапии Детской онкологической группы / пер. с англ. Я. А.

Янгутовой // Российский журнал детской гематологии и онкологии. – 2018. – № 1. – С. 11-16]<sup>4</sup>.

Считается также, что на уровне синтаксиса в тексте перевода упрощается авторская пунктуация исходного текста: переводчики склонны заменять более «слабые» знаки препинания более «сильными», что делает текст более структурированным, менее неоднозначным и легче воспринимаемым. По порядку усиления знаки представлены следующим образом: запятая → точка с запятой → точка [Baker, 1996: 181-183; Malmkjaer, 1997: 151-162].

Для эмпирической проверки характеристики *упрощение* на основе работ, посвященных исследованию феномена переводческого языка, был выбран ряд формальных индикаторов, поддающихся количественному анализу на материале корпуса текстов. Следующие индикаторы послужат операциональной основой для анализа *упрощения* в настоящем исследовании:

*Коэффициент лексической вариативности (lexical variety, или type-token ratio)*, отражающий разнообразие используемой в тексте лексики [Munday, 1998: 545-546; Rápai, 2004: 143-164; Corpas Pastor et al., 2008: 75-81; Ilisei et al., 2010: 503-511; Kunilovskaya, Kutuzov, 2015: 33-40; Lapshinova-Koltunski, 2015: 97; Volansky et al., 2015: 98-118; Kunilovskaya, Lapshinova-Koltunski, 2019: 47-56; Kunilovskaya et al., 2021: 101-112; Kunilovskaya, Corpas Pastor, 2021: 133-180; Pylypenko et al., 2021: 8596-8611; Кантышева, 2021: 142-143; François, Lefer, 2022: 50-70]. Показатель рассчитывается как соотношение уникальных типов ко всем токенам, где типами являются уникальные лемматизированные лексические единицы, а токенами – все словоформы текста. Существуют разные способы подсчета *коэффициента лексической вариативности* [Grieve, 2007: 251-270], однако наиболее часто в исследованиях по теме используются следующие формулы [Volansky et al., 2015: 98-118]:

*Базовая версия коэффициента:*

---

<sup>4</sup> Примеры действия описываемых характеристик взяты из текстов, вошедших в специально сформированный исследовательский корпус настоящей работы (См. Раздел 3.1). Если источник не указан и не сопровождается дополнительной информацией, пример смоделирован автором данного диссертационного исследования.

$$TTR = \frac{V}{N} \times 100, \quad (1)$$

где  $V$  – общее количество типов (объем словаря текста),

$N$  – общее количество словоформ (токенов) в тексте (здесь и далее множитель 100 вводится для преобразования десятичных дробей в проценты для удобства сравнения значений показателя).

Важным ограничением базовой формулы коэффициента *лексической вариативности* является то, что зависимость количества типов от длины текста является нелинейной [Viber, 2006: 252-253]: по мере того, как размер текста увеличивается, рост числа уникальных лексем (типов) снижается, поскольку повторяемость уже использованных в тексте единиц возрастает. Именно поэтому в исследовательской практике применяются модифицированные версии коэффициента, компенсирующие нелинейный характер зависимости между количеством типов и длиной текста. В связи с этим дополнительно рассчитываются версии коэффициента, позволяющие нивелировать влияние длины текста на значение показателя, обеспечивая тем самым сопоставимость результатов при анализе текстов разного объёма. Данные версии коэффициента представлены ниже.

*Логарифмическая версия коэффициента*, нивелирующая влияние длины текста на значение показателя [Grieve, 2007: 251-270]:

$$\text{Log } TTR = \frac{\log V}{\log N} \times 100, \quad (2)$$

где  $V$  – общее количество типов (объем словаря текста),

$N$  – общее количество словоформ (токенов) в тексте.

*Версия коэффициента, учитывающая количество *hapax legomena* (слов, встречающихся в тексте один раз)* [Honoré, 1979: 172-177]:

$$\text{Modified } TTR = \frac{100 \times \log N}{1 - \frac{V_1}{V}}, \quad (3)$$

где  $V$  – общее количество типов (объем словаря текста),

$V_1$  – показатель количества *hapax legomena* в тексте перевода,

$N$  – общее количество словоформ (токенов) в тексте.

*Стандартизированная версия коэффициента (Standardized TTR, или STTR)*, вычисляемая не для всего текста целиком, а для последовательных фрагментов

(чанков) фиксированной длины (в 1000 токенов) [Scott, 2004]. Для каждого чанка отдельно рассчитывается лексическая вариативность по базовой формуле (см. формулу 1 выше), после чего подсчитывается среднее значение данного показателя для всех чанков текста. При этом оставшийся чанк, объем которого составляет менее 1000 токенов, из анализа исключается.

*Коэффициент лексической плотности текста (lexical density)*, отражающий соотношение т. н. содержательных слов (глаголов, имен существительных, имен прилагательных и наречий) относительно общего объема текста в словарных токенах [Baker, 1996: 181-183; Laviosa- Braithwäite, 1997: 531-540; Laviosa, 1998: 557-570; Corpas Pastor et al., 2008: 75-81; Ilisei et al., 2010: 503-511; Kunilovskaya, Kutuzov, 2015: 33-40; Lapshinova-Koltunski, 2015: 97; Kunilovskaya, Lapshinova-Koltunski, 2019: 47-56; Kunilovskaya et al., 2021: 101-112; Kunilovskaya, Corpas Pastor, 2021: 133-180; Pylypenko et al., 2021: 8596-8611; Кантышева, 2021: 142-143; François, Lefer, 2022: 50-70]. Показатель рассчитывается по формуле:

$$LD = \frac{V_{lex}}{N} \times 100, \quad (4)$$

где  $V_{lex}$  – общее количество содержательных слов,

$N$  – общее количество словарных токенов в тексте [Ure, 1971: 443-452].

*Средний ранг используемой лексики (mean word rank)*, отражающий степень использования в тексте высокочастотной / низкочастотной лексики ПЯ [Volansky et al., 2015: 98-118]. Показатель вычисляется на основе ранжирования слов текста по частоте их встречаемости в Национальном корпусе русского языка [Савчук и др., 2024: 7-34]. Формула показателя:

$$MWR = \frac{\sum_{i=1}^n R_i}{n}, \quad (5)$$

где  $R_i$  – ранг  $i$ -го слова в тексте,

$n$  – общее количество словарных токенов.

Ранг слова приравнивается к его порядковому номеру в частотном списке русской лексики [Ляшевская, Шаров, 2009: <http://dict.ruslang.ru/freq.php>]. Как и в [Volansky et al., 2015: 98-118], при подсчете ранга слов используются первые 6000

слов списка. Максимальный ранг слова равняется 5000. Для слов, не вошедших в частотный список, применяются два подхода:

- (1) Слову присваивается максимальный ранг 6000;
- (2) Слово исключается из анализа и не учитывается при вычислении ранга.

*Средняя доля наиболее частотных знаменательных слов подкорпуса* в тексте (*list head ratio*), отражающая соотношение числа токенов, принадлежащих к наиболее частотным леммам подкорпуса (т. н. ядро наиболее частотной лексики, или *list head*), относительно общего объёма текста [Laviosa-Braithwaite, 1997: 531-540; Laviosa, 1998: 557-570; Kunilovskaya, Kutuzov, 2015: 33-40]. Показатель основан на предположении о том, что в переводах топ-список частотных слов подкорпуса покрывает бóльшую долю токенов текста в сравнении с непереводаемыми текстами. В настоящем исследовании применяется модифицированная версия показателя с учётом следующих параметров:

*Фильтрация лексики:* Из анализа исключается функциональная лексика (предлоги, союзы, частицы, местоимения), а также высокочастотные десемантизированные глаголы (типа *быть, являться, иметь, делать, стать, становиться, оказаться, казаться*).

*Критерий формирования list head:* Отбираются леммы слов, принадлежащих к знаменательным частям речи и встречающихся в подкорпусе с частотой не ниже 0,001 (или 0,1%) от общего числа токенов (после применения фильтрации). Для текста рассчитывается отношение числа токенов, входящих в это ядро (т.е. в *list head*), к общему числу токенов текста по формуле:

$$LHR = \frac{V}{N} \times 100, \quad (6),$$

где  $V$  – количество токенов в тексте, принадлежащих к ядру наиболее частотной знаменательной лексики подкорпуса,

$N$  – общее количество словарных токенов в тексте (включая функциональную лексику и десемантизированные глаголы).

С целью анализа *упрощения* на грамматическом (морфологическом и синтаксическом) уровне были выбраны следующие индикаторы:

*Средняя длина слов (mean word length)* в символах и в слогах [Ilisei et al., 2010: 503-511; Volansky et al., 2015: 98-118; Kunilovskaya, Kutuzov, 2017: 69-94; Pylypenko et al., 2021: 8596-8611; François, Lefer, 2022: 50-70].

*Средняя длина предложений (mean sentence length)* в символах и в токенах [Baker, 1996: 181-183; Laviosa- Braithwäite, 1997: 531-540; Laviosa, 1998: 557-570; Corpas Pastor et al., 2008: 75-81; Ilisei et al., 2010: 503-511; Volansky et al., 2015: 98-118; Kunilovskaya, Kutuzov, 2015: 33-40; Kunilovskaya, Kutuzov, 2017: 69-94; Kunilovskaya, Lapshinova-Koltunski, 2019: 47-56; Pylypenko et al., 2021: 8596-8611; François, Lefer, 2022: 50-70]. При подсчете длины в токенах токенами считаются также знаки пунктуации.

*Средняя длина клауз (mean clause length)* в токенах [Liu, Afzaal, 2021: 1-20; Wang et al., 2023: 135-151]. При подсчете этого и следующего индикатора под клаузой понимается любая группа, в том числе непредикативная, вершиной которой является глагол, а при отсутствии полнозначного глагола – связка или грамматический элемент, выполняющий роль связки [Тестелец, 2001: 256]. Иными словами, учитываются как финитные, так и нефинитные клаузы. Для синтаксической разметки и определения клауз в тексте в настоящей работе используется библиотека для обработки естественного русского языка *Natasha*<sup>5</sup>.

*Среднее число клауз на предложение (average number of clauses per sentence)*, являющееся показателем синтаксической сложности текста [Kunilovskaya, Lapshinova-Koltunski, 2019: 47-56; Kunilovskaya, Corpas Pastor, 2021: 133-180; Kunilovskaya et al., 2021: 101-112]. Показатель рассчитывается как отношение общего числа клауз в тексте к общему числу предложений.

*Средняя доля простых предложений (simple sentences ratio)* относительно общего числа предложений, представляющая собой дополнительный показатель синтаксической сложности текста [Corpas Pastor et al., 2008: 75-81; Ilisei et al., 2010: 503-511; Kunilovskaya, Lapshinova-Koltunski, 2019: 47-56; Kunilovskaya et al., 2021: 101-112; Kunilovskaya, Corpas Pastor, 2021: 133-180; François, Lefer, 2022: 50-70].

---

<sup>5</sup> Библиотека для обработки естественного русского языка *Natasha*: <https://pypi.org/project/natasha/>

*Индекс удобочитаемости Флэша (Flesch reading ease score)*, отражающий уровень сложности текста с точки зрения его восприятия (чем ниже показатель, тем более сложным считается текст) [Corpas Pastor et al., 2008: 77-78]. В целях расчета данного показателя была взята формула для русского языка [Оборнева, 2006]:

$$K = 206,836 - 60,1 \times W - 1,3 \times S, \quad (7),$$

где  $K$  – показатель сложности текста,

$W$  – средняя длина слов (в слогах),

$S$  – средняя длина предложений (в словах).

*Нормализованные частоты различных знаков препинания (.,!?:;...-( )" '«»'“”)*, рассчитываемые как абсолютная частота каждого знака препинания относительно всех токенов в тексте [Volansky et al., 2015: 98-118; Rodríguez-Castro, 2011: 40-61]. Анализ пунктуации позволяет оценить синтаксическую структуру текста, степень её сложности.

Приведенная совокупность формальных индикаторов обеспечивает комплексный анализ характеристики *упрощение* в переводных текстах. Далее обратимся к описанию следующей потенциально универсальной характеристики переводов – *нормализации*.

## 2.2. Характеристика нормализация и её индикаторы

Под термином *нормализация* (а также *стандартизация* или *консерватизм*) принято понимать тенденцию переводных текстов быть лексически и грамматически более стандартизированными и типизированными по сравнению с непереводаемыми произведениями на языке перевода. Считается, что в процессе перевода переводчики часто отдают предпочтение общепринятой «грамматичности», следуя типичным шаблонам ПЯ и даже преувеличивая их использование [Baker, 1993: 224; Baker 1996: 176, 183-184; Kenny, 2001: 65].

На грамматическом уровне *нормализация* может выражаться в более частом использовании типичных для языка перевода немаркированных грамматических структур и коллокаций [Baker 1996: 183-184], а также пунктуационных паттернов [Vanderauwera, 1985: 93; May, 1997: 1-20].

Помимо грамматического уровня, также принято говорить и о стандартизации перевода на лексическом уровне. Во-первых, считается, что стандартизация на данном уровне проявляется в переводе через меньшую повторяемость знаменательной лексики: переводчики склонны избегать повторения одних и тех же лексических единиц из-за устоявшегося представления о том, что разнообразие словарного запаса является отражением культурного уровня автора (и переводчика в случае с переводом) и что чрезмерная лексическая повторяемость может восприниматься читателем как показатель «бедности» языка [Ben-Ari, 1998: 68-78; Volansky et al., 2015: 98-118]. По этой причине повторяющиеся слова исходного текста нередко опускаются или же заменяются синонимичными вариантами в переводе.

Другим проявлением *нормализации* на лексическом уровне служит то, что стилистически окрашенные лексические единицы (единицы, относящиеся к поэтическим, разговорным, сленговым, жаргонным, диалектным словам и словосочетаниям, а также к терминам) в процессе перевода нередко заменяются на более общеупотребительные и нейтральные единицы ПЯ. В контексте настоящего исследования, посвященного переводам именно научных текстов, примечателен пример, приведенный в работе Л. Эверос [Øverås, 1996: 98], когда в переводе с английского на норвежский язык биологический термин *hearing organs* (*органы слуха*) был заменён на общеупотребительное слово *ørene* (*уши*), а зоологический термин *abdomen* (*брюшко*) – на общеупотребительное слово *magen* (*живот*). Однако следует отметить, что исследование Л. Эверос [Øverås, 1996] посвящено переводам художественных произведений, в которых такие замены могут быть в меньшей степени критичны, в то время как в текстах научного стиля подобная лексическая стандартизация (использование общеупотребительной лексики вместо специализированных терминов) может приводить к снижению точности передачи научных данных, к искажению смысла исходного текста, что необходимо учитывать в переводе.

С *нормализацией* тесно связан *закон растущей стандартизации* (*the law of growing standardization*), сформулированный Г. Тури [Toury, 2012: 303-310].

Данный закон гласит, что в переводе уникальные единицы и структуры исходного текста (*текстемы*) имеют тенденцию преобразовываться в единицы и структуры, более типичные для культурного и языкового репертуара принимающей культуры (*реперторемы*). Под репертуаром в данном случае понимается совокупность языковых и культурных единиц и структур, закреплённых в сознании носителей языка и доступных для использования в различных типизированных коммуникативных ситуациях. Единицы репертуара (*реперторемы*) представляют собой те элементы языка и культуры, которые воспринимаются как «обычные», соответствующие норме. В свою очередь, текстема – это единица, функционирующая в каком-то конкретном тексте, включённая в уникальную сеть внутритекстовых отношений.

В процессе перевода связи, в которые вступают текстемы исходного текста, нередко модифицируются, а иногда и вовсе утрачиваются из-за того, что переводчики заменяют текстемы исходного текста на более типичные и стандартизированные единицы языка перевода (*реперторемы*), не всегда способные обеспечить ту же самую связь или обеспечить её в той же мере, что и текстемы текста оригинала. Причина этого процесса кроется в том, что уникальные, ситуативно обусловленные связи исходного текста, созданного в иных культурных и языковых реалиях, труднее реконструировать в тексте перевода, из-за чего выбор переводчиков часто падает на более конвенциональные единицы и структуры ПЯ. Для примера рассмотрим следующее предложение научного текста:

*«Given the lack of empirical data, this assumption stands on shaky ground...»*  
(пример мой – О. У.).

Данное предложение может быть переведено следующим образом:

*«Учитывая отсутствие эмпирических данных, можно сказать, что данное предположение сомнительно...»* (перевод мой – О. У.).

В приведенном варианте перевода текстема исходного текста *stand on shaky ground*, являющаяся идиомой, была полностью утрачена и заменена на вариант, состоящий из одного-единственного слова *сомнительно*, к тому же, более нейтральный и лишенный какой-либо образности. Тем не менее данную текстему

можно было бы попытаться сохранить, используя её русскоязычный идиоматический аналог:

*«Учитывая отсутствие эмпирических данных, можно сказать, что данное предположение **держится на шатких основаниях...**»* (перевод мой – О. У.).

Следует также отметить, что *нормализация (стандартизация)* текста перевода, вероятно, может быть тесно связана с рассмотренной в предыдущем разделе характеристикой *упрощение*, так как в процессе замещения уникальных элементов и структур исходного текста (*текстем*) более типичными и общеупотребительными элементами и структурами языка перевода (*реперторемами*) текст перевода может становиться лексически и структурно более простым.

Принято считать, что степень проявления стандартизации в тексте перевода во многом зависит от статуса перевода в принимающей культуре [Toury 2012: 306-307]: чем более периферийным является его статус, тем сильнее проявляется стремление переводчика следовать нормам и репертуару языка перевода. Верно и обратное: при более высоком статусе перевода (в особенности это касается перевода художественных произведений) переводчики, обладая большей степенью свободы, могут отклоняться от доминирующих в языке перевода норм в пользу сохранения оригинальных текстовых связей. Если говорить о научном тексте, а в особенности о таком научном жанре, как научная статья, то нельзя не отметить, что хотя в данном случае и допускается некоторая степень варьирования используемых средств выражения, по большей части данный жанр характеризуется высокой степенью стандартизации, частым использованием конвенциональных лексических и грамматических средств выражения [Беляева и др., 2021a: 4-5; Беляева и др., 2021b: 23-27]. Данная особенность научных текстов делает отклонение от привычных форм выражения в переводе крайне нежелательным: в таких условиях переводчики, осознавая всю строгость требований, предъявляемых к научному стилю изложения, могут в еще большей степени стремиться адаптировать текст перевода к нормам ПЯ, тем самым ещё более усиливая степень

его стандартизации, что, например, подтверждается исследованиями переводов научных текстов с разных языков на финский язык [Mauranen, 2000: 119-141].

Перейдем к рассмотрению формальных индикаторов *нормализации*, взятых за основу анализа данной характеристики в рамках настоящего исследования:

*Повторяемость знаменательной лексики (lexical repetition)*, рассчитываемая как отношение числа знаменательных слов (а именно глаголов, имен существительных, прилагательных и наречий), встречающихся в тексте более 1 раза (т. е. знаменательных слов с частотой  $\geq 2$ ), к общему количеству токенов в тексте [Volansky et al., 2015: 98-118]. Формула показателя:

$$LR = \frac{\sum_{i=1}^m 1_{\{f(w_i) \geq 2\}}}{N}, \quad (8),$$

где  $w_i$  –  $i$ -я уникальная знаменательная лексема текста,

$m$  – общее число уникальных знаменательных слов в тексте,

$f(w_i)$  – частота уникального знаменательного слова  $w_i$ ,

$N$  – общее количество токенов в тексте;

$1_{\{f(w_i) \geq 2\}}$  – индикаторная функция, принимающая значение 1, если слово  $w_i$  встречается не менее двух раз в тексте, и значение 0, если слово  $w_i$  встречается только один раз.

*Показатель поточечной взаимной информации слов (Pointwise Mutual Information)*, отражающий силу ассоциативной связи между двумя соседствующими словами, то есть то, насколько чаще два слова встречаются совместно, чем можно было бы ожидать, если бы их появления были независимы друг от друга [Church, Hanks, 1990: 22-29; Manning, Schütze, 1999: 178-183; Oakes, 2012: 139-140]. В корпусных исследованиях данный показатель широко используется для выявления коллокаций [Захаров, Хохлова, 2010: 137-143], причем под коллокацией в данном случае понимается не только устойчивое фразеологическое сочетание, но и любая биграмма (последовательность из двух слов), демонстрирующая высокую степень ассоциативной связанности [Baroni, Bernardini, 2003: 5; Volansky et al., 2015: 98-118]. Рассчитываются две версии показателя [Oakes, 2012: 139-140].

Формула *обычной версии PMI* :

$$PMI(x, y) = \log_2 \left( \frac{P(x, y)}{P(x) \times P(y)} \right) = \log_2 \left( \frac{\frac{f(x, y)}{N \times v}}{\frac{f(x) \times f(y)}{N^2}} \right), \quad (9)$$

где  $N$  – общее количество токенов в подкорпусе;

$P(x, y)$  – вероятность совместного появления пары лемм слов  $x$  и  $y$ ;

$P(x)$  и  $P(y)$  – вероятности появления  $x$  и  $y$  в тексте;

$f(x, y)$  – частота совместного появления пары лемм слов  $x$  и  $y$ ;

$f(x)$  и  $f(y)$  – отдельные частоты появления лемм слов  $x$  и  $y$ ;

$v$  – среднее число реально доступных при анализе контекстных позиций на один токен при выбранной конфигурации окна (т. е. количество соседних слов, учтённых для одного токена). В случае окна, равного 1, поправка на  $v$  не применяется.

Формула *модифицированной версии PMI*:

$$Modified\ PMI(x, y) = f(x, y) \times PMI(x, y), \quad (10)$$

где  $f(x, y)$  – частота совместного появления пары лемм слов  $x$  и  $y$ ;

$PMI(x, y)$  – значение, полученное с помощью предыдущей формулы.

В настоящей работе при расчёте показателей поточечной взаимной информации ( $PMI$  и  $Modified\ PMI$ ) использовалось два варианта размера окна совместной встречаемости слов ( $w$ ):  $w=1$  и  $w=5$ . Под окном в данном случае понимается число соседних позиций, в которых фиксируется совместное появление пары слов  $x$  и  $y$ . При  $w=1$  учитываются только соседние пары слов, то есть биграммы в их классическом понимании. Такой вариант позволяет выявлять устойчивые сочетания слов, непосредственно соседствующих друг с другом в тексте. При  $w=5$  учитываются слова на дистанции пяти позиций справа от токена, что даёт возможность фиксировать более свободные связи между словами, которые разделены в тексте другими лексемами. Для примера рассмотрим два предложения:

(1) «Во многом массовая культура формирует также и общественное мнение» (пример мой – О. У.);

(2) «Гендерные стереотипы по-прежнему оказывают значительное влияние на распределение профессиональных ролей» (пример мой – О. У.).

При окне  $w=1$  коллокациями в этих предложениях будут являться биграммы *во многом, массовая культура, общественное мнение, гендерный стереотип, значительное влияние, профессиональная роль*. При окне  $w=5$  для тех же предложений данный подход позволит выявить коллокации *формировать мнение, оказывать влияние, распределение ролей*.

Поскольку в текстах фактическая длина анализируемого окна всегда ограничена границами предложений, в формулу *PMI* вводится поправка  $v$ . Эта величина отражает среднее число слов-соседей, которое реально оказалось доступно для каждого токена при обходе подкорпуса. То есть для каждого слова подсчитывается, сколько соседей удалось учесть в пределах заданного окна, а полученные значения усредняются по всему подкорпусу. При  $w=1$  формула сводится к классической версии *PMI* [Volansky et al., 2015: 98-118], в которой поправка  $v$  не применяется.

Первая из приведенных формул (*PMI*) отражает силу ассоциативной связи между словами, однако имеет известный недостаток: редкие случайные сочетания слов могут давать искусственно завышенные значения *PMI* [Manning, Schütze, 1999: 178-183; 139-140; Oakes, 2012: 139-140]. Вторая формула (*Modified PMI*) вводит поправку на частоту совместного появления лемм слов  $x$  и  $y$ , благодаря чему данный показатель менее чувствителен к случайным единичным сочетаниям слов и позволяет выявлять устойчивые коллокации, которые действительно регулярно встречаются в подкорпусе [Oakes, 2012: 139-140].

В частности интерес в рамках анализа характеристики *нормализация* представляют следующие показатели *PMI*:

*Среднее значение PMI / Modified PMI биграмм слов* по всему подкорпусу [Volansky et al., 2015: 98-118]. Данный показатель отражает степень ассоциативной связанности слов в подкорпусе: чем выше его значение, тем сильнее выражены связи между словами. Например, такие биграммы, как *социальное неравенство, девиантное поведение, клиническое исследование, сердечная недостаточность* будут обладать бóльшим значением показателя ассоциативной связанности в сравнении с биграммами слов типа *организация факторов, исследование давления*.

При более низком среднем значении показателя *PMI / Modified PMI* для подкорпуса *A* по сравнению с подкорпусом *B* можно говорить о том, что слова в подкорпусе *A* в среднем образуют менее устойчивые и менее типичные для языка коллокации.

*Доля биграмм слов, значение PMI / Modified PMI которых выше 0* [Volansky et al., 2015: 98-118]. Данный показатель отражает распространённость ассоциативно связанных пар в подкорпусе, то есть долю сочетаний слов с  $PMI > 0$ . Более высокая доля биграмм слов с  $PMI > 0$  в подкорпусе *A* по сравнению с подкорпусом *B* свидетельствует о большем числе коллокаций в подкорпусе *A*.

Наряду с *нормализацией* к потенциально универсальным характеристикам переводов относится также и *экспликация*. Рассмотрим данную характеристику подробнее в следующем разделе.

### 2.3. Характеристика экспликация и её индикаторы

Под *экспликацией* в контексте феномена переводческого языка принято понимать тенденцию к повышению объяснительного потенциала переводного текста в сравнении с текстом исходным [Pápai, 2004: 143-164; Puurtinen, 2004: 165-176; Volansky et al., 2015: 98-118].

Впервые термин *экспликация* был использован лингвистами Ж.-П. Вине и Ж. Дарбельне [Vinay, Darbelnet, 1995: 80, 342-343] в 1958 году для обозначения специального приема перевода, состоящего в стремлении переводчика сделать явным в переводе то, что в оригинальном тексте выражено имплицитно, то есть подразумевается, но может быть выведено на основании контекста и культурных знаний о ситуации, описываемой в оригинале.

На *экспликацию* как на общую особенность переводных текстов (а не как на прием перевода) указывал Г. Тури [Toury, 1980: 60], охарактеризовав ее как устойчивую тенденцию переводов к эксплицитному выражению имплицитно заложенной в исходном тексте информации. Причем данная тенденция, по мнению Г. Тури, проявляется в переводах безотносительно личности переводчика, задействованной языковой пары, жанровых особенностей переводимого текста, временного периода, в который выполняется перевод, и иных факторов.

Рассматривать *экспликацию* как гипотетически универсальную характеристику переводов предложила Ш. Блум-Кулька, опирающаяся на исследование трансформаций когезивных и когерентных отношений в процессе перевода [Blum-Kulka, 2000: 298-313]. Ш. Блум-Кулька выделила два её типа:

1) Первый тип проистекает из объективных различий между системами ИЯ и ПЯ. Так, при переводе имен существительных с русского языка на английский переводчику неминуемо приходится вводить артикли (a / the) или другие определительные элементы (притяжательные, указательные местоимения и т. д.), поскольку в грамматике английского языка существует формально выраженная категория определенности / неопределенности, отсутствующая в русском языке. В то время как при переводе имен существительных с английского на русский необходимо уточнять род имени существительного – категорию, отсутствующую в грамматике английского языка. Иными словами, такая *экспликация* обусловлена объективной необходимостью восстанавливать то, что отсутствует в исходном языке, но обязательно для выражения в языке перевода.

2) Второй тип *экспликации* проистекает из стремления переводчика – сознательного или бессознательного – сделать текст перевода более понятным для читателя, раскрывая или уточняя имплицитно присутствующие в оригинале смысловые связи. Именно этот тип *экспликации*, по мнению Ш. Блум-Кульки, следует рассматривать как характерную черту переводческого языка. Такая *экспликация*, как считает исследователь, хотя и присуща в большей мере работам начинающих переводчиков или переводчиков-любителей, нередко наблюдается и в работах переводчиков-профессионалов [Blum-Kulka, 2000: 300-302].

Механизмом реализации *экспликации* второго типа является повышение *когезии* переводного текста. Под *когезией* принято понимать особый тип внутритекстовой связи, обеспечивающей логическую последовательность (континуум) и связность излагаемых в тексте фактов [Гальперин, 2007: 73-74]. В результате частого использования в переводе различных когезионных средств происходит так называемый сдвиг на уровне связности (*cohesion shift*) [Blum-Kulka, 2000: 300-302] – текст перевода становится информационно более эксплицитным

(иногда избыточно эксплицитным) по сравнению с текстом оригинала и с непереводаемыми текстами на языке перевода.

Средства обеспечения когезии традиционно подразделяются на лексические и грамматические [Halliday, Hasan, 1976]. Лексическая когезия обеспечивается выбором лексических единиц, семантически или ассоциативно связанных с другими лексемами того же текста (лексический повтор, сино- и антонимия, гиперо- и гипонимия, холо- и меронимия, а также коллокация) [Halliday, Matthiessen, 2014: 642-650].

Грамматическая когезия включает два основных механизма: *референцию* (в значении М. А. К. Халлидея [Halliday, Hasan, 1976]), т. е. использование языковых единиц (чаще всего местоимений), отсылающих к упомянутым в тексте или внетекстовым референтам (лицам, предметам, явлениям, месту, времени или ситуации) [Halliday, Hasan, 1976: 226-273; Bloor T., Bloor M., 2013: 95-101], а также *конъюнкцию*, т. е. логическую связь между клаузами внутри сложных предложений, между отдельными предложениями и между более крупными фрагментами текста, реализуемую посредством использования союзов (союзных слов), а также ряда наречий, предложных конструкций и целых словосочетаний, которые, хотя и не относятся к грамматической категории союзов, выполняют схожую с ними функцию – связывают фрагменты текста и выражают разного рода отношения между ними [Halliday, Hasan, 1976: 226-273; Bloor T., Bloor M., 2013: 95-101].

Конъюнктивные средства когезии по существу представляют собой тот же класс языковых единиц, который, в иной терминологической традиции, принято называть метатекстовыми вводными компонентами [Вежбицкая, 1978: 402-421] или же дискурсивными маркерами (*discourse markers*) [Викторова 2013: 34-46; Викторова, 2014: 18-22]: и те и другие маркируют логико-смысловые отношения между фрагментами текста и обеспечивают его структурную и смысловую целостность. Различие между ними носит не предметный, а терминологический характер и обусловлено принадлежностью к разным исследовательским парадигмам. По причине того, что в корпусных исследованиях переводческого

языка для анализа *экспликации* широко используется понятийный аппарат дискурсивных маркеров [Párai, 2004: 143-164; Kunilovskaya, Kutuzov, 2017: 69-94; Kunilovskaya, Lapshinova-Koltunski, 2019: 47-56; Kunilovskaya et al., 2021: 101-112; Kunilovskaya, Corpas Pastor, 2021: 133-180; Кантышева, 2021: 140], в настоящей работе для операционализации конъюнктивной когезии будем пользоваться соответствующей терминологией.

Дискурсивные маркеры подразделяются на две крупных категории, соотносимые с разграничением локальной и глобальной связности текста [Дейк, Кинч, 1988]: маркеры, обеспечивающие связь на *локальном уровне* (внутри предложений и между ними), и маркеры, обеспечивающие связь на *глобальном уровне* (между более крупными фрагментами текста и текста как единого целого). Поскольку настоящее исследование проведено на материале переводов научных статей, для анализа конъюнктивной когезии была принята классификация дискурсивных маркеров научного текста [Когут 2014: 102; Болсуновская и др., 2015: 117-123; Когут, 2016: 157-163]. К маркерам локальной связности в научном тексте относят маркеры:

- введения новой или дополнительной информации (*кроме того, также, помимо этого, более того* и т. д.);
- пояснения или перефразирования упомянутого ранее (*другими словами, иначе говоря, то есть, а именно, точнее говоря* и т. д.);
- противопоставления или отхода от основной линии изложения (*однако, напротив, тем не менее, в то же время, между тем* и т. д.);
- введения примеров (например, *к примеру, для иллюстрации* и т. д.).

К маркерам *глобальной связности* относят маркеры [Когут, 2016: 159-160]:

- введения читателя в тему статьи / исследования (*в данной статье речь пойдет о, в данной статье мы рассмотрим, в этой статье мы обратимся к, настоящая работа посвящена* и т. д.);
- обеспечения порядка следования информации / аргументации в тексте (*во-первых, прежде всего, далее, наконец, для начала рассмотрим, теперь обратимся к* и т. д.);

- описания расположения материала в тексте (*как говорилось ранее / выше, как было показано ранее / выше, как уже было отмечено* и т. д.);
- выводов и заключения (*следовательно, поэтому, в результате, итак, таким образом* и т. д.).

Обе эти группы дискурсивных маркеров играют ведущую роль в организации научного текста и научного дискурса в целом, поскольку позволяют автору (а в случае с переводом – переводчику) организовать своего рода навигацию по тексту: обозначить переходы от одной линии рассуждения к другой, пояснить, уточнить, конкретизировать высказанную ранее мысль, привести примеры к сказанному ранее, сделать выводы относительно какой-то конкретной линии рассуждения, подытожить результаты всего исследования в целом и т. д.

Перейдем к рассмотрению формальных индикаторов, отобранных для анализа характеристики *экспликация* в рамках настоящей работы. Так, в целях анализа связности текстов на *лексическом уровне* было принято решение использовать два взаимодополняющих друг друга индикатора:

*Среднее косинусное сходство пар знаменательных слов*, отражающее степень семантической близости лексики в тексте: чем выше значение показателя, тем более семантически схожа и взаимосвязана лексика текста. Вычисление показателя косинусного сходства слов является эффективным способом измерения семантической близости лексики в векторных моделях языка [Kutuzov, Kuzmenko, 2017: 155-161]. Векторные представления слов (*word embeddings*) – это способ представления лексических единиц в виде многомерных числовых векторов, где каждое слово кодируется последовательностью действительных чисел. Векторные представления позволяют отразить семантические отношения между словами через их геометрические отношения в многомерном пространстве: слова со схожими значениями или употребляющиеся в похожих контекстах располагаются в векторном пространстве близко друг к другу. Степень семантической близости слов формализуется через косинусное сходство, вычисляемое как косинус угла между векторами слов [Kutuzov, Kuzmenko, 2017: 155-161]. Чем меньше этот угол, тем выше значение метрики. Максимальное значение равняется 1, что

соответствует полному совпадению направлений (что, например, наблюдается при повторе в тексте одного и того же слова).

Для вычисления векторов отдельных слов и дальнейшего вычисления косинусного сходства между парами слов в настоящем исследовании используется предварительно обученная модель распределенных представлений слов *Taiga*, предоставляемая в рамках проекта RusVectōrēs<sup>6</sup> [Kutuzov, Kuzmenko, 2017: 155-161]. Косинусное сходство между парами слов рассчитывается через веб-инструмент *similarity* того же проекта RusVectōrēs. Приведем несколько примеров его использования: для слов *препарат* и *лекарство* (синонимы) в тексте косинусное сходство составит 0.788, для *увеличивать* и *снижать* (антонимы) – 0.767, для *фрукт* и *ананас* (гипероним и гипоним) – 0.721, для *сердце* и *печень* (согипонимы) – 0.425, для *прилагать* и *усилие* (коллокация) – 0.592, тогда как для пары *прикладывать* и *самолёт* оно будет близко к нулю – 0.002.

*Показатель среднего косинусного сходства* для всех пар слов в тексте вычисляется следующим образом:

(1) На основе списка лемм знаменательных слов текста (имен существительных, глаголов, имен прилагательных и наречий) формируются все возможные пары слов. При этом из анализа исключаются служебные части речи (предлоги, союзы, частицы), местоимения и числительные;

(2) Для каждого слова в паре извлекается его векторное представление из модели *Taiga*;

(3) С помощью инструмента *similarity* [Kutuzov, Kuzmenko, 2017: 155-161] для каждой пары вычисляется косинусное сходство слов;

(4) Среднее косинусное сходство всех пар слов рассчитывается как взвешенное среднее всех значений косинусных сходств, где весом выступает частота совместного появления в тексте пары слов.

В случае, если в переводном подкорпусе по сравнению с непереводным связи внутри текстов на уровне лексики более явны и эксплицитны, следует ожидать увеличение доли семантически близких единиц, и, как следствие, более высокое

---

<sup>6</sup> Сайт проекта RusVectōrēs: <https://rusvectors.org/ru/>

значение среднего косинусного сходства, поскольку рост данного показателя будет свидетельствовать о том, что в переводах присутствует семантически более связанная лексика, в то время как более низкие значения будут указывать на бóльшую семантическую разнородность лексики.

*Среднее косинусное сходство пар соседних предложений*, отражающее степень семантической близости и плавность переходов между ними. Показатель вычисляется следующим образом:

(1) Каждое предложение кодируется вектором фиксированной размерности, рассчитываемым посредством усреднения векторных представлений всех знаменательных слов предложения (имен существительных, глаголов, имен прилагательных и наречий), то есть как среднее арифметическое векторов этих слов. Как и в предыдущем случае, из подсчета исключаются служебные части речи, местоимения и числительные;

(2) Вычисляется косинусное сходство между векторами каждой пары соседствующих предложений;

(3) Итоговый показатель определяется как среднее арифметическое всех полученных значений косинусных сходств между последовательными предложениями.

Если два соседних предложения содержат семантически близкую лексику, угол между их векторами будет небольшим, а косинусное сходство – высоким. В случае более высокого значения среднего косинусного сходства соседних предложений можно говорить о более сильной семантической связности между предложениями в тексте и о большей плавности переходов между ними. Для наглядности приведём два примера:

1) *«Развитие статистических методов многоуровневого моделирования позволяет сегодня проверять предположения о влиянии на здоровье не только индивидуального социального капитала и его общественных ресурсов, но и учитывать взаимодействие тех и других факторов. Учёт этого взаимодействия даёт возможность выявлять социальные контексты, в которых для здоровья особенно важна социальная поддержка, где оно в большей степени зависит от*

*развития гражданской активности, а также общественные условия, благоприятные для проявления зависимости самочувствия людей от их доверительного отношения к другим»* [Русинова Н. Л., Сафронов В. В. Значение социального капитала для здоровья в странах Европы // Журнал социологии и социальной антропологии. –2014. –Том 17. – № 3. – С. 112-33] – значение косинусного сходства данной пары предложений составило 0.882. Такой результат говорит о том, что оба предложения задействуют семантически связанную лексику и образуют плавный тематический переход.

2) «*Тектонические плиты перемещаются со средней скоростью около 1-10 см в год. Романтизм в литературе противопоставлял индивидуальное переживание рационализму Просвещения»* (пример мой – О. У.) – значение показателя этой пары предложений составило 0.282. Данный результат закономерен: предложения принадлежат разным предметным областям (геология и история литературы) и содержат малосвязанные лексические единицы.

Кроме представленных выше индикаторов, для анализа *лексической когезии* также могут быть использованы уже приведенные в предыдущем разделе индикаторы *лексической повторяемости (lexical repetition)* и *поточечной взаимной информации (PMI и Modified PMI)*. Лексический повтор представляет собой наиболее сильный тип лексической когезии [Halliday, Matthiessen, 2014: 644-645], поскольку повторное использование лексем создает явные семантические связи между фрагментами текста, формируя когезивные цепочки и обеспечивая его континуальность. Соответственно, повышение уровня лексической повторяемости свидетельствует об усилении референциальных связей между фрагментами текста, что делает текст более эксплицитным.

В то же время показатель *поточечной взаимной информации* отражает силу ассоциативной связанности между совместно встречающимися в тексте лексическими единицами в линейном отрезке речи, что непосредственно соотносится с коллокационной когезией [Halliday, Matthiessen, 2014: 648-650], также являющейся одним из типов лексической когезии. Повышение среднего значения *PMI* биграмм слов, а также увеличение доли биграмм с положительным

значением *PMI* в подкорпусе будет свидетельствовать об усилении коллокационной когезии, то есть о более частом использовании устойчивых словосочетаний, вносящих вклад в общую текстовую связность через создание более предсказуемых ассоциативных связей между элементами текста.

С целью анализа связности текстов на грамматическом уровне, а именно *референциальной* грамматической когезии [Halliday, Hasan, 1976] были выбраны следующие индикаторы:

*Нормализованная частота местоимений*, принадлежащих к таким разрядам, как личные, притяжательные, указательные и относительные местоимения. Показатель рассчитывается как абсолютная частота вхождений в текст местоимений того или иного разряда, деленная на общее число токенов в тексте.

*Показатель эксплицитного называния*, представляющий отношение общего числа личных и притяжательных местоимений к общему числу используемых в тексте именованных сущностей [Volansky et al., 2015: 98-118; François, Lefer, 2022: 50-70]. Данный индикатор позволяет количественно оценить, в какой степени в тексте перевода задействована стратегия явного указания на референт, о котором идет речь в тексте. С одной стороны, рост доли личных и притяжательных местоимений в тексте может быть показателем роста *референциальной когезии*. С другой стороны, её снижение может указывать на стратегию заменять местоимения исходного текста именами собственными в переводе, что также может быть интерпретировано как *экспликация*, направленная на устранение потенциальной неоднозначности.

С целью анализа *конъюнктивной* грамматической когезии было решено сосредоточиться на анализе упомянутых выше дискурсивных маркеров научного текста, а именно *маркеров организационного характера* [Когут, 2016: 157-163]. Анализировались следующие индикаторы:

1) *Нормализованные частоты маркеров локального уровня связности* (маркеры введения новой / дополнительной информации, пояснения, противопоставления и введения примеров);

2) *Нормализованные частоты маркеров глобального уровня связности* (маркеры введения в тему, порядка следования информации, расположения материала в тексте и выводов).

Нормализованные частоты дискурсивных маркеров подсчитывались путём деления абсолютной частоты маркеров определенной категории на общее число токенов в тексте. В случае, если маркер состоял из нескольких токенов (например, *в то же время, с одной стороны*), при подсчете показателя нормализованной частоты он считался за один токен.

По причине того, что исчерпывающего перечня дискурсивных маркеров научного текста, относящихся к упомянутым категориям, в научной литературе найти не удалось, в целях анализа такой перечень был сформирован вручную. Отбор маркеров производился путём анализа массива текстов научных статей, параллельно с этим был проведен обзор научной литературы по данной теме [Викторова 2013: 34-46; Викторова, 2014: 18-22; Болсуновская и др., 2015: 117-123; Когут 2014: 102; Когут, 2016: 157-163; Валеева, Руднева, 2016: 109-115].

Дополнительно в анализ были включены маркеры, выражающие причинно-следственные (*потому что, вследствие этого, причиной служит* и др.) и целевые отношения (*для того, чтобы; затем, чтобы; с расчетом на то, что* и др.) – хотя эти типы маркеров и не представлены в классификации дискурсивных маркеров научного текста [Когут, 2016: 157-163], они играют не менее важную роль в организации научного дискурса. Данные типы маркеров вошли в категорию дискурсивных маркеров локального уровня связности текста, а отбор единиц осуществлялся с опорой типологию средств конъюнктивной когезии М. А. К. Халлидея и Р. Хасан [Halliday, Hasan, 1976].

В итоговый список маркеров вошло 845 единиц, включая разные вариации и словоформы маркеров (типа *упомянутый / -ая / -ое / -ые ранее / раньше / выше*). Полный список единиц в работе не приводится ввиду ограничений объёма диссертации, однако для прозрачности методики некоторые примеры отобранных маркеров представлены в ПРИЛОЖЕНИИ А.

Помимо перечисленных выше индикаторов, относящихся к *грамматической* и *лексической* когезии, в рамках анализа *экспликация* также анализируются следующие индикаторы:

*Показатель частоты единичного именованного (single naming)*, подсчитываемый как отношение числа однословных именованных сущностей (т. е. состоящих из одного токена) ко всем именованным сущностям в тексте [Volansky et al., 2015: 98-118; Hu, Kübler, 2021:339-372]. Данный индикатор основывается на предположении о том, что при переводе предпочтение отдается полным версиям имен (например, *Мишель Фуко*), тогда как в непереводах чаще используется вариант из одного слова (просто *Фуко*).

*Средняя длина именованных сущностей (mean named entities length)* представляет собой среднее число токенов, составляющих именованные сущности [Volansky et al., 2015: 98-118; Hu, Kübler, 2021:339-372]. Данный индикатор основывается на том же предположении, что и предыдущий: ожидается, что в переводах чаще используются более длинные версии имен собственных.

Рассмотрев особенности реализации характеристики *экспликация* и используемые для её анализа индикаторы, перейдем к описанию особенностей и индикаторов еще одной важной характеристики переводческого языка – *интерференции*.

#### 2.4. Характеристика интерференция и её индикаторы

Изначально термин *интерференция* использовался исключительно в контексте овладения языком для обозначения отклонений от норм использования билингвом любого из известных ему языков [Weinreich, 1953:1-2]. При этом термин *билингв* в данном случае понимается в широком смысле – как человек, владеющий более чем одним языком, а языковая интерференция считается двунаправленной: подразумевается как отклонение от норм изучаемого языка под воздействием родного, так и отклонение от норм родного языка под влиянием изучаемого. Несложно в данном случае заметить некоторое пересечение термина *интерференция* с упоминавшимся в Главе 1 термином *interlanguage*, применяемым

для обозначения некоего промежуточного варианта языка, используемого обучающимися на этапе овладения иностранным языком.

Как уже упоминалось ранее, Г. Тури [Toury, 1979: 223-224] предложил использовать термин *interlanguage* в любой ситуации, в которой два языка вступают в непосредственный контакт друг с другом, порождая уникальные языковые формы, а в особенности – в ситуации перевода текста с одного языка на другой. В рамках этого подхода Г. Тури сформулировал так называемый *закон интерференции (the law of interference)*. Данный закон гласит, что некоторые особенности организации текста оригинала оказывают влияние на переводческие решения, как бы «навязывая» себя переводчику, что обуславливает перенос этих особенностей в текст перевода [Toury 2012: 310-315]. При этом принято разграничивать *положительную* и *отрицательную интерференцию*. Положительная *интерференция* выражается в увеличении частоты использования форм и структур, уже существующих в системе ПЯ и не нарушающих его норм и узуса, и не представляет собой никакой проблемы. Отрицательная *интерференция*, напротив, состоит в выборе форм и структур ПЯ, обнаруживающих формальное сходство с единицами исходного текста и на первый взгляд кажущихся подходящими переводческими эквивалентами, но при этом нарушающих языковую норму или узус ПЯ [Hisei, 2012: 35; Куниловская, 2015: 5].

Считается, что опытные переводчики менее подвержены влиянию особенностей текста оригинала, так как лучше владеют нормами ПЯ и с бóльшим успехом избегают необоснованного копирования форм и структур исходного текста в текст перевода, в то время как начинающие переводчики или переводчики-любители могут не замечать или не уметь корректно адаптировать особенности исходного текста в тексте перевода. Тем не менее считается, что так или иначе *интерференция* проявляется в работах как опытных, так начинающих переводчиков [Tirkkonen-Condit, 2005: 405-414]. При этом существует гипотеза [Mauranen, 2004: 65-82] о том, что *интерференция* в переводе может быть обусловлена не только непосредственным влиянием исходного текста на текст перевода, но и самим фактом использования в процессе перевода двух языковых

систем, поскольку в момент выполнения перевода чтение переводчиком исходного текста как бы активизирует в его сознании систему исходного языка, которая, в свою очередь, опосредованно влияет на переводческие решения и, как следствие, на сам результат перевода даже в том случае, когда в читаемом отрезке исходного текста не содержится конкретных элементов и структур, которые могли бы быть напрямую перенесены в текст перевода.

В переводе *интерференция* выражается в том, что под воздействием структурной организации исходного текста переводчик выбирает схожие грамматические формы ПЯ, копирует синтаксические, пунктуационные паттерны, следует тому же, что и в тексте оригинала, порядку слов в предложении. Проблема такой *интерференции* состоит в том, что выбранные под влиянием исходного текста грамматические формы и структуры могут нарушать нормы и узус ПЯ в одном или нескольких аспектах [Toury, 1979: 226]. Во-первых, они могут нарушать грамматические нормы ПЯ. Во-вторых, будучи грамматически корректными, они нередко могут нарушать узус, либо употребляясь с нетипичной для данного языка частотностью, либо используясь в нехарактерных для них контекстах. В-третьих, такая *интерференция* способна приводить к нарушению жанрово-стилистических норм ПЯ, когда выбранные варианты перевода, хотя и допустимы в языковой системе в целом и грамматически корректны, оказываются неуместными в рамках конкретного функционального стиля или жанра.

С целью анализа характеристики *интерференция* в настоящей работе был выбран следующий индикатор:

*Нормализованные частоты частеречных последовательностей*, или *частеречных n-грамм (part of speech n-grams)*. Данный индикатор является одним из наиболее информативных при анализе *интерференции*, поскольку позволяет фиксировать различия в распределении синтаксических паттернов между переводными и непереводными текстами [Baroni, Bernardini, 2006: 259-274; Corpas Pastor et al., 2008: 78; Kurokawa et al., 2009: 81-88; Bernardini, Baroni, 2005: 1-12; Volansky et al., 2015: 98-118; Pylypenko et al., 2021: 8598]. В рамках исследования анализируются частоты последовательностей из одного (униграммы), двух

(биграмм) или трех (триграмм) токенов, замененных соответствующими им частеречными тегами. Нормализованные частоты рассчитываются как отношение абсолютной частоты конкретной n-граммы к совокупному числу всех n-грамм в тексте.

Для частеречной разметки текста в рамках исследования используются библиотеки для обработки естественного языка – *Natasha*<sup>7</sup> и *pymorphy2*<sup>8</sup>. Полный перечень используемых в анализе тегов представлен в ПРИЛОЖЕНИИ Б. Частеречные теги задействованных в анализе библиотек были дополнены некоторыми специальными тегами, позволяющими проводить более детализированный анализ грамматических явлений, потенциально отражающих *интерференцию* при переводе с английского языка на русский. Рассмотрим подробнее, каким именно образом эти теги расширяют спектр анализа:

1. Анализ *придаточных предложений*, вводимых не только подчинительными союзами (*SCONJ*), но также и союзными словами, а именно союзными местоименными словами (*PRN\_CONJ\_W* типа *который, какой, чей*), а также союзными наречными словами (*ADV\_CONJ\_W* типа *где, когда, куда*);

2. Анализ *степеней сравнения имен прилагательных и наречий*, включающий не только синтетические формы сравнительной (*COMP*) и превосходной (*SPRL*) степени, но также и аналитические формы, представленные комбинациями тегов *COMP\_MOD + ADJF / ADVB* (*более точный / точно*) и *SPRL\_MOD + ADJF / ADVB* (*самый / наиболее точный; наиболее точно*);

3. Анализ *местоимений различных разрядов* и комбинаций с ними, включающий специальные теги для:

- личных местоимений в функции подлежащего (*PER\_PRN* типа *он, они, мы*) и дополнения (*OBJ\_PRN* типа *ему, им, нам*);
- притяжательных местоимений (*POSSV\_PRN* типа *его, их, наши*);
- указательных местоимений (*DMSTR\_PRN* типа *этот, тот*);
- возвратного местоимения (*RFLX\_PRN*: *себя*);

<sup>7</sup> Библиотека для обработки естественного русского языка *Natasha*: <https://pypi.org/project/natasha/>

<sup>8</sup> Морфологический анализатор для русского языка *pymorphy2*: <https://pymorphy2.readthedocs.io/en/stable/>

- определительных местоимений, выраженных определительными местоименными прилагательными / существительными (*DF\_ADJ\_N\_PRN* типа *весь, другой, каждый*) и наречиями (*DF\_ADV\_PRN* типа *везде, отовсюду, всегда*);

- неопределенных местоимений, выраженных неопределенными местоименными существительными (*INDF\_NOUN\_PRN* типа *некто, кто-то*), прилагательными (*INDF\_ADJ\_PRN* типа *некоторый, какой-то*), наречиями (*INDF\_ADV\_PRN* типа *как-то, где-то, когда-то*) и числительными (*INDF\_NUM\_PRN* типа *несколько, сколько-то*);

- отрицательных местоимений, выраженных отрицательными местоименными существительными (*NEG\_N\_PRN* типа *никто, ничто*), прилагательными (*NEG\_ADJ\_PRN* типа *никакой, ничей*), наречиями (*NEG\_ADV\_PRN* типа *никогда, нигде, никак*) и числительным (*NEG\_NUM\_PRN*: *нисколько*);

4. Анализ *модальных конструкций*, выраженных модальными глаголами с инфинитивом (*MDL\_VRB + INFN* типа *может означать*) и модальными словами с инфинитивом (*MDL\_WRD + INFN* типа *можно / нужно / необходимо учитывать*);

5. Анализ *вводных конструкций*, выраженных комбинацией вводного слова и запятой (*PARENTH + COMMA* типа *конечно, ...; например, ...; действительно, ...*);

6. Анализ *сослагательного наклонения*, выраженного сослагательной частицей *бы* (*SBJ\_PRCL*);

7. Анализ *неличных форм глагола* и комбинаций с ними. Вводятся специальные теги для фиксации частеречных последовательностей с инфинитивом (*INFN*), деепричастием (*GRND*), а также полной и краткой формой причастия (*PRTF / PRTS*):

- тег *GRND* в комбинации с тегом *COMMA* позволяет фиксировать деепричастные обороты (*COMMA + GRND*: *...[, рассмотрев] особенности..., перейдем к ...*);

- тег *INFN* позволяет фиксировать следующие паттерны: составные глагольные сказуемые (*VERB + INFN*: *начали исследовать, продолжает*

*развиваться*) и целевые конструкции с инфинитивом (*SCONJ + INFN: чтобы выявить, чтобы установить*);

- теги *PRTF* и *PRTS* позволяют фиксировать следующие паттерны: причастные обороты (*COMMA + PRTF: результаты [, полученные] в ходе эксперимента*); страдательные конструкции, состоящие из вспомогательного глагола и краткого причастия (*AUX + PRTS: были проанализированы, будет рассмотрено*); полные причастия в функции определения в препозиции к определяемому слову (*PRTF + NOUN: полученные результаты*).

Итак, рассмотрев особенности характеристики *интерференция*, а также используемый для её анализа индикатор, в следующем разделе перейдем к индикаторам, не вошедшим ни в одну из перечисленных выше характеристик, но позволяющим фиксировать проявления феномена переводческого языка, связанные с реализацией норм русскоязычного научного стиля.

## **2.5. Индикаторы переводческого языка, отражающие особенности научного стиля русского языка**

Помимо индикаторов *упрощения, нормализации, экспликации и интерференции*, в исследование включена группа дополнительных показателей, опирающихся на стилеобразующие характеристики русскоязычного научного текста (см. раздел 1.4).

Рассмотрим подробнее каждую группу дополнительных индикаторов и их связь со стилеобразующими особенностями научного стиля русского языка. Первая группа связана с анализом дискурсивных маркеров научного текста, выражающих отношение автора к предмету высказывания и обеспечивающих взаимодействие автора с читателем [Когут, 2016: 157-163]. Выбор данных категорий маркеров основывается на двух важных характеристиках научного стиля: объективности и некатегоричности изложения, с одной стороны, и его диалогичности – с другой.

Как упоминалось в разделе 1.4, объективность научного текста выражается через взвешенность даваемых автором оценок в отношении степени изученности обсуждаемой темы, результатов исследования, позиций других исследователей и

своей собственной [Пумпянский, 1981: 26-27; Комиссаров, 1990: 110; Паршин 2000: 63-62; Плещенко и др., 1999: 36; Кожина и др., 2008: 289-311; Yakhontova, 2020: 1-4; Гуревич, 2024: 19-20]. Данное требование находит отражение в использовании особых языковых средств, позволяющих автору выразить степень уверенности в высказываемых им суждениях, выразить свою позицию корректно и сдержанно, в рамках принятых в научном сообществе конвенций. Одновременно с этим научному стилю свойственна диалогичность, то есть направленность речи на адресата с целью привлечь его внимание к важным фрагментам в тексте, вовлечь его в совместное с автором рассуждение и т. д. [Кожина и др., 2008: 289-311]. С учетом этих особенностей были выбраны следующие индикаторы:

*Нормализованные частоты маркеров авторского начала в тексте* [Когут, 2016: 157-163], используемых для выражения:

- мнения автора (*на наш взгляд, мы склонны полагать*);
- оценки предмета обсуждения / отношения к нему (*любопытно, удивительно, вызывает сомнение*);
- высокой (*несомненно, безусловно, бесспорно*), средней (*возможно, предположительно, как нам кажется*) и низкой (*маловероятно, сомнительно, спорно, едва ли*) степеней уверенности.

*Нормализованные частоты маркеров взаимодействия автора с читателем* [Когут, 2016: 157-163], используемых для:

- призыва читателя к самостоятельному (*см. (смотрите); ср. (сравните); обратите внимание*) или совместному с автором (*рассмотрим, обратимся к, допустим*) действию;
- акцентирования внимания читателя на важной или дополнительной информации (*следует отметить; примечательно, что; заметим, что*);
- апелляции к фоновым знаниям (*общеизвестно, как известно*).

Как и в случае с маркерами *организационного характера*, используемыми в анализе характеристики *экспликация*, исчерпывающего списка маркеров рассматриваемых категорий найти не удалось. В связи с этим такой список был составлен вручную. В итоговый список маркеров, используемых для выражения

авторского начала в тексте, вошло 465 единиц (включая вариации и словоформы). Список маркеров, используемых для взаимодействия автора с читателем, включил 471 единицу (также включая вариации и словоформы). Некоторые примеры маркеров данных категорий представлены в ПРИЛОЖЕНИИ В.

Последующие группы отобранных индикаторов связаны с морфологическими характеристиками научного стиля русского языка. Поскольку в процессе перевода морфологические особенности исходного текста могут влиять на соответствующие особенности текста перевода, было принято решение о включении в анализ ряда индикаторов, позволяющих выявить, в какой мере получающиеся в процессе перевода тексты соответствуют требованиям научного стиля русского языка по своим морфологическим параметрам. Грамматические категории анализируемых частей речи определялись с помощью библиотеки для обработки естественного русского языка *Natasha*<sup>9</sup>.

Вторая подгруппа включает индикаторы, связанные с функционированием имён существительных:

*Доля имён существительных* (относительно всех словарных токенов) отражает степень реализации номинативного характера научного текста: сниженные значения в переводных текстах могут свидетельствовать о недостаточной номинализации вследствие интерференции, поскольку субстантивация менее характерна для англоязычных научных текстов, чем для русскоязычных [Комиссаров, 1990: 132; Бреус, 2000: 34-39; Балыгина, Круковская, 2018: 3-10; Yakhontova, 2020: 1-4].

*Доля имен существительных среднего рода* (относительно всех существительных). Выбор индикатора основывается на положении о том, что в русскоязычных научных текстах существительные среднего рода употребляются достаточно часто в силу преобладающего у них отвлеченно-обобщенного значения [Кожина и др., 2008: 293].

*Доля существительных в форме единственного числа* (относительно всех существительных) отражает характерную для русского научного текста

---

<sup>9</sup> Библиотека для обработки естественного русского языка *Natasha*: <https://pypi.org/project/natasha/>

закономерность: преобладание единственного числа, обусловленное высокой долей отглагольных существительных и употреблением конкретных имён в обобщающем значении [Плещенко и др., 1999: 40-41; Кожина и др., 2008: 293]. В английских научных текстах форма множественного числа употребляется значительно чаще [Комиссаров, 1990: 117], что создает предпосылки для систематических различий между переводными и непереводными текстами по данному показателю.

*Доля абстрактных имен существительных* (относительно всех существительных), включенная в анализ как показатель реализации характерного для научного стиля преобладания абстрактной лексики над конкретной. Абстрактные существительные определялись путем проверки конца лемматизированных существительных на наличие суффиксов / суффиксоидов абстрактных существительных: *-ние, -ствие, -тие* (образование, следствие, развитие), *-ция* (концентрация, реакция), *-ость* (зависимость, устойчивость), *-есть* (тяжесть), *-ство* (вещество, количество), *-изм* (механизм, организм), *-логия* (методология), *-графия* (спектрография), *-патия* (невропатия), *-метрия* (спектрометрия), *-номия* (таксономия) и т. д.

Третья группа индикаторов, связанная с функционированием имён прилагательных:

*Доля имён прилагательных* (относительно всех словарных токенов). Высокая доля прилагательных способствует реализации точности и информационной насыщенности научного изложения, поскольку прилагательные позволяют более точно охарактеризовать свойства и признаки описываемых объектов и явлений.

*Доля кратких имен прилагательных* (относительно всех прилагательных).

*Доля прилагательных в сравнительной и превосходной степенях сравнения* (относительно всех прилагательных).

Четвёртая группа индикаторов, связанная с особенностями функционирования глаголов:

*Доля глаголов* (относительно всех словарных токенов). При подсчете индикатора учитывались как личные, так и неличные глагольные формы.

Считается, что в русскоязычном научном тексте доля глаголов снижена в силу активной глагольной субстантивации [Комиссаров, 1990: 115]. Данный индикатор в совокупности с индикатором доли имен существительных позволяет оценить степень реализации номинативного характера научного текста.

*Доля глаголов в форме страдательного залога* (относительно всех глаголов). Как упоминалось в разделе 1.4, формы страдательного (пассивного) залога являются одним из ключевых средств, обеспечивающих объективно-нейтральную манеру изложения в научном тексте [Пумпянский, 1981: 11-12, 18; Комиссаров, 1990: 116; Арнольд, 2010: 341; Гуревич, 2024: 19-20], поскольку позволяют сместить фокус с личности исследователя на объект и результаты исследования, описываемые в тексте факты, подчеркнуть универсальность и воспроизводимость научных наблюдений и выводов [Ding, 1998: 117-135; Halliday, Martin, 2003: 212-213], кроме того, являются отражением авторской скромности. При этом растущая в современном английском языке тенденция к более частому употреблению активного залога в текстах научного стиля [Leong, 2021: 77-102; Banks, 2017: 1-17; Арцишевская, 2025: 104-113] создает предпосылки для систематических расхождений между исходными и переводными текстами по данному показателю.

*Доля причастий* (относительно всех глаголов (личных и неличных форм)), *доля причастий в препозиции перед существительным, являющимся главным словом* (относительно всех причастий), и *доля причастий в причастных оборотах* (относительно всех причастий). Причастия и причастные обороты являются характерными для русскоязычного научного стиля, так как позволяют выражать атрибутивные отношения более компактно.

*Доля деепричастий* (относительно всех глаголов (личных и неличных форм)). Деепричастные обороты позволяют лаконично выразить дополнительное действие или обстоятельство, сопровождающее основное действие, делая текст более емким. Поскольку в английском языке деепричастия как грамматическая категория отсутствуют, а аналогичные значения выражаются иными средствами, анализ деепричастий позволяет оценить степень адаптации перевода к нормам ПЯ.

*Доля глаголов несовершенного вида* (относительно всех глаголов (личных и неличных форм и среди личных форм глаголов). Выбор индикаторов основывается на том, что в русском языке глаголы несовершенного вида в научном тексте используются значительно чаще глаголов совершенного вида, поскольку обладают более отвлеченно-обобщенным значением [Плещенко и др., 1999: 42]. Поскольку категория вида в английском языке отсутствует, можно ожидать недостаточную представленность глаголов в этой форме в переводах.

*Доля глаголов в прошедшем / настоящем / будущем времени* (относительно личных форм глаголов).

Представленный набор дополнительных индикаторов обеспечивает возможность оценки степени соответствия переводных текстов нормам научного стиля русского языка. В совокупности с индикаторами основных характеристик (*упрощения, нормализации, экспликации и интерференции*) эти индикаторы формируют базу для более комплексного исследования феномена переводческого языка в англо-русских переводах научных статей.

## **Выводы по главе 2**

Во второй главе была решена задача выработки методологической базы для последующего анализа феномена переводческого языка на материале англо-русских переводов научных статей, были рассмотрены теоретические представления о потенциально универсальных характеристиках перевода, осуществлён отбор индикаторов для их количественного анализа. Так, для каждой характеристики были описаны механизмы её проявления на разных уровнях языка, а также соответствующие этой характеристике индикаторы:

1. Характеристика *упрощение* обусловлена стремлением переводчиков обеспечить предсказуемость восприятия текста перевода, минимизировать риски его неверной интерпретации. Для операционализации анализа характеристики было отобрано 15 формальных индикаторов, куда вошли: (1-4) разные версии коэффициента лексической вариативности (*TTR, Log TTR, Modified TTR, Standardized TTR*), (5) коэффициент лексической плотности, (6) средний ранг

используемой в тексте лексики, (7) средняя доля наиболее частотных знаменательных слов подкорпуса, (8-9) средняя длина слов в символах и в слогах, (10-12) средняя длина предложений в символах, токенах и клаузах, (13) средняя длина клауз в токенах, (14) средняя доля простых предложений, (15) индекс удобочитаемости Флэша и (16) нормализованные частоты разных знаков препинания.

2. Характеристика *нормализация* представляет собой тенденцию к более частому употреблению в тексте перевода типизированных форм и структур ПЯ. С целью её анализа выбрано 3 индикатора: (1) лексическая повторяемость знаменательной лексики, (2) среднее значение показателя поточечной взаимной информации (*Pointwise Mutual Information*) словарных биграмм в подкорпусе в двух вариациях (*PMI* и *Modified PMI*), а также (3) доля биграмм слов, значение *PMI / Modified PMI* которых выше 0.

3. Характеристика *экспликация* проявляется в повышении объяснительного потенциала переводного текста. С целью её анализа выбраны индикаторы, охватывающие лексические и грамматические средства обеспечения когезии. Лексическая когезия анализировалась с использованием индикаторов среднего косинусного сходства (1) пар слов и (2) соседних предложений в тексте. Грамматическая когезия – через индикаторы (3) эксплицитного называния, (4) единичного именованного, (5) средней длины именованных сущностей, а также нормализованные частоты дискурсивных маркеров (6) локального и (7) глобального уровня связности научного текста.

4. Характеристика *интерференция* состоит в переносе структурных особенностей исходного текста в текст перевода. С целью анализа характеристики был выбран 1 индикатор: нормализованные частоты частеречных n-грамм (униграмм, биграмм, триграмм). Кроме того, для более детализированного анализа переводных текстов было принято решение расширить анализ частеречных последовательностей специальными тегами, позволяющими фиксировать следующие структуры: придаточные предложения, вводимые союзами и союзными словами, синтетические и аналитические формы степеней сравнения имен

прилагательных и наречий, сочетания с местоимениями разных разрядов, составные глагольные сказуемые, составные глагольные сказуемые с модальными глаголами, конструкции с модальными словами, вводные конструкции, конструкции с сослагательным наклонением (с частицей *бы*), деепричастные и причастные обороты, конструкции с причастием в функции определения в препозиции к определяемому слову, целевые конструкции с инфинитивом, страдательные глагольные конструкции (страдательный залог).

5. В анализ были включены дополнительные индикаторы, расширяющие спектр анализа. Данная группа индикаторов была сформирована на основе стилеобразующих черт научного стиля русского языка. В группу вошло 19 индикаторов: нормализованные частоты дискурсивных маркеров (1) выражения авторского начала в тексте и (2) взаимодействия автора с читателем; (3) доля существительных в тексте; доля существительных (4) среднего рода, (5) в форме единственного числа; (6) доля абстрактных существительных; (7) доля прилагательных в тексте, (8) доля кратких прилагательных, доля прилагательных в формах (9) сравнительной и (10) превосходной степени; (11) доля глаголов в тексте, доля глаголов (12) в форме страдательного залога, (13) доля причастий относительно всех глаголов, (14) доля причастий в препозиции перед существительным, (15) доля причастий, входящих в состав причастных оборотов, (16) доля деепричастий относительно всех глаголов, (17) доля глаголов несовершенного вида среди всех глаголов и среди личных форм глаголов, (18) доля глаголов в форме прошедшего / настоящего / будущего времени и (19) нормализованные частоты местоимений разных разрядов.

Подводя итог, отметим, что комплекс из 46 отобранных формальных индикаторов обладает потенциалом для автоматизированного выявления характеристик феномена переводческого языка и обеспечивает методологическую основу для проведения эмпирического анализа, представленного в последующих главах настоящей работы.

## ГЛАВА 3. ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК ПЕРЕВОДЧЕСКОГО ЯЗЫКА НА МАТЕРИАЛЕ НАУЧНЫХ СТАТЕЙ

### 3.1. Формирование исследовательского корпуса

В основу практической части настоящего исследования положен корпусный метод, являющийся основным инструментом дескриптивных исследований перевода, включая исследования феномена переводческого языка [Baker, 1993: 223-250; Baker, 1996:175-186; Chesterman, 2004b: 49; Zanettin, 2013: 20-32; Куниловская, 2015: 13-14; Мельник, 2015: 35]. Данный метод обеспечивает эмпирическую базу для анализа переводческих характеристик на большом массиве текстовых данных, позволяя обнаруживать те закономерности переводов, которые было бы трудно установить, основываясь на анализе отдельных переводных произведений.

Чаще всего рамках исследований феномена переводческого языка принято использовать так называемый монолингвальный сопоставимый корпус текстов (*monolingual comparable corpora*) [Baker, 1995: 223-243; Laviosa, 1997: 289-320; Laviosa, 1998: 557-570; Olohan, 2001: 423-432; Olohan, 2002: 419-429; Baroni, Bernardini, 2003: 1-2; Snell-Hornby, 2006: 125-126; Granger, 2010: 14-21; Ilisei, 2012; Kunilovskaya, Kutuzov, 2017: 80-81 и др.]. В отличие от параллельного корпуса, монолингвальный сопоставимый корпус состоит не из оригинальных и переводных произведений, а подразделяется на два (или более) монологичных подкорпуса. Один из этих подкорпусов всегда содержит непере译ные тексты на языке перевода, а другой – независимые от них переводы на этот язык. Выбор сопоставимого корпуса при исследовании феномена переводческого языка обосновывается тем, что непере译ные тексты на языке перевода способны выступать как эталонные, репрезентативные образцы естественного функционирования данного языка в речи носителей, служа основой для определения языковой нормы, относительно которой уже оценивается степень отклонения переводных текстов. Кроме того, как подчеркивал А. Л. Пумпянский [Пумпянский, 1981: 30-31], наиболее надежным критерием истины является практика, то есть то, как язык функционирует в реальной речи носителей соответствующего литературного языка, поскольку то,

как язык используется большинством носителей в определенных коммуникативных ситуациях, нельзя считать неверным. Другим аргументом в пользу выбора сопоставимого корпуса является то, что с его помощью можно выявлять особые языковые паттерны, которые либо свойственны исключительно переводным текстам, либо встречаются в них значительно чаще или значительно реже, чем в непереводах произведений на ПЯ [Baker, 1995: 235].

Что касается методологического подхода, в настоящем исследовании используется подход «снизу вверх» (см. раздел 1.3), предполагающий ограничение спектра анализа некоторым подмножеством текстов (в данном случае – подмножеством текстов научных статей), отобранных на основе строго определенных критериев, с последующим постепенным расширением области применения формулируемых выводов. Выбор подхода «снизу вверх» обусловлен тем, что он обеспечивает более высокую степень эмпирической достоверности результатов и позволяет избежать преждевременных обобщений относительно универсальности выявляемых характеристик для всех переводов вообще. Кроме того, он предполагает осторожность делаемых выводов и их строгую привязку к конкретному исследовательскому материалу, что представляется критически важным в контексте дискуссии о самой применимости понятия ‘универсальность’ к характеристикам переводных текстов [Olohan, 2002: 419-429; Chesterman, 2004a: 1-13; Chesterman, 2004b: 33-49; Toury, 2004: 15-32].

Так, при формировании сопоставимого корпуса тексты переводов отбираются в соответствии с заранее заданными выборочными критериями, а непереводами подбираются таким образом, чтобы соответствовать текстам переводного подкорпуса. Единые критерии отбора обеспечивают приемлемый уровень сопоставимости между непереводами и переводными подкорпусами, позволяя исследователю сосредоточиться на различиях, проистекающих из самой природы переводных текстов (т. е. их переводной природы). Иными словами, такой подход позволяет минимизировать влияние факторов, не связанных непосредственно с переводческим процессом [Laviosa, 1997: 292-293; Laviosa, 1998: 557-570].

Важнейшим критерием при формировании и оценке любого корпуса текстов является его репрезентативность [Захаров, 2005], заключающаяся в необходимо-достаточном и пропорциональном представлении в нём текстов, относящихся к тому типу языка, для изучения которого этот корпус создается. Корпусы, создаваемые с целью репрезентации определённого типа текстов, строятся на основе заранее определённой выборочной рамки (*sampling frame*), задающей параметры исследуемой языковой популяции и очерчивающей пространство, внутри которого допустим сбор данных [McEneaney, Hardie: 2012: 8-11]. Такая выборочная рамка по сути представляет собой единые выборочные критерии, среди которых могут быть: вовлеченная в процесс перевода языковая пара и направление перевода; временной период и культура, в условиях которых выполняется перевод; тип переводимого текста (текст определенного стиля и жанра, определенной тематики); переводчик или группа переводчиков, объединенных общими признаками; единые редакционные нормы или требования определенного издательства / журнала или типа журналов, в которых публикуется перевод [Chesterman, 2004b: 40-41].

В настоящем исследовании спектр анализа ограничен непереводаемыми текстами научных статей и переводами научных статей с английского языка на русский, выполненными в определенный временной период и относящимися к определенным предметным областям. В целях обеспечения сопоставимости материала в основу выборочной рамки легли следующие критерии:

- принадлежность текстов к собственно научному стилю;
- соответствие текстов жанру «научная статья»;
- принадлежность текстов к единой тематике (социологической или медицинской).

Всего было сформировано два сопоставимых корпуса – корпус текстов научных статей по социологии и корпус научных статей по медицине. Рассмотрим каждый из них более подробно.

В корпус социологической тематики вошли:

1. Непере译ной подкорпус, состоящий из полнотекстовых версий научных статей социологической тематики на русском языке;
2. Переводной подкорпус полнотекстовых версий научных статей социологической тематики (англо-русские ручные переводы).

При формировании переводных подкорпусов в настоящем исследовании принципиальное значение имела верификация типа перевода (ручной vs. машинный), поскольку случайное включение в подкорпус ручных переводов текстов, выполненных с использованием систем МП, могло исказить результаты исследования характеристик, специфичных для ручных переводов. В связи с этим факт принадлежности текста к ручным переводам устанавливался на основании следующих критериев: (1) указание имени переводчика в тексте статьи или (2) в выходных данных журнала, либо (3) упоминание переводчика в составе редакционной коллегии журнала как штатного сотрудника.

Каждый подкорпус корпуса по социологии включил 59 текстов, а общий объем каждого из подкорпусов составил приблизительно 310 000 словоформ. Все статьи (как непере译ные, так и переведенные) были вручную извлечены из рецензируемых научных периодических изданий по социологии (см. СПИСОК ИСТОЧНИКОВ). При выборе текстов для включения в переводной подкорпус брались исключительно те статьи, чьи оригиналы были опубликованы на английском языке и чьи авторы на момент публикации были аффилированы с университетами или исследовательскими организациями Великобритании или США. Для наглядности данные о собранном корпусе представлены в Табл. 2.

Таблица 2 – Данные о корпусе текстов статей социологической тематики

Данные о подкорпусах	Непере译ные статьи (Non-translated)	Переведенные статьи (Human-translated)
Количество статей	59	
Объем в токенах	≈310 000	
Период публикации	с 1990-х по 2025 гг.	

В свою очередь, в корпус медицинской тематики вошли:

1. Непере译ной подкорпус, состоящий из полнотекстовых версий научных статей медицинской тематики на русском языке;

2. Переводной подкорпус полнотекстовых версий научных статей медицинской тематики (англо-русские *ручные переводы*);

3. Переводной подкорпус полнотекстовых версий научных статей медицинской тематики (англо-русские *машинные переводы*, выполненные системой *Google Translate*).

В данном случае в первые два подкорпуса (непереводных текстов и ручных переводов) вошло по 70 текстов, а общий объем каждого из них составил приблизительно 216 500 словоформ. В третий подкорпус (подкорпус МП) вошло 50 текстов, общий объем которых также равнялся примерно 216 500 словоформам.

По причине того, что подавляющее число медицинских исследований выполняются целой командой авторов, причем нередко члены этой команды могут быть из разных стран, при выборе статей для включения в переводные подкорпусы соблюдались следующие условия: (1) оригинал статьи должен был быть опубликован на английском языке, (2) первые несколько авторов или корреспондирующий автор статьи должны были быть аффилированы с британским или американским университетом или исследовательским центром.

Отбор текстов научных статей всех трех подкорпусов производился из рецензируемых медицинских периодических изданий (см. СПИСОК ИСТОЧНИКОВ). Данные о медицинском корпусе текстов представлены в Табл. 3.

Таблица 3 – Данные о корпусе текстов статей медицинской тематики

Данные о подкорпусах	Непереводные статьи (Non-translated)	Переводные статьи (Human-translated)	Переводные статьи (Machine-translated)
Количество статей	70		50
Объем в токенах	≈216 500		
Период публикации	2000-2025 гг.		

Сформированные корпуса текстов требовали специального инструментария для систематического извлечения и вычисления показателей индикаторов феномена переводческого языка, описанных в Главе 2 настоящей работы. С этой целью в рамках настоящего исследования был разработан программный комплекс – *анализатор феномена переводческого языка*<sup>10</sup>, позволяющий в

<sup>10</sup> Анализатор феномена переводческого языка. URL: <https://pypi.org/project/translationese-analyzer/>

автоматизированном режиме осуществлять количественный анализ текстов по широкому спектру индикаторов переводческих характеристик. Данный программный комплекс был разработан на базе языка программирования *Python* и, помимо анализа индикаторов характеристик переводов, предназначен также и для автоматизированного сбора корпусов переводных и непереводных текстов. Описание интерфейса программного комплекса приведено в ПРИЛОЖЕНИИ Г.

Разработка программного комплекса обусловлена несколькими соображениями. Во-первых, объём анализируемого материала (свыше 500 000 словоформ в социологическом корпусе и более 600 000 – в медицинском) делает ручную обработку текстов практически невыполнимой задачей. Во-вторых, автоматизация процесса вычисления показателей обеспечивает единообразие применяемых процедур ко всем текстам корпуса, что критически важно для обеспечения валидности получаемых результатов. В-третьих, существующие программные решения для корпусного анализа не предоставляют возможности одновременного измерения всего комплекса представленных Главе 2 индикаторов и требуют значительной адаптации для работы с русскоязычными текстами.

Архитектура разработанного программного комплекса предусматривает модульную структуру, в которой каждый модуль отвечает за извлечение и расчёт определённой группы индикаторов. Такая организация обеспечивает гибкость системы и возможность её дальнейшего расширения. Программный комплекс включает модули для анализа всех перечисленных в Главе 2 индикаторов. Функционирование программного комплекса предполагает последовательное выполнение нескольких этапов обработки текстов. На первом этапе осуществляется предварительная обработка корпуса, включающая токенизацию, лемматизацию и морфологическую разметку текстов. Данные операции реализованы с использованием современных библиотек для обработки естественного русского языка<sup>11,12</sup>. На втором этапе к размеченным текстам применяются алгоритмы расчёта индикаторов в соответствии с метриками,

---

<sup>11</sup> Морфологический анализатор для русского языка *pymorphy2*: <https://pymorphy2.readthedocs.io/en/stable/>

<sup>12</sup> Библиотека для обработки естественного русского языка *Natasha*: <https://pypi.org/project/natasha/>

представленными в Главе 2. Результаты измерений для каждого проанализированного текста сохраняются в SQLite-базу данных в виде таблиц, где каждая таблица связана с одной из характеристик переводческого языка и содержит вычисленные на этапе анализа значения индикаторов, а также метаданные о тексте. Табличная организация данных позволяет автоматически извлекать значения индикаторов и выполнять подсчет средних показателей по каждому из подкорпусов (переводному и непереводному). Кроме того, такая структура хранения данных обеспечивает возможность сопоставления результатов между подкорпусами и выявления систематических расхождений в количественных характеристиках переводных и непереводных текстов.

Формирование описанных выше исследовательских корпусов текстов (социологической и медицинской тематики) производилось с помощью разработанного программного комплекса. Все тексты обоих корпусов прошли предобработку: из них предварительно вручную были удалены аннотации, списки ключевых слов, таблицы, формулы, тексты сносок, рисунки и списки литературы. Автоматически (с помощью программы) удалялись внутритекстовые номера сносок и внутритекстовые ссылки на литературу. Результаты автоматической предобработки были дополнительно проверены и откорректированы вручную: устранялись случаи неполного удаления библиографических ссылок, исправлялись лишние переносы и склейки слов. Затем все тексты были проанализированы с помощью программного комплекса на наличие индикаторов переводческих характеристик, описанных в Главе 2.

### 3.2. Методы анализа результатов исследования

Полученные в результате анализа количественные данные послужили основой для дальнейшего статистического сопоставления показателей индикаторов между переводными и непереводными подкорпусами с целью выявления статистически значимых различий и оценки их практической значимости. Выбор статистических методов определялся характером распределения анализируемых данных, поскольку распределение лингвистических показателей в корпусных исследованиях различается в зависимости от природы анализируемого явления

Так, параметрические тесты, включая t-критерий Уэлча, предполагают приближенную нормальность распределения и гомогенность дисперсий (в случае классического t-критерия) либо корректируют её (как в случае Welch's t-test). Непараметрические методы, такие как критерий Манна–Уитни, применяются в тех случаях, когда распределение данных существенно отклоняется от нормального: при наличии выраженной асимметрии, значительных выбросов или большого количества нулевых значений.

Для сопоставления показателей индикаторов, демонстрирующих распределение, близкое к нормальному, применялся двухвыборочный двусторонний t-критерий Уэлча (*Welch's t-test*) [Welch, 1947: 28-35] для взвешенных наблюдений. Данный подход был адаптирован специально для анализа корпусов, тексты в которых могут иметь различную длину. В отличие от классического t-критерия, где все наблюдения имеют равный вес, в настоящем исследовании каждый текст получал вес, равный количеству содержащихся в нём словарных токенов. Данный подход обоснован тем, что более длинные тексты вносят больший вклад в итоговую статистику, поскольку имеют бóльшие веса.

Взвешенное среднее для каждой группы текстов рассчитывалось по формуле:

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \quad (11)$$

где  $\bar{x}_w$  – взвешенное среднее арифметическое значение показателя,

$x_i$  – значение показателя для  $i$ -го текста в подкорпусе,

$w_i$  – вес  $i$ -го текста (в словарных токенах),

$n$  – общее количество текстов в подкорпусе.

Взвешенная дисперсия также вычислялась с поправкой на неравномерность весов по следующей формуле:

$$s_w^2 = \frac{(\sum_{i=1}^n w_i (x_i - \bar{x}_w)^2) \times (\sum_{i=1}^n w_i)}{(\sum_{i=1}^n w_i)^2 - \sum_{i=1}^n w_i^2}, \quad (12)$$

где  $s_w^2$  – значение взвешенной дисперсии,

$\bar{x}_w$  – взвешенное среднее арифметическое значение показателя,

$x_i$  – значение показателя для  $i$ -го текста в подкорпусе,

$w_i$  – вес  $i$ -го текста (в словарных токенах),

$n$  – общее количество текстов в подкорпусе.

Для оценки эквивалентного числа равновесовых наблюдений использовался эффективный размер выборки по формуле Киша [Kish, 1965]

$$n_{eff} = \frac{(\sum_{i=1}^n w_i)^2}{\sum_{i=1}^n w_i^2}, \quad (13)$$

где  $n$  – общее количество текстов в подкорпусе,

$w_i$  – вес  $i$ -го текста (в словарных токенах).

Эффективный размер выборки отражает эквивалентное количество равновесовых наблюдений и учитывает снижение статистической мощности при неравномерном распределении весов.

Далее рассчитывалась  $t$ -статистика взвешенного Welch-теста по формуле:

$$t = \frac{\bar{x}_{w1} - \bar{x}_{w2}}{\sqrt{(s_{w1}^2/n_{eff1}) + (s_{w2}^2/n_{eff2})}}, \quad (14)$$

где  $\bar{x}_{w1}$  и  $\bar{x}_{w2}$  – взвешенные средние арифметические значения показателя для первого и второго подкорпуса соответственно,

$s_{w1}^2$  и  $s_{w2}^2$  – взвешенные дисперсии показателя для первого и второго подкорпуса соответственно,

$n_{eff1}$  и  $n_{eff2}$  – эффективные размеры выборки для первого и второго подкорпуса соответственно.

Количество степеней свободы ( $df$ ) определялось по формуле Уэлча для случая неравных дисперсий:

$$df = \frac{\left(\frac{s_{w1}^2}{n_{eff1}} + \frac{s_{w2}^2}{n_{eff2}}\right)^2}{\frac{(s_{w1}^2/n_{eff1})^2}{n_{eff1}-1} + \frac{(s_{w2}^2/n_{eff2})^2}{n_{eff2}-1}} \quad (15)$$

где  $s_{w1}^2$  и  $s_{w2}^2$  – взвешенные дисперсии показателя для первого и второго подкорпуса соответственно,

$n_{eff1}$  и  $n_{eff2}$  – эффективные размеры выборки для первого и второго подкорпуса соответственно.

Так, t-критерий проверял нулевую гипотезу об отсутствии различий между взвешенными средними значениями анализируемого индикатора по подкорпусам. Статистическая значимость оценивалась на уровне  $\alpha=0,05$ . При сравнении показателей между подкорпусами рассчитывались:

- $\Delta$  – разность взвешенных средних как мера практической значимости различий, выражаемая в единицах исследуемой метрики;
- *p-значение* – вероятность получить наблюдаемые или более экстремальные различия при условии истинности нулевой гипотезы;
- *95% CI  $\Delta$*  – 95%-ый доверительный интервал для разности взвешенных средних;
- *r* – размер наблюдаемого эффекта ( $r \sim 0,1$  – малый эффект,  $r \sim 0,3$  – средний,  $r \geq 0,5$  – большой эффект), вычисляемый по формуле:

$$r = \frac{|t|}{t^2 + df}, \quad (16)$$

где  $t$  – значение t-статистики критерия Уэлча,

$df$  – число степеней свободы, рассчитанное по формуле Уэлча (15).

Дополнительно подсчитывался коэффициент вариации (*coefficient of variation, CV*), представляющий отношение стандартного отклонения к среднему взвешенному значению, выраженное в процентах. Коэффициент вариации используется как универсальный показатель относительной изменчивости (колеблемости) данных. В отличие от стандартного отклонения, *CV* не зависит от единиц измерения, что делает его удобным инструментом для сравнения степени вариативности показателей. В научной литературе принято считать, что значение  $CV \leq 10\%$  указывает на крайне низкую изменчивость значений показателя между

элементами выборки (в данном случае – текстами),  $CV \approx 10-20\%$  свидетельствует о средней стабильности;  $CV \approx 20-30\%$  характеризует значительную вариативность; а превышение порога  $CV > 30\%$  считается высокой вариативностью и свидетельствует о неоднородности данных.

Индикаторы, сопоставление которых проводилось с использованием t-критерия Уэлча, приведены в Табл. 4.

Таблица 4 – Индикаторы, анализ которых осуществлялся с использованием двухвыборочного t-критерия Уэлча

Характеристика перевода	Формальные индикаторы
Упрощение	Лексическая плотность, лексическая вариативность, средний ранг слов, средняя доля наиболее частотных знаменательных слов подкорпуса, средняя длина слов, средняя длина предложений, средняя длина клаузы в токенах, средняя доля простых предложений, индекс удобочитаемости Флеша, нормализованные частоты знаков препинания
Нормализация	Коэффициент лексической повторяемости знаменательной лексики
Экспликация	Среднее косинусное сходство пар слов и соседних предложений текста
Интерференция	Нормализованные частоты частеречных последовательностей
Прочие индикаторы, отражающие особенности научного стиля	<ul style="list-style-type: none"> <li>- Доля имён существительных в тексте, доля имён существительных среднего рода, доля имён существительных в единственном числе, доля абстрактных имён существительных,</li> <li>- Доля имён прилагательных в тексте, доля кратких имён прилагательных, доля прилагательных в сравнительной и превосходной степенях,</li> <li>- Доля глаголов в тексте, доля глаголов в форме страдательного залога, доля причастий; доля причастий в препозиции перед существительным; доля причастий в составе причастных оборотов, доля деепричастий, доля глаголов несовершенного вида, доля глаголов в разных временных формах</li> <li>- Нормализованные частоты местоимений разных разрядов</li> </ul>

Для сопоставления показателей индикаторов, характеризующихся распределением, отличающимся от нормального, применялся непараметрический U-критерий Манна–Уитни (*Mann–Whitney U*) – ранговый непараметрический тест, проверяющий нулевую гипотезу о совпадении распределений двух независимых выборок (а при одинаковой форме распределений – о равенстве их медиан) [Mann, Whitney, 1947]. Взвешивание (как в случае с t-критерием Уэлча) не применялось, поскольку тест основан на рангах и в стандартной постановке не предполагает учёта весов наблюдений. Отсутствие необходимости учёта весов позволило

использовать при расчете  $U$ -статистики и  $p$ -значений готовый программный пакет для статистического анализа *GraphPad Prism 10*.

Дополнительно для каждого подкорпуса вычислялись робастные характеристики: медиана  $\tilde{x}$ , квартили  $Q_1$ ,  $Q_3$ , межквартильный размах  $IQR=Q_3-Q_1$  и медианное абсолютное отклонение  $MAD = \text{median}(|x_i - \tilde{x}|)$ .

Помимо  $p$ -значения по Манну–Уитни, с помощью инструмента *GraphPad Prism 10* вычислялись:

- Размер эффекта  $r$  ( $r \sim 0.1$  – малый эффект,  $\sim 0.3$  – средний,  $\geq 0.5$  – большой эффект);
- Оценка сдвига Ходжеса–Лемана  $HL$ , представляющая типичную разницу между сравниваемыми группами (подкорпусами) в исходных единицах показателя. Рассчитывается как медиана всех попарных разностей между наблюдениями из первой и второй выборок. Знак  $HL$  показывает направление (положительное значение  $HL$  означает, что в первой группе (т. е. в группе непереводных текстов) типичное значение индикатора на  $HL$  единиц больше). Если 95%-ый доверительный интервал  $HL$  не охватывает ноль, различие считается статистически значимым. В настоящей работе показатель  $HL$  выполняет в рамках критерия Манна–Уитни ту же функцию, что и разность взвешенных средних  $\Delta$  в рамках  $t$ -критерия Уэлча: служит робастной мерой практической значимости обнаруженных различий в исходных единицах измерения;
- Размер эффекта  $A_{12}$  [Vargha, Delaney, 2000], то есть вероятность того, что случайно выбранное значение из первой группы (непереводного подкорпуса) больше случайно выбранного значения из второй группы (переводного подкорпуса) (если  $A_{12} \approx 0.5$ , эффект отсутствует; чем дальше значение  $A_{12}$  от  $0.5$ , тем более сильным считается эффект). Формула эффекта  $A_{12}$ :

$$A_{12} = \frac{U_1}{n_1 n_2}, \quad (17)$$

где  $U_1$  – количество пар, в которых значение из первой группы больше значения из второй группы,

$n_1$  и  $n_2$  – объёмы выборок (количество текстов в них).

$U_1$  рассчитывается по формуле:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}, \quad (18)$$

где  $R_1$  – сумма рангов наблюдений первой группы в объединённой ранжированной выборке.

Индикаторы, сопоставление которых проводилось с использованием двустороннего U-критерия Манна–Уитни, приведены в таблице 5.

Таблица 5 – Индикаторы, анализ которых осуществлялся с использованием двустороннего U-критерия Манна–Уитни

Характеристика перевода	Формальные индикаторы
Экспликация	Показатели, основанные на употреблении именованных сущностей ( <i>explicit naming, single naming, mean named entities length</i> ); нормализованные частоты дискурсивных маркеров локального уровня и глобального уровня связности текста
Прочие показатели, отражающие особенности научного стиля	нормализованные частоты дискурсивных маркеров авторского начала в тексте и маркеров взаимодействия автора с читателем

Необходимость применения непараметрического метода для показателей, подразумевающих учёт именованных сущностей (*explicit naming, single naming, mean named entities length*), и для дискурсивных маркеров обусловлена общей чертой их распределения: оба типа показателей характеризуются значительным количеством нулевых или близких к нулю значений, что исключает предположение о нормальности. Что касается именованных сущностей, то в зависимости от темы статье часть текстов может содержать минимальное количество именованных сущностей или не содержать их вовсе, в то время как другие тексты могут демонстрировать их высокую концентрацию, что приводит к появлению множественных нулевых или близких к нулю значений показателей между текстами анализируемого подкорпуса. Дискурсивные маркеры представляют собой факультативные средства текстовой организации, их использование во многом определяется индивидуальными предпочтениями автора, коммуникативными задачами текста, вследствие этого тексты могут не содержать определённых категорий маркеров вовсе или содержать их в небольшом количестве, что приводит к избыточному количеству нулевых значений или значений, близких к нулю.

В следующем разделе перейдем к интерпретации выявленных в результате анализа различий между подкорпусами по каждому из формальных индикаторов, отражающих характеристики переводческого языка.

### 3.3. Интерпретация результатов анализа индикаторов феномена переводческого языка

Перейдем к интерпретации результатов анализа индикаторов переводческих характеристик, проведенного на материале сформированных в рамках исследования текстовых корпусов. Для удобства изложения результатов анализа по каждому из корпусов далее в работе используются следующие сокращения:

СОЦ-РП – Социологический корпус (непереводные тексты и *ручные переводы*);

МЕД-РП – Медицинский корпус (непереводные тексты и *ручные переводы*);

МЕД-МП – Медицинский корпус (непереводные тексты и *машинные переводы*).

#### Индикаторы характеристики упрощение

Полные результаты статистического анализа всех индикаторов *упрощения* представлены ПРИЛОЖЕНИИ Д. В таблице 6 приведены результаты анализа индикаторов, связанных с *упрощением* текста на лексическом уровне

Таблица 6 – Результаты анализа индикаторов *лексического упрощения*

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	<i>p</i>
Lexical density (%)				
СОЦ-РП	64.79	62.54	+2.25	***
МЕД-РП	66.75	64.84	+1.91	***
МЕД-МП	66.75	62.74	+4.01	***
TTR (%)				
СОЦ-РП	28.18	27.46	+0.72	<i>n.s.</i>
МЕД-РП	29.12	26.27	+2.85	**
МЕД-МП	29.12	24.01	+5.11	***

Продолжение таблицы 6

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	<i>p</i>
Log TTR(%)				
СОЦ-РП	85.04	84.94	+0.1	<i>n.s.</i>
МЕД-РП	84.56	83.27	+1.28	**
МЕД-МП	84.56	82.75	+1.81	***
Modified TTR				
СОЦ-РП	1954.77	1975.27	-20.5	<i>n.s.</i>
МЕД-РП	1760.88	1676.61	+84.27	**
МЕД-МП	1760.88	1729.73	+31.15	<i>n.s.</i>
STTR (%)				
СОЦ-РП	47.64	47.99	-0.35	<i>n.s.</i>
МЕД-РП	43.17	40.41	+2.75	**
МЕД-МП	43.17	38.55	+4.62	***
MWR1				
СОЦ-РП	1827.00	1689.31	+137.69	***
МЕД-РП	2631.42	2357.54	+273.88	***
МЕД-МП	2631.42	2529.52	+101.9	<i>n.s.</i>
MWR2				
СОЦ-РП	892.54	863.36	+29.18	*
МЕД-РП	1041.56	1027.10	+14.47	<i>n.s.</i>
МЕД-МП	1041.56	988.49	+53.07	**
List head ratio(%)				
СОЦ-РП	9.18	8.6	+0.58	<i>n.s.</i>
МЕД-РП	15.12	16.25	-1.13	<i>n.s.</i>
МЕД-МП	15.12	15.37	-0.25	<i>n.s.</i>

По показателю *лексической плотности* во всех трёх корпусах фиксируется статистически значимое снижение в переводах по сравнению непереводаемыми текстами, а именно в ручных переводах социологических ( $\Delta=+2.25$ ,  $p<0.001$ ,  $r\approx 0.4$ ) и медицинских статей ( $\Delta=+1.91$ ,  $p<0.001$ ,  $r\approx 0.31$ ) и в машинных переводах медицинских статей ( $\Delta=+4.01$ ,  $p<0.001$ ,  $r\approx 0.54$ ). Примечательно, что в МП снижение *лексической плотности* выражено в значительно большей степени ( $r\approx 0.54$  – большой эффект), чем в ручных переводах. Полученный результат

подтверждает гипотезу об *упрощении*: ручные и в особенности машинные переводы содержат больше функциональной и меньше знаменательной лексики. В случае с ручными переводами снижение *лексической плотности* может быть следствием более выраженной *экспликации* логико-смысловых отношений в сравнении с непереводаемыми текстами: *экспликация* связана с более частым использованием в переводе таких функциональных единиц языка, как союзы и местоимения. Кроме того, наблюдаемое снижение лексической плотности во всех случаях может быть следствием *интерференции*: английский язык как более аналитический по своей природе активнее использует разнообразные функциональные элементы, которые в процессе перевода иногда могут переноситься в текст перевода без необходимой адаптации к нормам ПЯ.

Индикаторы *лексической вариативности* демонстрируют зависимость от предметной области и типа перевода. В корпусе СОЦ-РП ни по одной из версий *лексической вариативности* статистически значимых различий между подкорпусами не обнаружено. В корпусе МЕД-РП, напротив, наблюдается систематическое снижение *лексической вариативности* по всем четырём версиям индикатора: *TTR* ( $\Delta=+2.85$ ,  $p<0.01$ ,  $r\approx 0.26$ ), *Log TTR* ( $\Delta=+1.28$ ,  $p<0.01$ ,  $r\approx 0.30$ ), *Modified TTR* ( $\Delta=+84.27$ ,  $p<0.01$ ,  $r\approx 0.27$ ), *STTR* ( $\Delta=+2.75$ ,  $p<0.01$ ,  $r\approx 0.28$ ). Следует отметить, что снижение показателя *Modified TTR* указывает на уменьшение в ручных переводах по медицине доли лексических единиц, встречающихся в тексте единожды (*hapax legomena*). Можно предположить, что данный результат свидетельствует о том, что в целях облегчения понимания текста читателем переводчики или опускают такие лексемы вовсе, или отдают предпочтение другим, более частотным вариантам ПЯ. Вместе с тем в случае, если снижение *Modified TTR* касается избегания редкой терминологии и / или её замены на более общеупотребительные варианты, данная тенденция может приводить к снижению смысловой точности текста: терминология, являющаяся, как правило, низкочастотной в масштабах языка, нередко маркирует тонкие концептуальные различия, существенные для точного понимания данных исследования. В корпусе МЕД-МП в подкорпусе машинных переводов снижение *лексической*

*вариативности* выражено в ещё большей степени по трём из четырёх версий индикатора: *TTR* ( $\Delta=+5.11$ ,  $p<0.001$ ,  $r\approx 0.53$ ), *Log TTR* ( $\Delta=+1.81$ ,  $p<0.001$ ,  $r\approx 0.47$ ), *STTR* ( $\Delta=+4.62$ ,  $p<0.001$ ,  $r\approx 0.49$ ). Таким образом, гипотеза об *упрощении* через снижение *лексической вариативности* подтверждается только для переводных подкорпусов в МЕД-РП и МЕД-МП, причём эффект значительно усиливается при машинном переводе.

Показатели *среднего ранга слов (MWR)* демонстрируют неоднородную картину по корпусам. Напомним, что более высокое значение *MWR* указывает на преобладание в текстах анализируемого подкорпуса низкочастотной (редкой) лексики, а более низкое значение – на более активное использование слов из верхней части частотного списка [Ляшевская, Шаров, 2009: <http://dict.ruslang.ru/freq.php>]. В корпусе СОЦ-РП в переводах наблюдается статистически значимое снижение по обеим версиям индикатора: *MWR1* ( $\Delta=+137.69$ ,  $p<0.001$ ,  $r\approx 0.40$ ) и *MWR2* ( $\Delta=+29.18$ ,  $p<0.05$ ,  $r\approx 0.23$ ). Данный результат подтверждает гипотезу об *упрощении*: переводчики склонны выбирать более частотные и, как следствие, с большей вероятностью знакомые читателю лексические единицы ПЯ [Hvelplund, 2011: 90], чтобы минимизировать риск неверной интерпретации текста. Кроме того, при высокой когнитивной нагрузке (учитывая принадлежность переводимых текстов к научной сфере) переводчики могут непреднамеренно прибегать к более частотным и, соответственно, более доступным в ментальном лексиконе единицам ПЯ [Liu et al., 2019: 125-149]. Различие в величине эффекта между *MWR1* ( $r\approx 0.40$  – средний эффект) и *MWR2* ( $r\approx 0.23$  – слабый, но ближе к среднему эффект) объясняется различием в обработке слов: при подсчёте *MWR1* словам за пределами первых 6000 слов частотного списка присваивается максимальный ранг 6000, тогда как при подсчёте *MWR2* они исключаются из анализа вовсе. Следовательно, ключевое различие между подкорпусами связано именно с использованием более редкой лексики ПЯ.

В корпусе МЕД-РП статистически значимые различия выявлены только по первой версии индикатора: *MWR1* ( $\Delta=+273.88$ ,  $p<0.001$ ,  $r\approx 0.46$ ). Результат означает, что разница между непереводаемым и переводным подкорпусами в данном случае

заключается преимущественно в использовании сверхредкой лексики (слова с рангом  $> 6000$ ). Следует отметить, что к сверхредкой лексике в масштабах языка относится значительная часть медицинской терминологии, поскольку терминологическая лексика сама по себе является необщеупотребительной и редкой для языка в целом. Можно предположить, что столь стремительное снижение *MWR1* в подкорпусе ручных переводов может указывать на замену редких терминов оригинала более общеупотребительными вариантами ПЯ. Так, например, в клинической практике общепринятыми являются термины *гипотиреоз* или *миопия*, в то время как их более общеупотребительные эквиваленты *снижение функции щитовидной железы* и *близорукость* не обладают необходимой терминологической точностью, хотя и могут на практике заменять закрепленные в медицинской литературе термины. Подобные замены способны приводить к снижению смысловой, терминологической точности текста и представляет особую проблему, которая должна учитываться при переводе медицинской литературы.

В корпусе МЕД-МП наблюдается обратная картина: по версии *MWR1* значимых различий не выявлено, тогда как по версии *MWR2* фиксируется статистически значимое снижение ( $\Delta=+53.07$ ,  $p<0.01$ ,  $r\approx 0.31$ ). Отсутствие различий по версии *MWR1* свидетельствует о том, что в целом МП сохраняет более редкую (в масштабах ПЯ) лексику. Однако если оценка показателя проводится в пределах первых 6000 единиц частотного списка русской лексики (*MWR1*) [Ляшевская, Шаров, 2009: <http://dict.ruslang.ru/freq.php>], система МП «выбирает» из него более частотные варианты.

По показателю *list head ratio*, отражающему среднюю долю лексем из списка наиболее частотных знаменательных слов подкорпуса в тексте, статистически значимых различий ни в одном из корпусов не обнаружено. Отсутствие различий по данному индикатору означает, что тексты непереводных и переводных подкорпусов во всех трёх случаях (СОЦ-РП, МЕД-РП, МЕД-МП) содержат сопоставимую долю наиболее частотной лексики соответствующего подкорпуса.

Далее перейдем к рассмотрению результатов анализа индикаторов, связанных с *упрощением* текста на грамматическом уровне (Табл. 7).

Таблица 7 – Результаты анализа индикаторов *грамматического упрощения*

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	<i>p</i>
Mean word length (chars)				
СОЦ-РП	6.83	6.63	+0.2	***
МЕД-РП	6.88	6.80	+0.08	<i>n.s.</i>
МЕД-МП	6.88	6.57	+0.31	***
Mean word length (syllables)				
СОЦ-РП	2.92	2.84	+0.08	**
МЕД-РП	2.96	2.92	+0.04	<i>n.s.</i>
МЕД-МП	2.96	2.90	+0.06	*
Mean sentence length (chars)				
СОЦ-РП	175.28	145.76	+29.52	***
МЕД-РП	155.56	157.69	-2.13	<i>n.s.</i>
МЕД-МП	155.56	167.43	-11.87	*
Mean sentence length (tokens)				
СОЦ-РП	30.38	25.58	+4.8	***
МЕД-РП	26.45	26.88	-0.42	<i>n.s.</i>
МЕД-МП	26.45	30.08	-3.63	**
Mean sentence length (clauses)				
СОЦ-РП	2.2	2.04	+0.2	<i>n.s.</i>
МЕД-РП	1.7	1.83	-0.13	**
МЕД-МП	1.7	1.9	-0.2	***
Mean clause length (tokens)				
СОЦ-РП	11.5	10.5	+1	***
МЕД-РП	13.5	12.8	+0.7	<i>n.s.</i>
МЕД-МП	13.5	13.76	-0.26	<i>n.s.</i>
Simple sentences ratio (%)				
СОЦ-РП	58.63	56.63	+2	<i>n.s.</i>
МЕД-РП	73.06	68.64	+4.42	**
МЕД-МП	73.06	64.92	+8.14	***
Flesch reading ease score (%)				
СОЦ-РП	-0.5	8.91	-9.41	***
МЕД-РП	2.16	3.32	-1.16	<i>n.s.</i>
МЕД-МП	2.16	4.27	-2.11	<i>n.s.</i>

Показатели *средней длины слов* демонстрируют выраженную зависимость от предметной области. В корпусе СОЦ-РП оба показателя статистически значимо снижаются в переводах: *длина в символах* ( $\Delta=+0.2$ ,  $p<0.001$ ,  $r\approx 0.32$ ) и *длина в слогах* ( $\Delta=+0.08$ ,  $p<0.01$ ,  $r\approx 0.31$ ). Хотя абсолютная величина различий невелика и эффект в обоих случаях является средним, данный результат свидетельствует о том, что в переводах социологических текстов используются несколько более короткие слова, что подтверждает гипотезу о морфологическом *упрощении*. В корпусе МЕД-РП статистически значимые различия отсутствуют по обоим показателям. Отсутствие морфологического упрощения в ручных переводах медицинских текстов может объясняться спецификой медицинского дискурса: терминология в обоих языках преимущественно представлена многоморфемными заимствованиями из греко-латинского словарного фонда, при этом возможности замены подобных терминов структурно более простыми синонимами ограничены императивом терминологической точности (ср. *dysplasia* → *дисплазия*, *presbyopia* → *пресбиопия*). В корпусе МЕД-МП в переводах наблюдается значимое снижение *длины слов в символах* ( $\Delta=+0.31$ ,  $p<0.001$ ,  $r\approx 0.46$ ) и *в слогах* ( $\Delta=+0.06$ ,  $p<0.05$ ,  $r\approx 0.22$ ). Использование более коротких слов в МП может быть вызвано тем, что современные системы МП (в частности системы, основанные на нейронном МП, в том числе Google Translate) ориентированы на выбор более высокочастотных языковых соответствий ПЯ [Vanmassenhove et al., 2019: 222-232], при этом, как известно, более частотные лексемы языка являются в среднем более короткими, чем менее частотные [Corral, Serra, 2020: 1-14]. Данное наблюдение согласуется с приведенными выше результатами анализа среднего ранга слов (*MWR2*) для МП.

Показатели *средней длины предложений* демонстрируют разнонаправленные тенденции в зависимости от корпуса. В корпусе СОЦ-РП только два показателя *средней длины предложений* статистически значимо снижаются в переводе: *в символах* ( $\Delta=+29.52$ ,  $p<0.001$ ,  $r\approx 0.46$ ) и *в токенах* ( $\Delta=+4.8$ ,  $p<0.001$ ,  $r\approx 0.44$ ). Примечательно в этом контексте то, что *число клауз на предложение* и *доля простых предложений* между переводными и непереводными статьями по социологии статистически значимо не различаются, при том что в переводах

*средняя длина клауз* в токенах снижается ( $\Delta=+1$ ,  $p<0.001$ ,  $r\approx 0.43$ ). Данный результат говорит о том, что в данном случае *упрощение* на синтаксическом уровне реализуется не столько через изменение структурной сложности предложений (количества клауз), сколько через изменение количества слов внутри клаузы: при сохранении числа клауз может снижаться средняя длина каждой отдельной клаузы в токенах за счёт *упрощения* её внутренней структуры через использования более компактных словосочетаний. Примечательно также, что коэффициент вариативности показателя *средней длины предложений* (см. ПРИЛОЖЕНИЕ Д) в непереводных текстах значительно выше аналогичного показателя в переводах ( $CV=25.3\%$  против  $CV=13.7\%$  для *длины предложений в символах* и  $CV=25.9\%$  против  $CV=12.8\%$  для *длины предложений в токенах*), что говорит о том, что *средняя длина предложений* между отдельными текстами внутри непереводного социологического подкорпуса варьируется в гораздо большей степени, чем в подкорпусе переводов. Снижение коэффициента вариации практически вдвое может указывать на стандартизацию (*нормализацию*) синтаксических структур: вероятно, переводчики стремятся придерживаться некоторого стандарта в отношении длины предложений.

В корпусе МЕД-РП по показателям *средней длины предложений* (в символах и в токенах) и *средней длины клаузы* (в токенах) различий не выявлено. В то же время в переводах наблюдается некоторое увеличение *среднего числа клауз на предложение* ( $\Delta=-0.13$ ,  $p<0.01$ ,  $r\approx 0.28$ ), сопровождаемое значительным снижением *доли простых предложений* ( $\Delta=+4.42$ ,  $p<0.01$ ,  $r\approx 0.26$ ). Результат указывает на усложнение синтаксической структуры высказываний и может быть интерпретирован как проявление *экспликации*: более сложные синтаксические конструкции (СПП, ССП) позволяют переводчику лучше структурировать информацию и устанавливать более явные логико-смысловые связи между фрагментами высказывания.

В корпусе МЕД-МП наблюдается увеличение *средней длины предложений* в переводах по всем показателям: в символах ( $\Delta=-11.87$ ,  $p<0.05$ ,  $r\approx 0.26$ ), токенах ( $\Delta=-3.63$ ,  $p<0.01$ ,  $r\approx 0.43$ ) и клаузах ( $\Delta=-0.20$ ,  $p<0.001$ ,  $r\approx 0.47$ ); при этом *средняя*

длина клауз (в токенах) статистически значимо не меняется. В то же время в МП стремительно снижается доля простых предложений ( $\Delta=+8.147$ ,  $p<0.001$ ,  $r\approx 0.49$ ). Вероятнее всего, данный результат говорит о том, что система МП воспроизводит синтаксические структуры исходного текста, не выполняя необходимого разбиения исходного английского предложения на более короткие и структурно более простые в переводе там, где это необходимо, что вытекает из их неспособности адекватно реализовывать структурные синтаксические перестройки.

По показателю *удобочитаемости Флеша* различия фиксируются только в корпусе СОЦ-РП: наблюдается значительное увеличение индекса удобочитаемости в ручных переводах по социологии ( $\Delta=-9.41$ ,  $p<0.001$ ,  $r\approx 0.44$ ). Поскольку более высокое значение индекса указывает на бóльшую лёгкость восприятия текста, данный результат подтверждает гипотезу об *упрощении*.

Перейдём к интерпретации результатов анализа *нормализованных частот пунктуационных знаков*. В таблице 8 приведены результаты анализа только тех знаков, по которым были выявлены статистически значимые отличия.

Таблица 8 – Результаты анализа *пунктуационных знаков*

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	<i>p</i>
. – точка (%)				
СОЦ-РП	3.38	3.83	-0.45	***
МЕД-МП	3.76	3.37	+0.39	***
-- тире (%)				
СОЦ-РП	0.70	0.55	+0.16	*
МЕД-РП	0.55	0.27	+0.28	***
МЕД-МП	0.55	0.07	+0.48	***
) – закрывающая скобка (%)				
СОЦ-РП	0.88	0.67	+0.21	*
МЕД-РП	1.81	1.29	+0.52	***
МЕД-МП	1.81	2.12	-0.31	*
( – открывающая скобка (%)				
СОЦ-РП	0.85	0.66	+0.19	*
МЕД-РП	1.78	1.25	+0.53	***
МЕД-МП	1.78	2.12	-0.34	***

Продолжение таблицы 8

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	$p$
« – открывающие кавычки-ёлочки (%)				
СОЦ-РП	1.28	0.71	+0.56	***
» – закрывающие кавычки-ёлочки (%)				
СОЦ-РП	1.28	0.71	+0.56	***
, – запятая (%)				
МЕД-МП	6.40	7.04	-0.64	***

Показатель частоты употребления *точка* демонстрирует разнонаправленные тенденции в корпусах СОЦ-РП и МЕД-МП. В корпусе СОЦ-РП наблюдается статистически значимое увеличение частоты употребления *точек* в переводах ( $\Delta=-0.45$ ,  $p<0.001$ ,  $r\approx 0.35$ ). Данный результат свидетельствует о том, что в ручных переводах по социологии в среднем используется больше предложений, что в целом согласуется с описанным ранее сокращением *средней длины предложений*: поскольку структурно более сложные предложения, как правило, являются более длинными, увеличение частоты точек говорит о *синтаксическом упрощении*, а именно о разбиении более сложных предложений исходного текста на структурно более простые в тексте перевода. Кроме того, как отмечалось в разделе 2.1, точка считается наиболее «сильным» знаком препинания, а при упрощении пунктуации переводчики заменяют более «слабые» знаки препинания более «сильными».

В корпусе МЕД-МП наблюдается статистически значимое снижение частоты употребления *точек* ( $\Delta=+0.39$ ,  $p<0.001$ ,  $r\approx 0.34$ ) в МП, что согласуется с увеличением *средней длины предложений* и указывает на то, что система МП воспроизводит синтаксические структуры исходного текста без необходимого разбиения предложений на более короткие и структурно простые, что является проявлением *интерференции* синтаксических структур ИЯ.

Во всех трёх корпусах как в ручных, так и в машинных переводах наблюдается снижение частоты использования *тире*: в ручных переводах социологических ( $\Delta=+0.16$ ,  $p<0.05$ ,  $r\approx 0.25$ ) и медицинских статей ( $\Delta=+0.28$ ,  $p<0.001$ ,  $r\approx 0.44$ ) и в машинных переводах медицинских статей ( $\Delta=+0.48$ ,  $p<0.001$ ,  $r\approx 0.79$ ). При этом в

текстах МП эффект является очень большим ( $r \approx 0.79$ ). Снижение частоты использования *тире* во всех случаях может быть связано с функциональной асимметрией данного знака в английском и русском языках [Кауимова, 2015]. В русском языке *тире* часто выполняет функцию маркирования эллиптических конструкций, где оно замещает опущенные элементы предложения (например, глагол-связку: *Социология – наука об ...*, всё сказуемое: *В первую группу вошло 50 человек, во вторую – 30*), а также служит средством передачи различных логико-семантических отношений (противопоставления, следствия, причины и т. д.) в тех случаях, когда они не выражены союзами или другими лексическими средствами [Кауимова, 2015: 156-157]. В английском языке *тире* преимущественно ориентировано на создание риторического эффекта, управление вниманием читателя и выполняет следующие дискурсивно-организующие функции: введение парентетических вставок, дополнительного авторского комментария, создание эмфатических пауз, соединение связанных по смыслу, но синтаксически независимых частей предложения и т. д. При переводе с английского на русский переводчики сталкиваются с необходимостью передачи дискурсивных функций английского *тире* средствами русского языка. И поскольку полное функциональное соответствие между *тире* в данных языках отсутствует, при переводе происходит его замещение другими пунктуационными или лексическими средствами, выражающими те же логико-смысловые отношения (другими знаками пунктуации, союзами и т. п.). Одновременно с этим поскольку специфические для русского языка случаи эллипсиса, требующие постановки *тире*, в исходных текстах не употребляются, то и в переводных текстах *тире* встречаются реже.

Показатели частоты употребления *круглых скобок* демонстрируют принципиально различные тенденции в ручных и машинных переводах. В корпусах с ручными переводами (СОЦ-РП и МЕД-РП) наблюдается снижение частоты использования как *открывающих*, так и *закрывающих скобок*: в ручных переводах социологических статей: *открывающие* ( $\Delta = +0.19$ ,  $p < 0.05$ ,  $r \approx 0.19$ ) и *закрывающие скобки* ( $\Delta = +0.21$ ,  $p < 0.05$ ,  $r \approx 0.22$ ); в ручных переводах медицинских статей: *открывающие* ( $\Delta = +0.53$ ,  $p < 0.001$ ,  $r \approx 0.38$ ) и *закрывающие скобки* ( $\Delta = +0.52$ ,  $p < 0.001$ ,

$r \approx 0.37$ ). Снижение частоты использования *скобок* в ручных переводах отражает тенденцию к редукции парентетической сложности: круглые скобки часто используются для введения дополнительной информации, уточнений и пояснений. Снижение их частоты, вероятно, отражает стремление переводчиков сделать изложение более линейным, интегрируя представленную в скобках информацию в основной текст. Следовательно, данный результат может быть интерпретирован как проявление *упрощения*.

В корпусе МЕД-МП наблюдается противоположная тенденция – увеличение частоты использования *открывающих* ( $\Delta = -0.34$ ,  $p < 0.001$ ,  $r \approx 0.25$ ) и *закрывающих скобок* ( $\Delta = -0.31$ ,  $p < 0.05$ ,  $r \approx 0.24$ ). Такой результат, вероятно, объясняется тем, что, в отличие от человека, система МП не осуществляет интеграции заключённой в скобках информации в основной текст, перенося её в текст перевода в исходном виде.

По показателю частоты употребления *кавычек-ёлочек* наблюдается статистически значимое снижение в ручных переводах в корпусе СОЦ-РП: *открывающие* ( $\Delta = +0.56$ ,  $p < 0.001$ ,  $r \approx 0.39$ ) и *закрывающие кавычки* ( $\Delta = +0.56$ ,  $p < 0.001$ ,  $r \approx 0.39$ ). Результат может быть следствием того, что исходным англоязычным текстам использование данного знака в его русскоязычном варианте («») не свойственно. В английском языке для маркирования прямой речи, цитат используются иные типы кавычек (" "). Кроме того, прямая речь исходного текста может передаваться как косвенная в переводе. В то же время в корпусах медицинских статей статистически значимых различий по данному показателю выявлено не было, что закономерно: для медицинских статей включение фрагментов прямой речи в текст (высказывания респондентов, цитаты из интервью и т. д.) не характерно, как это характерно для социологических статей.

Показатель частоты употребления *запятой* оказался статистически значимо выше в машинных переводах в корпусе МЕД-МП ( $\Delta = -0.64$ ,  $p < 0.001$ ,  $r \approx 0.31$ ). Результат согласуется с уже упоминавшимся увеличением *средней длины предложений* и *среднего числа клауз на предложение* в подкорпусе МП. Кроме того, результат может быть следствием *интерференции*: как будет видно далее, в

МП значительно растет частота использования сложноподчинённых предложений, вводимых союзами или союзными словами.

### Индикаторы характеристики *нормализация*

Один из ключевых показателей, используемых при анализе характеристики *нормализация*, – коэффициент лексической повторяемости (*lexical repetition*), отражающий степень повторного использования одних и тех же знаменательных лексических единиц в тексте. Результаты анализа представлены в таблице 9, полные результаты статистического анализа приведены в ПРИЛОЖЕНИИ Е.

Таблица 9 – Результаты анализа *лексической повторяемости*

Lexical repetition (%)				
Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT–T)	<i>p</i>
СОЦ-РП	9.65	9.3	+0.35	*
МЕД-РП	9.49	9.02	+0.47	<i>n.s.</i>
МЕД-МП	9.49	7.73	+1.76	***

В корпусе СОЦ-РП показатель *лексической повторяемости* в переводных текстах оказался статистически значимо ниже, чем в непереводных ( $\Delta=+0.35$ ,  $p<0.05$ ,  $r\approx 0.21$ ), что подтверждает действие *нормализации*: переводчики демонстрируют тенденцию к избеганию повторного использования одних и тех же знаменательных лексем на протяжении текста. В корпусе МЕД-МП в машинных переводах также наблюдается снижение показателя *лексической повторяемости* ( $\Delta=+1.76$ ,  $p<0.001$ ,  $r\approx 0.17$ ), что обусловлено тем, что системы МП могут переводить одну и ту же лексему текста оригинала по-разному в зависимости от локального контекста. Стоит отметить, что снижение *лексической повторяемости* затрагивает такой важный аспект текстуальной связности, как лексическая когезия (см. раздел 2.3): *лексический повтор* является одним из средств её обеспечения, а снижение лексических повторов способно приводить к снижению лексической (в частности терминологической) консистентности и связности всего текста в целом, что в особенности важно учитывать при постредактировании МП медицинской литературы.

Перейдём к рассмотрению показателя *поточечной взаимной информации*, а именно двух его версий – обычной (*PMI*) и модифицированной (*Modified PMI*). Напомним, что показатель *поточечной взаимной информации* отражает степень ассоциативной связанности всех пар слов подкорпусе. Более высокие значения показателя указывают на более устойчивые коллокационные связи между биграммами слов в подкорпусе. Напомним также, что метрика обычной версии *PMI* чувствительна к редким событиям: даже если слова встречаются в подкорпусе совместно всего лишь один раз и при этом их совместная встречаемость случайна, при достаточно низких индивидуальных частотах входящих в биграмму слов всё словосочетание (т. е. биграмма) получает искусственно завышенное значение *PMI*, что может несколько исказить общую картину по данному показателю. В то время как модифицированная версия (*Modified PMI*) компенсирует данный эффект за счёт перемножения значения *PMI* биграммы на абсолютную частоту совместной встречаемости составляющих эту биграмму слов ( $PMI(x, y) \times f(x, y)$ ), тем самым отфильтровывая случайные и малорепрезентативные сочетания. Поэтому *Modified PMI* является более надёжным индикатором устойчивых коллокационных связей.

Анализ *поточечной взаимной информации* проводился при двух размерах контекстного окна  $w$ : при  $w=1$  учитывались непосредственно соседствующие друг с другом слова (собственно коллокации), при  $w=5$  – слова на расстоянии до 5 позиций друг от друга. Для каждого условия использовались три порога минимальной совместной встречаемости (*min. cooccurrence*=1, 3 и 5). Пороги 3 и 5 позволяют отфильтровать случайные сочетания: единицы с совместной встречаемостью меньше данного порога из расчета исключаются. Начнем с результатов анализа при размере окна  $w=1$ : в таблице 10 приведены среднее значение показателя *PMI / Modified PMI (Average)* и доля биграмм в подкорпусе, значение показателей *PMI / Modified PMI* которых оказалось выше 0 (*Share > 0*).

Таблица 10 – Результаты анализа *поточечной взаимной информации* (при  $w=1$ )

Корпус	<i>min. cooccurrence=1</i>				<i>min. cooccurrence=3</i>				<i>min. cooccurrence=5</i>			
	Average		Share > 0		Average		Share > 0		Average		Share > 0	
	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T
PMI												
СОЦ-РП	5.7	5.55	96%	96%	3.66	3.59	93%	93%	3.54	3.47	93%	93%
МЕД-РП	6.12	5.60	97%	96%	4.88	4.47	95%	94%	4.8	4.46	96%	95%
МЕД-МП	6.12	5.83	97%	96%	4.88	4.79	95%	95%	4.8	4.76	96%	95%
Modified PMI												
СОЦ-РП	8.32	8.30	96%	96%	27.85	27.5	93%	93%	47.4	46.4	93%	93%
МЕД-РП	10.9	10.6	97%	96%	37.11	37.4	95%	94%	61.18	62.1	96%	95%
МЕД-МП	10.9	11.4	97%	96%	37.11	39.2	95%	95%	61.18	64.6	96%	95%

Для начала рассмотрим значения показателей для обычной версии *PMI*. Во всех трёх корпусах при всех трёх порогах совместной встречаемости в переводах наблюдаются более низкие *средние значения PMI* по сравнению с непереводами. При этом *доля биграмм с положительным значением PMI* либо остаётся одинаковой, либо незначительно снижается в переводе (в пределах 1%) во всех условиях для всех корпусов, что указывает на сопоставимую долю ассоциативно связанных пар слов в переводах и непереводах. В совокупности эти результаты свидетельствуют о том, что различия между переводными и непереводами подкорпусами проявляются преимущественно в интенсивности связей между биграммами, т. е. в устойчивости ассоциативных паттернов: в переводах (ручных и машинных) при  $w=1$  биграммы, состоящие из непосредственно соседствующих слов, в целом обладают меньшей ассоциативной связанностью, чем биграммы в непереводах.

Модифицированная версия *Modified PMI* демонстрирует более неоднородную картину. В корпусе СОЦ-РП доля биграмм с *Modified PMI > 0* остается одинаковой в обоих подкорпусах при всех порогах совместной встречаемости, а значения *Modified PMI* для переводных текстов снижаются по сравнению с непереводами, что согласуется с результатами анализа обычной версии *PMI*: в ручных переводах по социологии используются менее устойчивые коллокации. В корпусе МЕД-РП доля биграмм с *Modified PMI > 0* в МП снижается на 1% при всех условиях совместной встречаемости. При этом при *min.*

*cooccurrence*=1 в переводах наблюдается снижение *среднего значения Modified PMI*, однако при увеличении порога совместной встречаемости до 3 и 5 случаев ситуация меняется: переводные тексты начинают демонстрировать более высокие средние значения показателя. Данный результат говорит о том, что в ручных медицинских переводах используется несколько меньше ассоциативно связанных пар слов, а также о том, что коллокационная структура ручных переводов отличается от непереводаемых текстов несколько меньшим разнообразием используемых коллокаций, но большей интенсивностью (устойчивостью) связей между компонентами этих коллокаций. В то время как непереводаемые тексты демонстрируют большее разнообразие ассоциативно связанных редких пар слов, но используемые коллокации в них менее шаблонны (устойчивы).

В корпусе МЕД-МП эффект выражен ещё сильнее: *Modified PMI* в машинных переводах оказался значительно выше при всех трёх порогах совместной встречаемости. Доля биграмм с *Modified PMI* > 0 при этом либо снижается на 1% (при *min. cooccurrence*=1 и 3), либо остается на том же уровне (при *min. cooccurrence*=2). Поскольку *Modified PMI* является более надёжным индикатором устойчивых коллокационных связей, чем обычный *PMI*, данный результат свидетельствует о том, что МП тяготеет к воспроизведению более предсказуемых, шаблонных словосочетаний ПЯ, избегая нестандартных и нетипичных коллокаций.

Теперь рассмотрим результаты анализа *поточечной взаимной информации* при увеличенном размере контекстного окна до 5 позиций ( $w=5$ ) (Табл. 11).

Таблица 11 – Результаты анализа *поточечной взаимной информации* (при  $w=5$ )

Корпус	min. cooccurrence=1				min. cooccurrence=3				min. cooccurrence=5			
	Average		Share > 0		Average		Share > 0		Average		Share > 0	
	NT	T	NT	T	NT	T	NT	T	NT	T	NT	T
<b>PMI</b>												
СОЦ-РП	4.19	4.11	93%	92%	1.6	1.5	81%	80%	1.26	1.18	77%	77%
МЕД-РП	4.42	3.96	93%	92%	2.44	2.04	85%	83%	2.11	1.77	82%	80%
МЕД-МП	4.42	4.22	93%	92%	2.44	2.48	85%	85%	2.11	2.16	82%	82%
<b>Modified PMI</b>												
СОЦ-РП	5.26	5.19	93%	92%	9.65	9.34	81%	80%	14.23	13.7	77%	77%
МЕД-РП	6.36	5.8	93%	92%	15.4	13.8	85%	83%	23.4	21.16	82%	80%
МЕД-МП	6.36	6.49	93%	92%	15.4	16.3	85%	85%	23.4	24.8	82%	82%

В корпусах СОЦ-РП и МЕД-РП при всех трёх условиях минимальной совместной встречаемости наблюдается снижение средних значений обеих версий показателя в ручных переводах. При этом доля биграмм с *PMI / Modified PMI* выше нуля в переводах либо несколько снижается, либо остается на том же уровне. Такая динамика свидетельствует о том, что в случае учёта более удалённых соседей непереводные тексты, как правило, содержат несколько больше ассоциативно связанных пар слов (коллокаций) и что сами связи между словами в этих текстах более устойчивы. В корпусе МЕД-МП при *min. cooccurrence=1* наблюдается снижение обычной версии *PMI* в машинных переводах. Однако при дальнейшем ужесточении порога до 3 и 5 случаев в МП фиксируются более высокие, чем в непереводных текстах, средние значения *PMI*. Это же подтверждается данными по показателю *Modified PMI*, средние значения которого при всех трёх условиях (*min. cooccurrence=1, 3 и 5*) выше в МП. Доля биграмм с *PMI / Modified PMI* выше нуля при *min. cooccurrence=3 и 5* остается одинаковой, а при *min. cooccurrence=1* снижается на 1%.

### **Индикаторы характеристики экспликация**

Перейдем к рассмотрению результатов анализа индикаторов *экспликации*. Полные результаты статистического анализа всех индикаторов *экспликации* приведены в ПРИЛОЖЕНИИ Ё. Начнем с индикаторов, связанных с употреблением имен собственных (Табл. 12).

Таблица 12 – Результаты анализа индикаторов *эксплицитного* названия, *единичного* именованного и *средней* длины именованных сущностей

Корпус	Non-translated (median)	Translated (median)	HL	<i>p</i>
Explicit naming (%)				
СОЦ-РП	187.9	300	-81.09	**
МЕД-РП	164.9	157.9	+10.1	<i>n.s.</i>
МЕД-МП	164.9	99.59	+51.87	**
Single naming (%)				
СОЦ-РП	58.82	66.67	-6.94	<i>n.s.</i>
МЕД-РП	50.00	66.67	-16.67	**
МЕД-МП	50.00	72.78	-21.46	***

Продолжение таблицы 12

Корпус	Non-translated (median)	Translated (median)	HL	<i>p</i>
Mean named entities length (tokens)				
СОЦ-РП	1.53	1.45	+0.052	<i>n.s.</i>
МЕД-РП	1.91	1.56	+0.44	***
МЕД-МП	1.911	1.562	+0.35	***

оказатель *эксплицитного называния*, отражающий отношение числа личных и притяжательных местоимений к общему числу именованных сущностей, подтверждает действие *экспликаци* только для корпуса СОЦ-РП: в переводах по социологии растёт употребление личных и притяжательных местоимений относительно имён собственных (HL=-81.09,  $p < 0.01$ ,  $r = 0.24$ ,  $A_{12} = 0.36$ ), что может свидетельствовать о стремлении переводчиков-людей прибегать к использованию местоимений для облегчения идентификации именованных сущностей. В корпусе МЕД-МП показатель *эксплицитного называния* оказался ниже для подкорпуса МП (HL=+51.87,  $p < 0.01$ ,  $r \approx 0.24$ ,  $A_{12} \approx 0.64$ ). Однако, как будет показано далее, в подкорпусе МП доля личных и притяжательных местоимений оказалась выше, чем в непереводе подкорпусе. Следовательно, можно предположить, что в исходных англоязычных статьях именованные сущности используются реже, чем в непереводе русскоязычных текстах, что и обуславливает снижение показателя *эксплицитного называния* в подкорпусе МП.

Показатель *единичного именования*, оценивающий долю однословных именованных сущностей в тексте, вопреки ожиданиям, увеличивается в ручных медицинских переводах (HL=-16.67,  $p < 0.01$ ,  $r = 0.27$ ,  $A_{12} \approx 0.34$ ) – это означает, что в них используются более короткие версии имён собственных. В машинных медицинских переводах эффект выражен еще сильнее (HL=-21.46,  $p < 0.001$ ,  $r \approx 0.47$ ,  $A_{12} \approx 0.23$ ). В обоих случаях, вероятно, результат отражает различия в использовании именованных сущностей между английскими и русскоязычными оригиналами. Показатель *средней длины именованных сущностей* (в токенах) снижается в подкорпусе ручных (HL=+0.44,  $p < 0.001$ ,  $r \approx 0.35$ ,  $A_{12} \approx 0.7$ ) и машинных

переводов по медицине ( $HL=+0.35$ ,  $p<0.001$ ,  $r\approx 0.34$ ,  $A_{12}\approx 0.7$ ), что согласуется с результатами анализа *единичного именованя*.

Перейдём к рассмотрению результатов анализа *дискурсивных маркеров глобального и локального уровня связности* научного текста. В таблице 13 приведены результаты только по тем категориям маркеров, для которых были зафиксированы статистически значимые различия между подкорпусами ( $p<0.05$ ).

Таблица 13 – Результаты анализа *дискурсивных маркеров локального и глобального уровня связности*

Корпус	Non-translated (median)	Translated (median)	HL	<i>p</i>
<b>Локальный уровень связности текста</b>				
Введение новой информации (%)				
СОЦ-РП	0.94	0.85	+0.14	**
МЕД-РП	0.85	0.98	-0.15	**
Противопоставление (%)				
СОЦ-РП	0.84	0.97	-0.12	*
МЕД-РП	0.54	0.7	-0.19	***
Введение примеров (%)				
МЕД-РП	0.07	0.16	-0.084	***
МЕД-МП	0.07	0.16	-0.073	***
Причинно-следственные отношения (%)				
МЕД-РП	0.47	0.53	-0.076	*
Целевые отношения (%)				
МЕД-РП	0.57	0.86	-0.22	***
МЕД-МП	0.57	1.36	-0.73	***
<b>Глобальный уровень связности текста</b>				
Обеспечение порядка следования информации (%)				
МЕД-МП	0.4	0.55	-0.134	*
Выводы и заключение (%)				
МЕД-МП	0.22	0.18	+0.05	*

По показателю частоты использования маркеров *введения новой информации* (маркеры типа *a также, кроме того, наряду с этим* и т. д.) в корпусах СОЦ-РП и МЕД-РП наблюдаются противоположные тенденции. В СОЦ-РП видно небольшое

снижение частоты использования маркеров данного типа в переводах ( $HL=+0.14$ ,  $p<0.01$ ,  $r=0.26$ ,  $A_{12}=0.65$ ). Этот результат может объясняться действием *интерференции* английского языка: как показывают исследования дискурсивных маркеров научного текста [Валеева, Руднева, 2016: 109-115], авторы англоязычных статей в меньшей степени полагаются на маркеры введения новой информации, в то время как для русскоязычного академического стиля характерно их более активное использование для поддержания текстовой связности. Полученный результат свидетельствует о несколько более плавном и последовательном введении новой информации в непереводных русскоязычных научных статьях по социологии в сравнении с переводными. С другой стороны, в корпусе МЕД-РП наблюдается обратная тенденция: маркеры *введения новой информации* используются в ручных переводах по медицине несколько чаще, чем в непереводных текстах ( $HL=-0.15$ ,  $p<0.01$ ,  $r\approx 0.23$ ,  $A_{12}\approx 0.37$ ), что может говорить о стремлении переводчиков к несколько более явному обозначению переходов между высказываниями.

В подкорпусах ручных переводов наблюдается рост частоты использования маркеров *противопоставления* (маркеры типа *тем не менее, в то же время, напротив, однако* и др.), а именно в ручных переводах по социологии ( $HL=-0.12$ ,  $p<0.05$ ,  $r\approx 0.21$ ,  $A_{12}=0.38$ ) и по медицине ( $HL=-0.19$ ,  $p<0.001$ ,  $r\approx 0.35$ ,  $A_{12}\approx 0.3$ ). Данный результат может объясняться двумя факторами. Во-первых, *интерференцией* исходного английского текста [Kunilovskaya, 2017], где контраст между различными положениями, как правило, выражается более эксплицитно (имеется в виду частое употребление маркеров типа *however, although, in fact*). Во-вторых, как показывают исследования [Hoek et al., 2017], отношения *противопоставления* являются когнитивно более сложными для понимания в сравнении с другими типами отношений, из-за чего они чаще выражаются посредством явного маркирования логических связей между фрагментами текста, что согласуется с гипотезой об *экспликации* [Kunilovskaya et al., 2021: 109].

Наблюдается увеличение частоты использования маркеров *введения примеров* (маркеры типа *например, к примеру* и др.) в ручных ( $HL=-0.084$ ,  $p<0.001$ ,

$r \approx 0.38$ ,  $A_{12} \approx 0.28$ ) и машинных ( $HL = -0.073$ ,  $p < 0.001$ ,  $r \approx 0.35$ ,  $A_{12} \approx 0.29$ ) медицинских переводах. Примечательно, что количественные характеристики данного показателя практически идентичны для обоих типов перевода: медианные значения обоих переводных подкорпусов совпадают, величины сдвига Ходжеса–Лемана близки, размеры эффекта сопоставимы – все это позволяет предположить, что рост частоты маркеров *введения примеров* в ручных переводах обусловлен не *экспликацией*, а *интерференцией* исходных англоязычных текстов. Если бы увеличение частоты маркеров данного типа было следствием стремления переводчика-человека к повышению эксплицитности текста перевода, то между ручными и машинными переводами наверняка наблюдались бы существенные различия в пользу ручных переводов.

Некоторый рост частоты использования маркеров *причинно-следственных отношений* (маркеры типа *потому что, поэтому, по этой причине, вследствие* и т. д.) в подкорпусе ручных медицинских переводов ( $HL = -0.076$ ,  $p < 0.05$ ,  $r \approx 0.17$ ,  $A_{12} \approx 0.4$ ) говорит в пользу гипотезы об *экспликации*, поскольку данный тип маркеров делает аргументацию более последовательной и логически выстроенной.

Наблюдается рост частоты использования маркеров *целевых отношений* (маркеры типа *для того чтобы, с целью, с тем чтобы, во избежание* и т. д.) в ручных ( $HL = -0.22$ ,  $p < 0.001$ ,  $r \approx 0.28$ ,  $A_{12} \approx 0.34$ ) и машинных ( $HL = -0.73$ ,  $p < 0.001$ ,  $r \approx 0.70$ ,  $A_{12} \approx 0.09$ ) медицинских переводах. В ручных переводах умеренный рост маркеров данного типа может объясняться как *экспликацией* целевых отношений, так и *интерференцией* английских целевых конструкций. Можно предположить, что в случае с МП столь резкий рост частоты маркеров *целевых отношений* и столь крупный размер эффекта ( $r \approx 0.70$ ) свидетельствуют о калькировании английских целевых конструкций. В результате этого в текстах МП их количество становится неестественно большим, что требует особого внимания при постредактировании МП, а именно контроля частоты их употребления и уместности их использования в качестве варианта перевода.

Что касается маркеров глобального уровня текстовой связности, то в корпусе МЕД-МП в машинных переводах наблюдается некоторое увеличение частоты

использования маркеров, обеспечивающих *порядок следования информации* (маркеры типа *во-первых, затем, сначала, наконец* и пр.) ( $HL=-0.134$ ,  $p<0.05$ ,  $r\approx 0.22$ ,  $A_{12}\approx 0.37$ ) и некоторое снижение частоты употребления маркеров *выводов и заключения* (маркеры типа *таким образом, в заключение, итак* и др.) ( $HL=+0.05$ ,  $p<0.05$ ,  $r\approx 0.21$ ,  $A_{12}\approx 0.62$ ). В случае с МП, где отсутствует намеренное вмешательство человека, такие различия между подкорпусами, вероятнее всего, отражают особенности организации англоязычных оригиналов.

Перейдём к рассмотрению индикаторов лексической когезии (Табл. 14).

Таблица 14 – Результаты анализа индикаторов *лексической когезии*

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	$p$
Word pairwise cosine				
СОЦ-РП	0.172	0.160	+0.012	***
МЕД-РП	0.174	0.175	-0.001	<i>n.s.</i>
МЕД-МП	0.174	0.170	+0.004	*
Sentence cosine (adjacent)				
СОЦ-РП	0.700	0.655	+0.045	***
МЕД-РП	0.683	0.690	-0.007	<i>n.s.</i>
МЕД-МП	0.683	0.694	-0.011	<i>n.s.</i>

Напомним, что показатель *среднего косинусного сходства пар слов* (*word pairwise cosine*) отражает степень семантической близости между всеми парами знаменательных слов в тексте, а показатель *среднего косинусного сходства соседних предложений* (*sentence cosine adjacent*) – степень семантической близости между последовательно идущими друг за другом предложениями текста.

В корпусе СОЦ-РП в переводах фиксируется снижение *среднего косинусного сходства пар слов* ( $\Delta=+0.012$ ,  $p<0.001$ ,  $r\approx 0.39$ ) и *среднего косинусного сходства соседних предложений* ( $\Delta=+0.045$ ,  $p<0.001$ ,  $r\approx 0.41$ ). Это значит, что ручные переводы по социологии в целом состоят из семантически несколько менее связанной лексики и что соседние предложения в них характеризуются менее плавными тематическими переходами. В корпусе МЕД-МП наблюдается некоторое снижение *среднего косинусного сходства между парами слов* в МП

( $\Delta=0.004$ ,  $p<0.05$ ,  $r\approx 0.24$ ), что также указывает на небольшое снижение общей семантической связности используемой лексики по сравнению с непереводаемыми русскоязычными статьями, хотя лексическая связность на уровне идущих друг за другом предложений значимо не отличается.

### Индикаторы характеристики *интерференция*

Теперь перейдем к результатам анализа частеречных последовательностей, продемонстрировавших наибольшие статистически значимые различия. Полные результаты статистического анализа приведены в ПРИЛОЖЕНИИ Ж.

Ниже приведены результаты анализа *частеречных униграмм, биграмм и триграмм* в таблицах 15-17 соответственно.

Таблица 15 – Результаты анализа *частеречных униграмм*

N-грамма	Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	<i>p</i>
NOUN	СОЦ-РП	29.44	27.48	+1.96	***
	МЕД-РП	34.83	33.07	+1.76	***
	МЕД-МП	34.83	31.89	+2.94	***
ADJF	СОЦ-РП	9.76	8.96	+0.80	**
	МЕД-МП	8.92	7.48	+1.44	***
VERB	СОЦ-РП	4.95	5.71	-0.76	***
ADVB	СОЦ-РП	1.91	2.19	-0.28	**
	МЕД-РП	1.27	1.62	-0.36	***
INFN	СОЦ-РП	1.58	1.87	-0.29	**
	МЕД-РП	0.7	1.1	-0.4	**
	МЕД-МП	0.7	0.55	+0.15	*
PRTF	МЕД-РП	1.78	2.05	-0.27	***
	МЕД-МП	1.78	2.1	-0.32	**
SCONJ	СОЦ-РП	2.02	2.61	-0.59	***
	МЕД-РП	1.09	1.54	-0.45	***
	МЕД-МП	1.09	1.46	-0.37	***
PRTS	МЕД-МП	1.01	1.14	-0.13	*
AUX	МЕД-РП	0.44	0.64	-0.20	***
	МЕД-МП	0.44	0.97	-0.53	***
COPULA	СОЦ-РП	0.16	0.26	-0.10	***
	МЕД-РП	0.15	0.22	-0.08	**
	МЕД-МП	0.15	0.33	-0.18	***

Продолжение таблицы 15

N-грамма	Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	$p$
MDL_VRB	СОЦ-РП	0.33	0.48	-0.14	***
	МЕД-РП	0.27	0.48	-0.21	***
MDL_WRD	СОЦ-РП	0.27	0.33	-0.06	*
	МЕД-РП	0.15	0.26	-0.11	**
	МЕД-МП	0.15	0.06	+0.09	***
PER_PRN	СОЦ-РП	0.6	1.02	-0.42	***
	МЕД-РП	0.14	0.29	-0.15	***
	МЕД-МП	0.14	0.43	-0.29	***
DMSTR_PRN	СОЦ-РП	2.13	2.5	-0.37	***
	МЕД-РП	1.23	1.72	-0.49	***
PRN_CONJ_W	СОЦ-РП	0.73	0.93	0.20	***
	МЕД-РП	0.27	0.46	-0.19	***
	МЕД-МП	0.27	0.53	-0.26	***
DF_ADJ_N_PRN	МЕД-РП	0.4	0.6	-0.2	***
	МЕД-МП	0.4	0.55	-0.15	***
INDF_ADJ_PRN	СОЦ-РП	0.11	0.18	-0.07	***
	МЕД-РП	0.05	0.09	-0.04	***
INDF_NUM_PRN	СОЦ-РП	0.04	0.06	-0.02	**
SBJ_PRCL	СОЦ-РП	0.07	0.15	-0.08	***
	МЕД-РП	0.01	0.03	-0.02	***
RFLX_PRN	МЕД-РП	0.04	0.06	-0.02	*
SPRL	СОЦ-РП	0.09	0.13	-0.04	**

Таблица 16 – Результаты анализа частеречных биграмм

N-грамма	Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	$p$
NOUN NOUN	СОЦ-РП	5.80	4.93	+0.87	***
	МЕД-РП	9.51	8.94	+0.57	*
	МЕД-МП	9.51	7.81	+1.71	***
NOUN ADJF	СОЦ-РП	2.33	1.86	+0.47	***
	МЕД-РП	2.47	2.15	+0.32	**
	МЕД-МП	2.47	1.40	+1.07	***
ADJF NOUN	МЕД-МП	7.38	6.14	+1.24	***
COMMA SCONJ	СОЦ-РП	1.21	1.74	-0.53	***
	МЕД-РП	0.79	1.04	-0.25	***
	МЕД-МП	0.79	0.96	-0.17	**

Продолжение таблицы 16

N-грамма	Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	$p$
PER_PRN VERB	СОЦ-РП	0.29	0.54	-0.25	***
	МЕД-РП	0.05	0.15	-0.1	***
	МЕД-МП	0.05	0.31	-0.26	***
COMMA PRN_CONJ_W	СОЦ-РП	0.44	0.62	-0.18	***
	МЕД-РП	0.14	0.30	-0.16	***
	МЕД-МП	0.14	0.38	-0.24	***
ADVB VERB	СОЦ-РП	0.32	0.43	-0.11	***
MDL_VRB INFN	СОЦ-РП	0.22	0.32	-0.10	***
	МЕД-РП	0.19	0.32	-0.13	***
S_START PER_PRN	СОЦ-РП	0.07	0.17	-0.10	***
COMP_MOD ADJF	СОЦ-РП	0.11	0.18	-0.07	***
DMSTR_PRN NOUN	МЕД-РП	0.60	0.79	-0.19	***
PREP DMSTR_PRN	МЕД-РП	0.47	0.64	-0.17	***
AUX PRTS	МЕД-РП	0.37	0.51	-0.15	***
	МЕД-МП	0.37	0.79	-0.42	***
COMMA PRTF	МЕД-РП	0.68	0.81	-0.13	*
	МЕД-МП	0.68	0.94	-0.25	***

Таблица 17 – Результаты анализа *частеречных триграмм*

N-грамма	Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	$p$
NOUN ADJF NOUN	СОЦ-РП	1.94	1.59	+0.35	**
	МЕД-РП	2.13	1.86	+0.27	**
	МЕД-МП	2.13	1.20	+0.93	***
NOUN COMMA NOUN	СОЦ-РП	0.94	0.63	+0.32	***
	МЕД-РП	1.01	0.62	+0.39	***
	МЕД-МП	1.01	0.71	+0.30	***
NOUN NOUN NOUN	СОЦ-РП	1.03	0.79	+0.24	**
	МЕД-МП	2.32	1.61	+0.71	***
ADJF NOUN NOUN	СОЦ-РП	1.22	0.99	+0.23	***
	МЕД-РП	1.91	1.72	+0.19	*
	МЕД-МП	1.91	1.51	+0.40	***
NOUN NOUN ADJF	СОЦ-РП	0.41	0.26	+0.15	***
	МЕД-РП	0.56	0.42	+0.14	**
	МЕД-МП	0.56	0.24	+0.32	***

## Продолжение таблицы 17

N-грамма	Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	<i>p</i>
NOUN COMMA SCONJ	СОЦ-ПП	0.40	0.53	-0.13	***
	МЕД-ПП	0.28	0.37	-0.09	***
	МЕД-МП	0.28	0.35	-0.07	**
NOUN COMMA PRN_CONJ_W	СОЦ-ПП	0.33	0.48	-0.15	***
	МЕД-ПП	0.11	0.26	-0.15	***
	МЕД-МП	0.11	0.28	-0.17	***
VERB COMMA SCONJ	СОЦ-ПП	0.20	0.32	-0.12	***
	МЕД-МП	0.12	0.20	-0.08	***
COMMA SCONJ NOUN	СОЦ-ПП	0.30	0.45	-0.15	***
	МЕД-ПП	0.25	0.32	-0.07	*
DMSTR_PRN COMMA SCONJ	СОЦ-ПП	0.24	0.37	-0.13	***
COMMA PRN_CONJ_W VERB	СОЦ-ПП	0.19	0.28	-0.09	***
	МЕД-ПП	0.07	0.15	-0.08	***
	МЕД-МП	0.07	0.19	-0.12	***
NOUN MDL_VRB INFN	МЕД-ПП	0.09	0.16	-0.07	**
NOUN COMMA PRTF	МЕД-ПП	0.61	0.74	-0.13	**
	МЕД-МП	0.61	0.84	-0.23	***
PREP DMSTR_PRN NOUN	МЕД-ПП	0.22	0.30	-0.08	***
PREP DMSTR_PRN COMMA	МЕД-ПП	0.11	0.19	-0.08	***
S_START PER_PRN VERB	МЕД-МП	0.01	0.10	-0.09	***
PER_PRN VERB NOUN	МЕД-МП	0.01	0.10	-0.09	***
NOUN AUX PRTS	МЕД-МП	0.19	0.37	-0.18	***
COMMA PRTF PREP	МЕД-ПП	0.26	0.38	-0.12	***
	МЕД-МП	0.26	0.45	-0.19	***
AUX PRTS PREP	МЕД-МП	0.11	0.43	-0.32	***

Общей тенденцией для всех трёх переводных подкорпусов является снижение доли имён существительных и разнообразных комбинаций с ними. Так, в корпусе СОЦ-ПП в ручных переводах наблюдается следующая картина:

- *NOUN* ( $\Delta=+1.96$ ,  $p<0.001$ ,  $r\approx 0.41$ );
- *NOUN NOUN* ( $\Delta=+0.87$ ,  $p<0.001$ ,  $r\approx 0.38$ );
- *NOUN ADJF* ( $\Delta=+0.47$ ,  $p<0.001$ ,  $r\approx 0.35$ );
- *NOUN ADJF NOUN* ( $\Delta=+0.35$ ,  $p<0.001$ ,  $r\approx 0.30$ );
- *NOUN NOUN NOUN* ( $\Delta=+0.24$ ,  $p<0.001$ ,  $r\approx 0.32$ );

- *ADJF NOUN NOUN* ( $\Delta=+0.23$ ,  $p<0.001$ ,  $r\approx 0.42$ );
- *NOUN NOUN ADJF* ( $\Delta=+0.15$ ,  $p<0.001$ ,  $r\approx 0.45$ ).

В корпусе МЕД-РП в ручных переводах медицинских статей динамика схожа:

- *NOUN* ( $\Delta=+1.76$ ,  $p<0.001$ ,  $r\approx 0.43$ );
- *NOUN NOUN* ( $\Delta=+0.57$ ,  $p<0.05$ ,  $r\approx 0.20$ );
- *NOUN ADJF* ( $\Delta=+0.32$ ,  $p<0.01$ ,  $r\approx 0.24$ );
- *NOUN ADJF NOUN* ( $\Delta=+0.27$ ,  $p<0.01$ ,  $r\approx 0.24$ );
- *ADJF NOUN NOUN* ( $\Delta=+0.19$ ,  $p<0.05$ ,  $r\approx 0.23$ );
- *NOUN NOUN ADJF* ( $\Delta=+0.14$ ,  $p<0.01$ ,  $r\approx 0.30$ ).

В корпусе МЕД-МП в машинных переводах медицинских статей ситуация аналогична, а величина эффекта еще более выражена:

- *NOUN* ( $\Delta=2.94$ ,  $p<0.001$ ,  $r\approx 0.65$ );
- *NOUN NOUN* ( $\Delta=+1.71$ ,  $p<0.001$ ,  $r\approx 0.59$ );
- *ADJF NOUN* ( $\Delta=+1.24$ ,  $p<0.001$ ,  $r\approx 0.43$ );
- *NOUN ADJF* ( $\Delta=+1.07$ ,  $p<0.001$ ,  $r\approx 0.69$ );
- *NOUN ADJF NOUN* ( $\Delta=+0.93$ ,  $p<0.001$ ,  $r\approx 0.69$ );
- *NOUN NOUN NOUN* ( $\Delta=+0.71$ ,  $p<0.001$ ,  $r\approx 0.56$ );
- *ADJF NOUN NOUN* ( $\Delta=+0.40$ ,  $p<0.001$ ,  $r\approx 0.47$ );
- *NOUN NOUN ADJF* ( $\Delta=+0.32$ ,  $p<0.001$ ,  $r\approx 0.66$ ).

Снижение частоты приведенных комбинаций свидетельствует о значительном снижении степени номинативности переводных текстов (в особенности текстов МП), что, вероятно, является следствием *интерференции* английского языка: несмотря на то что научные тексты в обоих языках отличаются большей степенью номинализации (субстантивации) в сравнении с текстами других функциональных стилей, в англоязычном научном дискурсе номинализация выражена в меньшей степени (см. раздел 1.4). Соответственно, снижение номинативности переводных текстов говорит о нарушении номинативного характера научного стиля русского языка и требует особого учета при переводе и постредактировании.

При этом в корпусе СОЦ-ПП в подкорпусе ручных переводов тенденция к снижению номинативности сопровождается ростом доли глаголов (*VERB*:  $\Delta=-0.76$ ,  $p<0.001$ ,  $r\approx 0.49$ ) и инфинитивов (*INFN*:  $\Delta=-0.29$ ,  $p<0.01$ ,  $r\approx 0.29$ ), что свидетельствует об увеличении степени глагольности переводных статей по социологии. Кроме того, в них наблюдается рост доли наречий (*ADVB*:  $\Delta=-0.28$ ,  $p<0.01$ ,  $r\approx 0.28$ ) и комбинаций наречий с глаголами (*ADVB VERB*:  $\Delta=-0.11$ ,  $p<0.001$ ,  $r\approx 0.39$ ), сопровождающийся уменьшением доли полных имен прилагательных (*ADJF*:  $\Delta=+0.80$ ,  $p<0.01$ ,  $r\approx 0.26$ ) и доли приведенных выше комбинаций прилагательных с существительными. Данная тенденция объясняется тем, что прилагательные функционируют в качестве определений при именах существительных, тогда как наречия часто выступают в роли обстоятельств при глагольных сказуемых. Снижение номинативности переводов, в свою очередь, закономерно влечёт за собой перераспределение зависимых от неё атрибутивных и обстоятельственных отношений. При этом в медицинском подкорпусе ручных переводов так же, как и в подкорпусе ручных переводов по социологии, наблюдается рост доли наречий (*ADVB*:  $\Delta=-0.36$ ,  $p<0.001$ ,  $r\approx 0.38$ ) и инфинитивов (*INFN*:  $\Delta=-0.40$ ,  $p=0.001$ ,  $r\approx 0.37$ ), хотя различий в использовании глаголов и имен прилагательных между переводными и непереводными медицинскими статьями выявлено не было. В медицинском подкорпусе машинных переводов, в свою очередь, наблюдается значительное снижение доли имен прилагательных (*ADJF*:  $\Delta=+1.44$ ,  $p<0.01$ ,  $r\approx 0.42$ ), в то время как значимых различий в использовании личных форм глаголов и наречий не выявлено. При этом доля инфинитивов снижается (*INFN*:  $\Delta=+0.15$ ,  $p<0.01$ ,  $r\approx 0.21$ ), что контрастирует с тенденциями, выявленными в подкорпусах ручных переводов.

В ручных переводах социологических статей наблюдается увеличение доли модальных глаголов (*MDL\_VRB*:  $\Delta=-0.14$ ,  $p<0.001$ ,  $r\approx 0.40$ ) и их комбинаций с инфинитивом (*MDL\_VRB INF*:  $\Delta=-0.10$ ,  $p<0.001$ ,  $r\approx 0.39$ ). Аналогичная тенденция наблюдается и в ручных переводах медицинских статей: рост доли модальных глаголов (*MDL\_VRB*:  $\Delta=-0.21$ ,  $p<0.001$ ,  $r\approx 0.44$ ), их комбинаций с инфинитивом (*MDL\_VRB INFN*:  $\Delta=-0.13$ ,  $p<0.001$ ,  $r\approx 0.37$ ), а также комбинаций *NOUN MDL\_VRB*

*INFN* ( $\Delta=-0.07$ ,  $p<0.01$ ,  $r\approx 0.33$ ). Помимо этого, в обоих подкорпусах ручных переводов фиксируется некоторый рост доли модальных слов: в переводных статьях по социологии (*MDL\_WRD*:  $\Delta=-0.06$ ,  $p<0.05$ ,  $r\approx 0.25$ ) и по медицине (*MDL\_WRD*:  $\Delta=-0.11$ ,  $p<0.01$ ,  $r\approx 0.33$ ). Вероятнее всего, данный результат является следствием *интерференции* модальных синтаксических структур английского языка, в котором конструкции с модальными глаголами являются наиболее продуктивным средством выражения деонтической и эпистемической модальности (конструкции типа *must be noted, should be considered; may require, might indicate* и т. д.) [Рецкер, 2016: 168-169] и, по всей видимости, используются более активно и регулярно, чем соответствующие им средства выражения модальности в русском языке (ср. [Ozyumenko, 2012: 37-44]). Кроме того, как показывают исследования [Бутенко, Авагян, 2021: 46-55], английские модальные конструкции при переводе на русский язык часто передаются не только при помощи глаголов, но и посредством модальных слов (слова типа *необходимо, нужно, можно, нельзя* и т. д.), являющихся в русском языке наиболее частым средством выражения модальности [Рецкер, 2016: 168-169]. Все это обуславливает рост доли модальных глаголов и модальных слов, а также комбинаций с ними в ручных переводах социологических и медицинских научных статей. Интересно при этом, что в корпусе МЕД-МП значимых различий в использовании модальных глаголов и комбинаций с ними выявлено не было, в то время как модальные слова, напротив, в МП используются реже, чем в непереводах по медицине (*MDL\_WR*:  $\Delta=+0.09$ ,  $p<0.001$ ,  $r\approx 0.48$ ).

Во всех трех переводных подкорпусах фиксируется рост доли глаголов-связок: в ручных переводах социологических (*COPULA*:  $\Delta=-0.10$ ,  $p<0.001$ ,  $r\approx 0.36$ ) и медицинских статей (*COPULA*:  $\Delta=-0.08$ ,  $p<0.01$ ,  $r\approx 0.33$ ) и в особенности в МП медицинских статей (*COPULA*:  $\Delta=-0.18$ ,  $p<0.001$ ,  $r\approx 0.60$ ). Вероятно, увеличение доли глаголов-связок, используемых при формировании составных именных сказуемых, обусловлено калькированием английских конструкций с глаголом *to be*: в английском языке глагол-связка в составе составного именного сказуемого обязательна, и, по всей видимости, она нередко переносится в текст перевода даже

в случае, когда может быть опущена (Ср.: *the method is effective* → *метод является эффективным* вместо *метод эффективен*).

В корпусах СОЦ-РП и МЕД-РП наблюдается увеличение доли неопределённых местоименных прилагательных (слова типа *некоторый, какой-либо* и т. д.) в ручных переводах социологических (*INDF\_ADJ\_PRN*:  $\Delta=-0.07$ ,  $p<0.001$ ,  $r\approx 0.41$ ) и медицинских статей (*INDF\_ADJ\_PRN*:  $\Delta=-0.04$ ,  $p<0.001$ ,  $r\approx 0.33$ ). Кроме того, в ручных переводах по социологии также наблюдается некоторый рост неопределённых местоименных числительных (слова типа *несколько*) (*INDF\_NUM\_PRN*:  $\Delta=-0.02$ ,  $p<0.01$ ,  $r\approx 0.27$ ), а в ручных переводах по медицине – рост доли определительных местоимений (слова типа *каждый, все, любой* и т. д.) (*DF\_ADJ\_N\_PRN*:  $\Delta=-0.20$ ,  $p<0.001$ ,  $r\approx 0.46$ ). В свою очередь, в машинных переводах медицинских статей так же, как и в ручных, фиксируется рост доли определительных местоимений т. д. (*DF\_ADJ\_N\_PRN*:  $\Delta=-0.15$ ,  $p<0.001$ ,  $r\approx 0.41$ ). По всей видимости, наблюдаемая тенденция к более активному использованию перечисленных разрядов местоимений в переводных подкорпусах может быть следствием *интерференции*: в английском языке с его обязательным требованием к маркированию категории определенности / неопределенности часто используются местоимения типа *some, any, all, each, every* и т. д., которые могут переноситься в текст перевода в виде соответствующих им неопределённых и определительных местоимений в русского языка.

Помимо этого, ручных переводах медицинских статей наблюдается некоторое увеличение частоты использования возвратного местоимения (*себя*) (*RFLX\_PRN*:  $\Delta=-0.02$ ,  $p<0.05$ ,  $r\approx 0.20$ ), что может быть вызвано калькированием возвратных местоимений, используемых в исходном английском тексте.

В корпусах МЕД-РП и МЕД-МП фиксируется увеличение доли вспомогательных глаголов в ручных (*AUX*:  $\Delta=-0.20$ ,  $p<0.001$ ,  $r\approx 0.36$ ) и в особенности машинных переводах медицинских статей (*AUX*:  $\Delta=-0.53$ ,  $p<0.001$ ,  $r\approx 0.74$ ). Одновременно с этим в подкорпусе МП наблюдается рост доли кратких причастий (*PRTS*:  $\Delta=0.13$ ,  $p<0.05$ ,  $r\approx 0.20$ ) и их комбинаций со вспомогательным глаголом:

- *AUX PRTS* ( $\Delta=-0.42$ ,  $p<0.001$ ,  $r\approx 0.70$ );
- *AUX PRTS PREP* ( $\Delta=-0.32$ ,  $p<0.001$ ,  $r\approx 0.85$ );
- *NOUN AUX PRTS* ( $\Delta=-0.18$ ,  $p<0.001$ ,  $r\approx 0.56$ );

В то же время в подкорпусе ручных переводов по медицине наблюдается увеличение доли комбинации вспомогательного глагола и краткого причастия (*AUX PRTS*:  $\Delta=-0.15$ ,  $p<0.001$ ,  $r\approx 0.31$ ). Полученные результаты свидетельствуют о систематическом калькировании английских аналитических пассивных конструкций в текст перевода. В целом частое использование пассивных конструкций в медицинских текстах обусловлено жанрово-стилистическими конвенциями: как отмечалось в разделе 1.4, страдательный залог является одним из ключевых средств реализации объективно-нейтральной манеры изложения в научном тексте, позволяя сместить фокус с личности исследователя на объект и результаты исследования, а также подчеркнуть универсальность и воспроизводимость научных наблюдений. Медицинский дискурс с его требованиями точности, объективности и верифицируемости описываемых в исследовании данных активно использует данное средство. При переводе на русский язык аналитические пассивные конструкции английского оригинала (типа *was conducted*, *were measured*) могут переноситься в текст перевода в виде соответствующих по структуре страдательных конструкций ПЯ (*было проведено*, *были измерены*), в то время как допустимым и не имеющим аналогов в английском языке средством передачи того же страдательного значения являются глаголы с постфиксом *-ся* (*проводилось*, *измерялись*). При этом, как видно из результатов анализа, в машинных переводах калькирование страдательных конструкций выражено в гораздо большей степени, чем в ручных.

Примечательно, что во всех трёх корпусах в переводах фиксируется значительное увеличение доли подчинительных союзов: в ручных переводах социологических (*SCONJ*:  $\Delta=-0.59$ ,  $p<0.001$ ,  $r\approx 0.56$ ) и медицинских статей (*SCONJ*:  $\Delta=-0.45$ ,  $p<0.001$ ,  $r\approx 0.45$ ), а также в МП медицинских статей (*SCONJ*:  $\Delta=-0.37$ ,  $p<0.001$ ,  $r\approx 0.43$ ).

В подкорпусе ручных социологических переводов при этом также фиксируется увеличение доли следующих последовательностей:

- *COMMA SCONJ* ( $\Delta=-0.53$ ,  $p<0.001$ ,  $r\approx 0.63$ );
- *COMMA SCONJ NOUN* ( $\Delta=-0.15$ ,  $p<0.001$ ,  $r\approx 0.57$ );
- *NOUN COMMA SCONJ* ( $\Delta=-0.13$ ,  $p<0.001$ ,  $r\approx 0.47$ );
- *DMSTR\_PRN COMMA SCONJ* ( $\Delta=-0.13$ ,  $p<0.001$ ,  $r\approx 0.44$ );
- *VERB COMMA SCONJ* ( $\Delta=-0.12$ ,  $p<0.001$ ,  $r\approx 0.45$ ).

В подкорпусе ручных медицинских переводов наблюдается похожая тенденция:

- *COMMA SCONJ* ( $\Delta=-0.25$ ,  $p<0.001$ ,  $r\approx 0.35$ );
- *NOUN COMMA SCONJ* ( $\Delta=-0.09$ ,  $p<0.001$ ,  $r\approx 0.32$ );
- *COMMA SCONJ NOUN* ( $\Delta=-0.07$ ,  $p<0.05$ ,  $r\approx 0.23$ ).

Аналогичная тенденция фиксируется и в подкорпусе машинных медицинских переводов:

- *COMMA SCONJ* ( $\Delta=-0.17$ ,  $p<0.01$ ,  $r\approx 0.28$ );
- *VERB COMMA SCONJ* ( $\Delta=-0.08$ ,  $p<0.001$ ,  $r\approx 0.41$ );
- *NOUN COMMA SCONJ* ( $\Delta=-0.07$ ,  $p<0.01$ ,  $r\approx 0.29$ ).

Так, в случае с ручными переводами рост доли подчинительных союзов и комбинаций с ними может интерпретироваться двояко: с одной стороны, как проявление *экспликации* логико-смысловых отношений между частями высказываний, а с другой – как *интерференция* синтаксических структур англоязычных оригиналов. Попытаться разграничить вклад этих двух проявлений феномена переводческого языка можно косвенно через сопоставление результатов анализа в подкорпусах ручных и машинных переводов. Поскольку в случае с машинным переводом вряд ли можно говорить об осознанном или неосознанном стремлении системы МП к *экспликации* содержащихся имплицитно в исходном тексте логико-смысловых отношений, рост доли подчинительных союзов в нем следует рассматривать как проявление *интерференции* – воспроизведения системой МП сложноподчиненных структур исходного текста. В подкорпусах ручных переводов, в свою очередь, помимо *интерференции*, вклад в полученный

результат частично способна вносить так же и *экспликация* как результат осознанного или бессознательного стремления переводчиков-людей к повышению степени эксплицитности переводного текста. На это косвенно указывает и тот факт, что доля подчинительных союзов в ручных переводах по сравнению с непереводными увеличивается в большей степени, чем в машинном переводе, кроме того, величина эффекта  $r$  в ручных переводах также выражена в несколько большей мере ( $\Delta=-0.59$ ,  $r\approx 0.56$  в СОЦ-РП vs  $\Delta=-0.45$ ,  $r\approx 0.45$  в МЕД-РП vs  $\Delta=-0.37$ ,  $r\approx 0.43$  в МЕД-МП). Вместе с тем следует подчеркнуть, что предложенное разграничение наблюдаемых проявления переводческого языка носит гипотетический характер. Окончательные выводы о соотношении *интерференции* и *экспликации* требуют анализа параллельного корпуса, который позволил бы сопоставить переводные тексты с их англоязычными оригиналами и установить, в какой мере рост доли подчинительных союзов обусловлен переносом структур исходного текста, а в какой – *экспликацией* логико-смысловых отношений.

Во всех трех переводных подкорпусах фиксируется рост доли местоименных союзных слов (слова типа *который*, *чей*): в ручных переводах социологических ( $PRN\_CONJ\_W$ :  $\Delta=-0.20$ ,  $p<0.001$ ,  $r\approx 0.35$ ) и медицинских статей ( $PRN\_CONJ\_W$ :  $\Delta=-0.19$ ,  $p<0.001$ ,  $r\approx 0.50$ ), а также в МП медицинских статей ( $PRN\_CONJ\_W$ :  $\Delta=-0.26$ ,  $p<0.001$ ,  $r\approx 0.59$ ). В социологическом переводном подкорпусе также растет доля различных частеречных комбинаций с местоименными союзными словами:

- $COMMA\ PRN\_CONJ\_W$  ( $\Delta=-0.18$ ,  $p<0.001$ ,  $r\approx 0.43$ );
- $NOUN\ COMMA\ PRN\_CONJ\_W$  ( $\Delta=-0.15$ ,  $p<0.001$ ,  $r\approx 0.40$ );
- $COMMA\ PRN\_CONJ\_W\ VERB$  ( $\Delta=-0.09$ ,  $p<0.001$ ,  $r\approx 0.41$ ).

В подкорпусе ручных медицинских переводов наблюдается схожая тенденция:

- $COMMA\ PRN\_CONJ\_W$  ( $\Delta=-0.16$ ,  $p<0.001$ ,  $r\approx 0.55$ );
- $NOUN\ COMMA\ PRN\_CONJ\_W$  ( $\Delta=-0.15$ ,  $p<0.001$ ,  $r\approx 0.56$ );
- $COMMA\ PRN\_CONJ\_W\ VERB$  ( $\Delta=-0.08$ ,  $p<0.001$ ,  $r\approx 0.43$ );

Аналогичная тенденция фиксируется и в подкорпусе машинных переводов по медицине:

- *COMMA PRN CONJ W* ( $\Delta=-0.24$ ,  $p<0.001$ ,  $r\approx 0.71$ );
- *NOUN COMMA PRN CONJ W* ( $\Delta=-0.17$ ,  $p<0.001$ ,  $r\approx 0.67$ );
- *COMMA PRN CONJ W VERB* ( $\Delta=-0.12$ ,  $p<0.001$ ,  $r\approx 0.60$ ).

Из приведенных данных видно, что наблюдаемое увеличение доли местоименных союзных слов и комбинаций с ними в машинном переводе выражено в большей степени, чем в ручных переводах. И поскольку системы МП, как уже говорилось, едва ли обладают стремлением к *экспликации* имплицитно содержащейся в исходном тексте информации, данный результат свидетельствует об *интерференции*, а именно о переносе структуры английских придаточных определительных и ограничительных конструкций (Defining и Non-Defining Relative Clauses) в текст машинного перевода. Что касается ручных переводов, то здесь, как и в случае с увеличением доли подчинительных союзов и комбинаций с ними, по всей видимости, можно говорить как об *интерференции*, так и об *экспликации*, что тем не менее так же требует дополнительной верификации на материале параллельных англоязычных текстов.

Во всех трех переводных подкорпусах наблюдается увеличение доли личных местоимений: в ручных переводах социологических (*PER\_PRN*:  $\Delta=-0.42$ ,  $p<0.001$ ,  $r\approx 0.57$ ) и медицинских статей (*PER\_PRN*:  $\Delta=-0.15$ ,  $p<0.001$ ,  $r\approx 0.45$ ), а также в машинных переводах медицинских статей (*PER\_PRN*:  $\Delta=-0.29$ ,  $p<0.001$ ,  $r\approx 0.72$ ). Кроме того, во всех трех переводных подкорпусах фиксируется рост доли комбинаций личного местоимения с личной формой глагола: в подкорпусе ручных социологических переводов (*PER\_PRN VERB*:  $\Delta=-0.25$ ,  $p<0.001$ ,  $r\approx 0.58$ ), в подкорпусах ручных (*PER\_PRN VERB*:  $\Delta=-0.10$ ,  $p<0.001$ ,  $r\approx 0.48$ ) и машинных медицинских переводов (*PER\_PRN VERB*:  $\Delta=-0.26$ ,  $p<0.001$ ,  $r\approx 0.77$ ). Помимо этого, в подкорпусе социологических ручных переводов фиксируется увеличение доли личных местоимений в начале предложений (*S\_START PER\_PRN*:  $\Delta=-0.10$ ,  $p<0.001$ ,  $r\approx 0.56$ ), а в подкорпусе медицинских машинных переводов – рост последовательностей *S\_START PER\_PRN VERB* ( $\Delta=-0.09$ ,  $p<0.001$ ,  $r\approx 0.59$ ) и *PER\_PRN VERB NOUN* ( $\Delta=-0.09$ ,  $p<0.001$ ,  $r\approx 0.78$ ). Увеличение в переводах доли личных местоимений, местоимений в начале предложения (в функции

подлежащего) и комбинаций личного местоимения со следующей за ним личной формой глагола (в функции подлежащего и сказуемого), вероятно, обусловлено влиянием англоязычных конвенций научного письма, где использование личных местоимений при выражении личности исследователя более приемлемо и в ряде случаев более предпочтительно, чем в русскоязычной научной традиции (См. раздел 1.4). Это приводит к переносу в текст перевода исходной структуры с личным местоимением и личной формой глагола, в то время как допустимыми и более соответствующими нормам русского научного стиля являются формы страдательного залога или глаголы на постфикс *-ся* в страдательном значении (Ср.: *We found / can see a clear correlation between the variables* → *Мы обнаружили / можем наблюдать чёткую корреляцию между переменными* (полный перенос структуры) вместо, например, *Выявлена чёткая корреляция между переменными* или *Наблюдается чёткая корреляция между переменными*).

В подкорпусе ручных переводов по социологии наблюдается существенный рост доли указательных местоимений (*DMSTR\_PRN*:  $\Delta=-0.37$ ,  $p<0.001$ ,  $r\approx 0.40$ ). В подкорпусе ручных переводов по медицине наблюдается еще больший рост доли указательных местоимений (*DMSTR\_PRN*:  $\Delta=-0.49$ ,  $p<0.001$ ,  $r\approx 0.47$ ), а также рост разнообразных частеречных последовательностей с ними:

- *DMSTR\_PRN NOUN* ( $\Delta=-0.19$ ,  $p<0.001$ ,  $r\approx 0.35$ );
- *PREP DMSTR\_PRN* ( $\Delta=-0.17$ ,  $p<0.001$ ,  $r\approx 0.42$ );
- *PREP DMSTR\_PRN NOUN* ( $\Delta=-0.08$ ,  $p<0.001$ ,  $r\approx 0.37$ );
- *PREP DMSTR\_PRN COMMA* ( $\Delta=-0.08$ ,  $p<0.001$ ,  $r\approx 0.43$ ).

Примечательно, что в подкорпусе МП аналогичная тенденция не прослеживается. Рост доли указательных местоимений и последовательностей с ними в обоих подкорпусах ручных переводов при одновременном отсутствии данной тенденции в подкорпусе МП свидетельствует о том, что, вероятнее всего, данный результат является проявлением *экспликации*: переводчики-люди стремятся к более явному маркированию референциальных и логико-смысловых связей между элементами текста с помощью указательных местоимений (*этот, данный, такой* и т. д.). Более выраженный рост доли указанных

последовательностей в медицинском подкорпусе по сравнению с социологическим ( $\Delta=-0.49$  vs.  $\Delta=-0.37$ ) может объясняться спецификой медицинского дискурса, в котором требование к точности изложения данных исследования особенно важно: медицинский научный текст содержит информацию, от которой зависит здоровье пациентов, и неверная интерпретация представленных в таком тексте данных может повлечь серьёзные последствия – все это обуславливает повышенное стремление переводчиков к эксплицитному маркированию того, к какому именно предмету – показателю, методу, группе пациентов и т. д. – относится то или иное высказывание. Указательные местоимения выступают в этом случае надёжным средством обеспечения референциальной однозначности, минимизирующим риск ошибочного понимания текста.

В подкорпусе ручных медицинских переводов фиксируется увеличение доли полных причастий *PRTF* ( $\Delta=-0.27$ ,  $p<0.001$ ,  $r\approx 0.28$ ) и комбинаций с ними:

- *COMMA PRTF* ( $\Delta=-0.13$ ,  $p<0.05$ ,  $r\approx 0.23$ );
- *NOUN COMMA PRTF* ( $\Delta=-0.13$ ,  $p<0.01$ ,  $r\approx 0.26$ );
- *COMMA PRTF PREP* ( $\Delta=-0.12$ ,  $p<0.001$ ,  $r\approx 0.35$ );

В подкорпусе МП научных статей по медицине ситуация аналогична: растёт доли полных причастий (*PRTF*:  $\Delta=-0.32$ ,  $p<0.01$ ,  $r\approx 0.31$ ) и следующих комбинаций:

- *COMMA PRTF* ( $\Delta=-0.25$ ,  $p<0.001$ ,  $r\approx 0.42$ );
- *COMMA PRTF PREP* ( $\Delta=-0.19$ ,  $p<0.001$ ,  $r\approx 0.53$ );
- *NOUN COMMA PRTF* ( $\Delta=-0.23$ ,  $p<0.001$ ,  $r\approx 0.42$ ).

В совокупности рост доли перечисленных последовательностей указывает на более частое употребление в переводных медицинских подкорпусах обособленных причастных оборотов. Данный результат, вероятнее всего, является следствием *интерференции* и обусловлен невозможностью использования в английском языке номинативных рамочных конструкций (см. раздел 1.4.), строящихся по схеме [*причастие в функции определения* + (*предложная конструкция*)] + *имя существительное*. Так, например, в английском языке невозможны конструкции типа *определённые на предыдущем этапе методологические ограничения* – \*

*defined at the previous step methodological limitations* (правильный вариант: *methodological limitations defined at the previous step*).

Во всех трёх переводных подкорпусах наблюдается тенденция к снижению доли последовательности *NOUN COMMA NOUN*, то есть к снижению частоты использования перечислений имен существительных через запятую: в ручных переводах социологических ( $\Delta=+0.32$ ,  $p<0.001$ ,  $r\approx 0.44$ ) и медицинских статей ( $\Delta=+0.39$ ,  $p<0.001$ ,  $r\approx 0.50$ ) и в МП медицинских статей ( $\Delta=+0.30$ ,  $p<0.001$ ,  $r\approx 0.37$ ).

В обоих подкорпусах ручных переводов фиксируется рост частоты употребления сослагательной частицы *бы*, а именно в ручных переводах по социологии (*SBJ\_PRCL*:  $\Delta=-0.08$ ,  $p<0.001$ ,  $r\approx 0.46$ ) и по медицине (*SBJ\_PRCL*:  $\Delta=-0.02$ ,  $p<0.001$ ,  $r\approx 0.36$ ). Данный результат может быть обусловлен *интерференцией* английских конструкций с модальными глаголами *would* и *could* в эпистемическом значении, которые широко используются в англоязычном научном дискурсе в качестве средств хеджирования [Salager-Meyer, 1994: 149-170; Nyland, 1998: 106-107]. При переводе на русский язык подобные конструкции часто передаются сослагательным наклонением (*this would suggest* → *это могло бы свидетельствовать о*; *could be attributed to* → *можно было бы объяснить*; *would indicate* → *это указывало бы на*), что, вероятно, и обуславливает рост частоты использования частицы *бы* в ручных переводах.

В социологическом корпусе в ручных переводах фиксируется рост частоты использования синтетических форм превосходной степени имени прилагательного (*SPRL*:  $\Delta=-0.04$ ,  $p<0.01$ ,  $r\approx 0.29$ ), что может быть следствием *интерференции*: в случае, когда в исходном тексте использована синтетическая форма превосходной степени, она может быть перенесена в текст перевода в виде соответствующей ей синтетической формы русского языка (*the fastest* → *быстрейший* вместо *наиболее быстрый*). Как отмечалось в разделе 1.4, использование синтетических форм превосходной степени сравнения в русском научном тексте достаточно редко, при этом формы с суффиксами *-ейш-*, *-айш-* (*быстрейший*, *высочайший*, *типичнейший*) почти не используются из-за свойственной им эмоционально-экспрессивной окрашенности. В связи с этим при переводе и постредактировании научных текстов

следует обращать внимание на употребление форм превосходной степени сравнения и при необходимости заменять их формами, более соответствующими стилистическим нормам русского научного стиля (*наиболее быстрый / высокий / типичный*).

Кроме того, в социологическом переводном подкорпусе наблюдается рост частоты использования аналитических форм сравнительной степени (*COMP\_MOD ADJF*:  $\Delta = -0.07$ ,  $p < 0.001$ ,  $r \approx 0.41$ ), что может объясняться либо *интерференцией* аналитических форм степеней сравнения английского языка, либо *нормализацией*, при которой переводчики стремятся употреблять более типичные для русского научного текста (т. е. аналитические) формы сравнительной степени.

### **Индикаторы, отражающие особенности научного стиля русского языка**

Перейдём к индикаторам, отобраным с опорой на стилеобразующие характеристики русского научного стиля. Полные результаты статистического анализа индикаторов данной группы приведены в ПРИЛОЖЕНИИ 3.

Начнем интерпретацию результатов анализа с дискурсивных маркеров *авторского начала в тексте* и маркеров *взаимодействия автора с читателем*. В таблице 18 приведены результаты анализа маркеров, по которым были выявлены статистически значимые различия.

Таблица 18 – Результаты анализа *дискурсивных маркеров авторского начала в тексте и взаимодействия автора с читателем*

Корпус	Non-translated (median)	Translated (median)	HL	<i>p</i>
<b>Маркеры авторского начала в тексте</b>				
Высокая степень уверенности (%)				
СОЦ-РП	0.26	0.38	-0.12	***
МЕД-РП	0.15	0.26	-0.10	***
Средняя степень уверенности (%)				
СОЦ-РП	0.47	0.66	-0.19	***
МЕД-РП	0.29	0.47	-0.19	***

Продолжение таблицы 18

Корпус	Non-translated (median)	Translated (median)	HL	<i>p</i>
<b>Маркеры взаимодействия автора с читателем</b>				
Акцентирование внимания (%)				
СОЦ-РП	0.027	0.019	+0.013	*
Апелляция к фоновым знаниям (%)				
МЕД-МП	0.042	0	+0.028	***

Во подкорпусах ручных переводов наблюдается тенденция к увеличению частоты дискурсивных маркеров *высокой* (маркеры типа *должен, безусловно, несомненно, определенно* и др.) и *средней степени уверенности* (маркеры типа *может, возможно, вероятно, предположительно, похоже, видимо* и др.). В переводном социологическом подкорпусе по показателю *нормализованной частоты маркеров высокой степени уверенности* различия составляют  $HL=-0.12$ ,  $p<0.001$ ,  $r\approx 0.35$ ,  $A_{12}\approx 0.30$ , а по показателю *нормализованной частоты маркеров средней степени уверенности* –  $HL=-0.19$ ,  $p<0.001$ ,  $r=0.36$ ,  $A_{12}\approx 0.29$ . В то время как в переводном медицинском подкорпусе различия по показателю *нормализованной частоты маркеров высокой степени уверенности* составляют  $HL=-0.10$ ,  $p<0.001$ ,  $r\approx 0.28$ ,  $A_{12}\approx 0.34$ , а по показателю *нормализованной частоты маркеров средней степени уверенности* –  $HL=-0.19$ ,  $p<0.001$ ,  $r\approx 0.34$ ,  $A_{12}\approx 0.30$ . Можно предположить, что в обоих случаях наблюдаемая тенденция обусловлена влиянием англоязычных исходных текстов, для которых характерен крайне богатый спектр средств выражения эпистемической возможности или необходимости [Biber et al., 2021: 957-978]. Кроме того, результат согласуется с приведенными ранее результатами анализа частеречных последовательностей, продемонстрировавших рост частоты употребления модальных глаголов (*MDL\_VRB*), модальных слов (*MDL\_WRD*) и конструкций *MDL\_VRB INF* в ручных переводах.

В ручных переводах по социологии фиксируется некоторое снижение частоты использования маркеров *акцентирования внимания* ( $HL=+0.013$ ,  $p<0.05$ ,  $r\approx 0.22$ ,  $A_{12}\approx 0.63$ ). Результат может объясняться влиянием исходных текстов англоязычных научных статей, для которых в меньшей мере, чем для

русскоязычных, свойственно использование данного типа маркеров (Ср. [Валеева, Руднева, 2016]).

Наконец, в корпусе МЕД-МП в машинных переводах наблюдается снижение частоты использования маркеров *апелляции к фоновым знаниям* (маркеры типа *как известно, общеизвестно, как принято считать* и т. д.) ( $HL=+0.076$ ,  $p<0.001$ ,  $r\approx 0.39$ ,  $A_{12}\approx 0.7$ ). Такой результат может быть обусловлен тем, что подобные конструкции в русскоязычных научных статьях используются для подтверждения легитимности выдвигаемых в тексте положений несколько чаще, чем в англоязычных статьях (Ср. [Zanina, 2016]).

Теперь перейдем к результатам анализа использования *имен существительных* (Табл. 19).

Таблица 19 – Результаты анализа *имен существительных*

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	$p$
Nouns to tokens ratio (%)				
СОЦ-РП	36.12	33.38	+2.74	***
МЕД-РП	41.27	38.90	+2.37	***
МЕД-МП	41.27	37.41	+3.86	***
Singular number nouns to all nouns (%)				
СОЦ-РП	71.30	67.04	+4.26	***
МЕД-МП	72.49	66.82	+5.67	***
Abstract nouns to all nouns (%)				
СОЦ-РП	42.38	39.19	+3.19	**
МЕД-РП	36.8	39.16	-2.36	**

Результаты анализа *доли имен существительных* в тексте подтверждают выявленную ранее тенденцию к снижению номинативной плотности переводных текстов во всех трех корпусах. Так, доля имён существительных снижается в ручных переводах социологических ( $\Delta=+2.74$ ,  $p<0.001$ ,  $r\approx 0.46$ ) и медицинских статей ( $\Delta=+2.37$ ,  $p<0.001$ ,  $r\approx 0.40$ ) и в особенности в машинных переводах медицинских статей ( $\Delta=+3.86$ ,  $p<0.001$ ,  $r\approx 0.59$ ).

Другое статистически значимое различие касается *имен существительных в форме единственного числа*: их доля существенно снижается в ручных переводах

социологических статей ( $\Delta=4.26$ ,  $p<0.001$ ,  $r\approx 0.37$ ) и в машинных переводах медицинских статей ( $\Delta=+5.67$ ,  $p<0.001$ ,  $r\approx 0.58$ ). Такое снижение может быть связано с тем, что в английском научном тексте множественное число используется более активно, чем в русском, в том числе для выражения общего понятия или именованной некоторой совокупности предметов (*patients, cells*), тогда как в русском языке в этой функции предпочтение обычно отдается существительным в форме единственного числа (*пациент, клетка*) (см. раздел 1.4). Повышенная частота форм множественного числа в переводах может, таким образом, быть следствием калькирования форм множественного числа исходного текста.

В корпусах СОЦ-РП и МЕД-РП в переводных подкорпусах фиксируются разнонаправленные тенденции в отношении доли абстрактных имен существительных: в подкорпусе ручных переводов по социологии данный показатель снижается ( $\Delta=+3.19$ ,  $p<0.01$ ,  $r\approx 0.26$ ), в то время как в подкорпусе ручных переводов по медицине – увеличивается ( $\Delta=-2.36$ ,  $p<0.01$ ,  $r\approx 0.23$ ). Можно предположить, что снижение доли абстрактных имен существительных в переводах по социологии может быть либо следствием снижения уровня номинативности переводов (например, если вместо слов типа *применение, корреляция* используются глаголы *применять, коррелировать*), либо того, что в исходных статьях их доля уступает доле конкретных имен существительных, либо показателем некоторого упрощения текста перевода для восприятия: как известно, абстрактные существительные являются более сложными для восприятия, чем конкретные [Roxbury et al., 2014]. Интересным при этом является не нарушающее норм русского научного стиля увеличение доли абстрактных существительных в ручных переводах медицинских статей. Вместе с тем установление точных причин наблюдаемых тенденций требует дальнейших исследований на материале параллельных подкорпусов.

Перейдем к результатам анализа имен прилагательных. В таблице 20 представлены результаты анализа индикаторов, продемонстрировавших статистически значимые различия.

Таблица 20 – Результаты анализа *имен прилагательных*

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	<i>p</i>
Adjectives to tokens ratio (%)				
СОЦ-РП	15.75	14.39	+1.36	***
МЕД-РП	15.63	14.62	+1.01	**
МЕД-МП	15.63	13.39	+2.24	***
Short adjectives to all adjectives (%)				
СОЦ-РП	4.66	6.31	-1.65	***
МЕД-РП	2.96	4.36	-1.39	***
Superlative adjectives to all adjectives (%)				
МЕД-МП	1.11	1.53	-0.42	*

Во всех трёх корпусах наблюдается статистически значимое снижение доли *прилагательных* в переводах, а именно в ручных переводах статей по социологии ( $\Delta=+1.36$ ,  $p<0.001$ ,  $r\approx 0.36$ ) и медицине ( $\Delta=+1.01$ ,  $p<0.01$ ,  $r\approx 0.24$ ) и в машинных переводах медицинских статей ( $\Delta=+2.24$ ,  $p<0.001$ ,  $r\approx 0.49$ ). Отметим, что прилагательные, наряду с существительными, обеспечивают номинативность текста, а их пониженная частота в переводах коррелирует со снижением доли имен существительных и, следовательно, может быть следствием *интерференции*.

В корпусах СОЦ-РП и МЕД-РП наблюдается увеличение доли кратких имен прилагательных в ручных переводах как статей по социологии ( $\Delta=-1.65$ ,  $p<0.001$ ,  $r\approx 0.43$ ), так и по медицине ( $\Delta=-1.39$ ,  $p<0.001$ ,  $r\approx 0.36$ ). Можно предположить, что данный результат есть следствие переноса в текст перевода англоязычных составных именных сказуемых с именем прилагательным (ср.: *the indicators are comparable* → *показатели сопоставимы*), однако данное утверждение требует проверки на англоязычном параллельном подкорпусе текстов.

В подкорпусе МП медицинских статей наблюдается некоторое увеличение доли прилагательных в форме превосходной степени ( $\Delta=-0.42$ ,  $p<0.05$ ,  $r\approx 0.27$ ), что в ряде случаев может нарушать нормы русского научного стиля с его требованием к объективности и аккуратности делаемых в рамках исследования высказываний и что должно учитываться на этапе постредактирования.

Перейдем к индикаторам, связанным с использованием разных глагольных форм. В таблице 21 представлены данные по индикаторам, анализ которых показал статистически значимые различия между подкорпусами.

Таблица 21 – Результаты анализа использования *глагольных форм*

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	<i>p</i>
Verbs to tokens ratio (%)				
СОЦ-РП	11.86	12.88	-1.02	***
МЕД-РП	9.62	10.60	-0.98	***
МЕД-МП	9.62	10.13	-0.51	*
Passive verbs ratio (%)				
СОЦ-РП	20.54	17.74	+2.79	**
МЕД-МП	35.27	39.39	-4.12	**
Participles to all verbs ratio (%)				
СОЦ-РП	22.20	17.99	+4.21	***
Participles before nouns to all participles (%)				
МЕД-МП	24.48	21.06	+3.42	***
Future verbs to all finite verbs (%)				
СОЦ-РП	2.92	4.30	-1.39	**
МЕД-РП	0.83	1.42	-0.60	*
Present verbs to all finite verbs (%)				
МЕД-МП	57.02	38.27	+18.75	***
Past verbs to all finite verbs (%)				
МЕД-МП	41.84	60.21	-18.37	***
Imperfective verbs to all verbs ratio (%)				
МЕД-МП	62.74	54.01	+8.73	***
Imperfective finite verbs to all finite verbs (%)				
МЕД-МП	83.37	74.99	+8.38	***

Во всех трёх переводных подкорпусах фиксируется увеличение доли глаголов: в ручных переводах статей по социологии ( $\Delta=-1.02$ ,  $p<0.001$ ,  $r\approx 0.38$ ) и по медицине ( $\Delta=-0.98$ ,  $p<0.001$ ,  $r\approx 0.36$ ) и в машинных переводах статей по медицине ( $\Delta=-0.51$ ,  $p<0.05$ ,  $r\approx 0.23$ ). Данный результат полностью согласуется с описанным ранее снижением номинативности переводных текстов.

В корпусе СОЦ-РП наблюдается значительное снижение доли страдательных форм глагола ( $\Delta=+2.79$ ,  $p<0.01$ ,  $r\approx 0.28$ ). Прежде всего данный результат может быть обусловлен *интерференцией* английского языка, современная норма академического письма которого, допускает широкое использование активного залога и местоимений 1-го лица (*I / we*) (см. раздел 1.4.), в то время как для русскоязычных научных текстов в подобных случаях более предпочтительными являются формы страдательного залога (Ср. *In this paper we examine...* и *В статье рассматривается...*; *We collected data from* и *Собраны данные из...*). Данный вывод также согласуется с результатами анализа частеречных n-грамм (увеличение доли n-граммы *PER\_PRN* больше всего выражено в социологическом подкорпусе ручных переводов), а также подкрепляется приведенными ниже результатами анализа индикаторов, связанных с употреблением разных разрядов местоимений. Существенную роль, по-видимому, играет и предметная специфика социологических текстов: в них значительно чаще, чем в медицинских, речь идёт о людях, социальных группах, участниках опросов, информантах, исследователях и т. д., что в большей степени способствует использованию личных местоимений с глаголами в активном залоге. Кроме того, снижение доли страдательных форм глагола может отражать тенденцию к *упрощению*, поскольку страдательные (пассивные) формы считаются когнитивно более сложными для восприятия [Ferreira, 2003], в связи с чем могут заменяться действительными (активными) формами в переводе.

В МЕД-МП в подкорпусе машинных переводов, напротив, фиксируется значительный рост доли форм страдательного залога ( $\Delta=-4.12$ ,  $p<0.01$ ,  $r\approx 0.28$ ). В целом высокая частота использования страдательных конструкций в медицинских текстах закономерна, поскольку страдательный залог является для них ключевым средством реализации объективно-нейтральной манеры изложения, особенно важной при описании данных медицинских исследований. При этом увеличение доли форм страдательного залога в МП по сравнению с непереводаемыми статьями, вероятнее всего, связано с уже упоминавшимся снижением номинализации глагольных конструкций исходного текста: английские пассивные конструкции

(типа *was performed, were analyzed*) передаются в МП с помощью форм страдательного залога (*было проведено / проводилось, были проанализированы / анализировались*), тогда как в непереводных текстах, по всей видимости, вместо них авторы нередко используют отглагольные существительные (*проведение чего-либо, измерение чего-либо*) – данный вывод косвенно подтверждается представленными ранее результатами анализа частеречных последовательностей: в МП фиксируется значительное увеличение доли комбинаций с аналитическими страдательными конструкциями (*AUX PRTS*:  $\Delta=-0.42$ ,  $p<0.001$ ,  $r\approx 0.70$ ; *AUX PRTS PREP*:  $\Delta=-0.32$ ,  $p<0.001$ ,  $r\approx 0.85$ ; *NOUN AUX PRTS*:  $\Delta=-0.18$ ,  $p<0.001$ ,  $r\approx 0.56$ ).

В корпусе СОЦ-ПП фиксируется значительное уменьшение доли причастий в ручных переводах ( $\Delta=+4.21$ ,  $p<0.001$ ,  $r\approx 0.43$ ), что коррелирует как с описанным ранее повышением доли местоименных союзных слов (*PRN\_CONJ\_W*) и комбинаций с ними, а также с приведенным ниже повышением доли относительных местоимений и может быть интерпретировано как проявление синтаксической *интерференции*: вероятно, переводчики воспроизводят англоязычную модель выражения атрибутивных отношений посредством придаточных определительных предложений, не используя в полной мере потенциал причастных конструкций русского языка [Kunilovskaya, Kutuzov, 2017: 32-33; Kunilovskaya, Corpas Pastor, 2021: 134-135]. При этом, как известно, причастия являются характерным для научного стиля средством компрессии информации, позволяющим избегать более громоздких придаточных предложений, а также рассматриваются как маркер более высокого регистра.

Интересно также, что в обоих подкорпусах ручных переводов фиксируется увеличение доли глаголов в форме будущего времени, а именно в ручных социологических ( $\Delta=-1.39$ ,  $p<0.01$ ,  $r\approx 0.36$ ) и медицинских переводах ( $\Delta=-0.60$ ,  $p<0.05$ ,  $r\approx 0.24$ ). Установление причин наблюдаемой тенденции требует анализа параллельного корпуса текстов, который позволил бы выявить, какие именно конструкции исходных текстов передаются формами будущего времени в переводе. С практической точки зрения полученный результат заслуживает

особого внимания при постредактировании ручных переводов, поскольку формы будущего времени не столь типичны для русского научного стиля.

В корпусе МЕД-МП в машинных переводах фиксируется значительное снижение доли причастий, стоящих перед определяемым ими именем существительным ( $\Delta=+3.42$ ,  $p<0.001$ ,  $r\approx 0.30$ ), являющимся главным словом<sup>13</sup>. Данный результат, вероятно, является следствием *интерференции* используемых в англоязычных научных текстах конструкций типа *the method proposed by...*, *the data obtained during...*, *the approach used in...* – такие конструкции в процессе МП на русский язык сохраняют свою структуру: *метод, предложенный (кем-то)*; *данные, полученные в ходе (чего-то)*; *подход, использованный в...* При этом возможные варианты перевода типа *предложенный (кем-то) метод*; *полученные в ходе (чего-то) данные*; *используемый в ... подход* система МП, вероятнее всего, не «выбирает» или делает это реже. Как уже упоминалось, подобные рамочные конструкции не имеют аналога в английском языке, а конструкции типа \* *the proposed by...method*, \* *the obtained during...data* и \* *the used here approach* в английском языке невозможны. В результате при переводе на русский система МП калькирует структуру исходных англоязычных конструкций. Данная закономерность может учитываться при постредактировании МП научных текстов: вероятно, в ряде случаев постпозитивные причастные обороты могут быть преобразованы в соответствующие им номинативные рамочные конструкции с причастием (препозитивные определения) там, где это представляется допустимым.

Особое внимание обращает на себя стремительное снижение в машинных переводах доли *глаголов несовершенного вида относительно всех глаголов* ( $\Delta=+8.73$ ,  $p<0.001$ ,  $r\approx 0.55$ ) и *относительно личных форм глаголов* ( $\Delta=+8.38$ ,  $p<0.001$ ,  $r\approx 0.51$ ). Вероятно, данная тенденция в МП фиксируется из-за того, что, поскольку в английском языке у глаголов отсутствует категории вида, система МП не всегда способна определить, какой именно видовой категории должен соответствовать глагол исходного текста в ПЯ (Ср. [Manakhimova et al., 2025: 872]).

<sup>13</sup> Учитываются в том числе и те случаи, когда причастие стоит перед определяемым им именем существительным, но разделено с ним несколькими словами (например, *приведенный в качестве примера случай*)

Кроме того, как было рассмотрено в разделе 1.4., частое использование глаголов совершенного вида для научного стиля русского языка не свойственно. Предпочтение, как правило, отдается глаголам несовершенного вида как передающим значение процессуальности, повторяемости или большей объективности описываемого действия. Следовательно, при постредактировании результатов МП особенно важной является проверка правильности передачи видовых значений, в частности замена глаголов совершенного вида на глаголы несовершенного вида там, где это необходимо.

Наиболее примечательными являются различия в распределении временных форм глаголов в корпусе МЕД-МП между непереводаемым подкорпусом и подкорпусом машинных переводах. Так, в МП наблюдается существенное снижение доли глаголов в форме настоящего времени ( $\Delta=+18.75$ ,  $p<0.001$ ,  $r\approx 0.53$ ), сопровождаемое значительным увеличением доли глаголов в форме прошедшего времени ( $\Delta=-18.37$ ,  $p<0.001$ ,  $r\approx 0.51$ ). Такое расхождение можно объяснить факторами, связанными с *интерференцией* английского языка и спецификой работы систем МП. Английский научный стиль допускает более активное использование глаголов в форме прошедшего времени, тогда как в русском научном дискурсе предпочтение отдается формам настоящего времени в отвлечённо-обобщённом значении (см. раздел 1.4). Система МП, по всей видимости, систематически калькирует временные формы исходного текста, воспроизводя прошедшее время даже там, где в русской научной традиции форма настоящего времени была бы более уместна.

Наконец, перейдем к результатам анализа *местоимений разных разрядов* (Табл. 22).

Таблица 22 – Результаты анализа *местоимений разных разрядов*

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	<i>p</i>
Personal / possessive pronouns ratio (%)				
СОЦ-РП	2.38	3.06	-0.68	***
МЕД-РП	0.82	1.02	-0.19	*
МЕД-МП	0.82	1.04	-0.22	*

Продолжение таблицы 22

Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	<i>p</i>
Demonstrative pronouns ratio (%)				
МЕД-РП	1.76	2.42	-0.66	***
Relative pronouns ratio (%)				
СОЦ-РП	2.04	2.62	-0.59	***
МЕД-РП	1.24	1.53	-0.29	***
МЕД-МП	1.24	1.65	-0.41	***
Defining pronouns ratio (%)				
МЕД-РП	0.5	0.7	-0.20	***
МЕД-МП	0.5	0.69	-0.19	***
Indefinite pronouns ratio (%)				
МЕД-РП	0.13	0.21	-0.075	***
МЕД-МП	0.13	0.17	-0.045	*

Во всех трех корпусах фиксируется рост доли *личных и притяжательных местоимений* в переводе, а именно в ручных переводах статей по социологии ( $\Delta=-0.68$ ,  $p<0.001$ ,  $r\approx 0.41$ ) и медицине ( $\Delta=-0.19$ ,  $p<0.05$ ,  $r\approx 0.21$ ) и в машинных переводах медицинских статей ( $\Delta=-0.22$ ,  $p<0.05$ ,  $r\approx 0.26$ ). Данный результат, вероятнее всего, является следствием *интерференции*, поскольку, как уже упоминалось, современная норма англоязычного академического письма допускает и в ряде случаев поощряет использование местоимений первого лица для эксплицитного обозначения личности исследователя, тогда как в русскоязычном научном дискурсе их употребление традиционно сведено к минимуму в соответствии с требованием деперсонализации научного изложения.

Кроме того, во всех трех корпусах фиксируется увеличение доли относительных местоимений в переводе: в ручных переводах по социологии ( $\Delta=-0.59$ ,  $p<0.001$ ,  $r\approx 0.50$ ) и медицине ( $\Delta=-0.29$ ,  $p<0.001$ ,  $r\approx 0.31$ ) и в МП по медицине ( $\Delta=-0.41$ ,  $p<0.001$ ,  $r\approx 0.44$ ). Данный результат согласуется с увеличением доли местоименных союзных слов во всех переводных подкорпусах и, по всей видимости, обусловлен воспроизведением структуры английских ограничительных определительных (Defining Relative Clauses) и / или

распространительных (Non-Defining Relative Clauses) придаточных (ср. [Kunilovskaya, Kutuzov, 2017: 32-33; Kunilovskaya, Corpas Pastor, 2021: 134-135]).

В корпусах МЕД-РП и МЕД-МП наблюдается следующая тенденция: растёт доля *неопределённых* ( $\Delta=-0.075$ ,  $p<0.001$ ,  $r\approx 0.34$ ) и *определятельных местоимений* ( $\Delta=-0.20$ ,  $p<0.001$ ,  $r\approx 0.37$ ) в ручных переводах, а также доля *неопределённых* ( $\Delta=-0.045$ ,  $p<0.05$ ,  $r\approx 0.22$ ) и *определятельных* ( $\Delta=-0.19$ ,  $p<0.001$ ,  $r\approx 0.4$ ) *местоимений* в машинных переводах. Данный результат, вероятно, обусловлен *интерференцией*: поскольку, как уже упоминалось, в английском языке у имён существительных присутствует категория определённости / неопределённости, в большинстве случаев она требует своего маркирования с помощью разнообразных определятельных элементов (типа *every, each, any, some* и т. п.), которые при переводе могут переноситься в русский текст в виде соответствующих им местоимений русского языка (типа *каждый, любой / всякий, некоторый*).

В подкорпусе ручных переводов медицинских статей фиксируется рост доли *указательных местоимений* ( $\Delta=-0.66$ ,  $p<0.001$ ), что, вероятно, является проявлением *экспликации* референциальных связей в тексте и согласуется с представленными ранее результатами анализа частеречных последовательностей.

### Выводы по главе 3

В рамках настоящей главы были решены следующие практические задачи:

1. Сформированы три монолингвальных сопоставимых корпуса: СОЦ-РП (непереводные и ручные переводы социологических статей), МЕД-РП (непереводные и ручные переводы медицинских статей) и МЕД-МП (непереводные и машинные переводы медицинских статей).

2. Разработан специализированный программный комплекс для исследования феномена переводческого языка, обеспечивший автоматизированное измерение показателей 46 индикаторов переводческих характеристик, а также их сопоставление между подкорпусами.

3. Изложена и применена методика статистического анализа полученных данных, включающая t-критерий Уэлча и U-критерий Манна–Уитни, а также

расчёт величины эффекта  $r$  для оценки практической значимости выявленных различий.

4. В ходе анализа установлено, что проанализированные характеристики переводных текстов проявляются во всех исследованных корпусах, однако степень и характер их проявления варьируются в зависимости от предметной области и типа перевода. Кроме того, установлено, что проявления характеристик неоднозначны с точки зрения их влияния на качество перевода и могут быть разделены на две группы:

(а) Проявления, требующие учёта при переводе и постредактировании, поскольку связаны со стилистическими отклонениями от норм русского научного стиля и могут оказывать влияние на качество перевода:

- Снижение степени номинативности и рост степени глагольности научных переводов. Поскольку высокая степень номинативности – стилеобразующая черта русского научного стиля, переводчику и постредактору следует целенаправленно контролировать её степень, чаще прибегая к номинализации (субстантивации) глаголов там, где это необходимо.

- Снижение среднего ранга используемой лексики во всех переводных подкорпусах, свидетельствующее о тенденции к выбору более частотных и общеупотребительных единиц языка перевода. При переводе и постредактировании следует проверять, не заменяется ли терминологическая лексика исходного текста более общеупотребительными, но менее точным эквивалентами ПЯ в переводе.

- Нарушение видо-временной соотнесенности глагольных форм в МП. Система МП калькирует английские временные формы глаголов оригинала, что ведёт к нехарактерному для русского научного стиля обилию форм прошедшего времени в переводе. Кроме того, по причине отсутствия у глагола категории вида в английском языке система МП часто некорректно соотносит глагол исходного текста с требуемой видовой категорией глагола в ПЯ, что ведёт к нетипичному для русских научных текстов обилию глаголов совершенного вида. При постредактировании МП необходима целенаправленная проверка временных /

видовых значений с заменой неверно переведенных глаголов на глаголы настоящего времени / несовершенного вида там, где необходимо.

- Структурная переусложненность предложений в МП, проистекающая из неспособности систем МП выполнять надлежащие синтаксические перестройки – разбиение длинных сложных предложений на несколько более простых в переводе, в связи чем в МП наблюдается значительное снижение доли простых предложений, а также увеличение числа клауз на предложение.

- Увеличение частоты употребления глаголов-связок во всех переводных подкорпусах (особенно в МП). В русском языке в настоящем времени глагол-связка нередко опускается, в то время как в английском языке её присутствие грамматически обязательно. При переводе и постредактировании следует контролировать частоту использования глаголов-связок.

- Снижение лексической повторяемости знаменательных единиц, особенно выраженное в подкорпусе МП медицинских статей. Системы МП, по всей видимости, могут передавать одну и ту же лексическую единицу оригинала различными способами на протяжении одного и того же текста, из-за чего снижается лексическая связность текста. При постредактировании МП по медицине необходимо целенаправленно проверять единообразие передачи знаменательных, а в особенности терминологических единиц в тексте.

- Увеличение доли модальных глаголов и модальных конструкций *модальный глагол + инфинитив* в ручных переводах. При переводе и постредактировании рекомендуется чаще варьировать способы выражения модальности, используя не только модальные глаголы, но и модальные слова, вводные конструкции и иные средства выражения модальности.

- Избыточное использование аналитических страдательных конструкций (*AUX PARTS*) в подкорпусах ручных и в особенности машинных переводов медицинских статей, являющееся следствием переноса английских аналитических пассивных форм (*was conducted* → *был проведён*, *were measured* → *были измерены*). При переводе и постредактировании рекомендуется чаще

варьировать аналитические конструкции и возвратные глаголы в страдательном значении (*был проведён → проводился, были измерены → измерялись*).

- Значительное увеличение доли личных местоимений в функции подлежащего, в том числе в начале предложения, а также комбинаций *местоимение + личная форма глагола (PER\_PRN VERB)* во всех переводных подкорпусах. Данная тенденция обусловлена *интерференцией*: англоязычные конвенции научного письма допускают более широкое использование в тексте местоимений первого лица для обозначения личности исследователя, в то время как для русскоязычной научной традиции это не столь типично и не всегда желательно. При переводе и постредактировании рекомендуется заменять конструкции с личным местоимением и личной формой глагола на страдательные конструкции (например: *мы обнаружили связь → была обнаружена связь; мы наблюдаем чёткую корреляцию → наблюдается чёткая корреляция*).

- Рост доли синтетических форм превосходной степени прилагательных (типа *важнейший, крупнейший*) в социологическом переводном подкорпусе. Синтетические формы превосходной степени почти не свойственны русскому научному стилю, предпочтение отдается аналитическим формам (*наиболее важный, наиболее значимый*). При переводе и постредактировании следует контролировать выбор формы превосходной степени, отдавая предпочтение аналитическим формам, за исключением случаев, когда синтетическая форма является устоявшимся способом выражения (например, *дальнейший анализ*).

- Значительное снижение доли существительных в форме единственного числа в социологическом переводном подкорпусе и в медицинском подкорпусе МП, вероятно, обусловленное калькированием форм множественного числа: в английском научном тексте множественное число используется активнее, чем в русском, в том числе для обозначения общего понятия или именованной совокупности предметов, тогда как в русской научной традиции в аналогичной функции предпочтение отдаётся форме единственного числа. При переводе и постредактировании в контекстах с аналогичной функцией следует чаще заменять форму множественного числа на форму единственного.

(б) Проявления, фиксирующие отличие переводов от непереводаемых текстов и либо не влияющие на качество перевода, либо требующие дальнейших исследований для установления характера их влияния на качество перевода:

- Снижение лексической плотности во всех переводных подкорпусах (в особенности в МП), свидетельствующее об увеличении доли служебной (функциональной) лексики и снижении доли знаменательной лексики.

- Снижение лексической вариативности в подкорпусах ручных и машинных переводов медицинских статей.

- Значительное увеличение индекса удобочитаемости Флеша (на 9.41%) в переводах научных статей по социологии, коррелирующее с уменьшением средней длины слов и предложений.

- Редкое использование *type* как в машинных, так и в ручных переводах, обусловленное функциональной асимметрией данного знака в английском и русском языках. В русском языке *type* используется в нетипичных для английского функциях: для замены глагола-связки, всего сказуемого, для передачи разных логико-семантических отношений (следствия, причины и т. д.), из-за чего оно часто не воспроизводится и не используется в переводе.

- Снижение частоты использования открывающих и закрывающих круглых скобок в ручных переводах, отражающее тенденцию к редукции парентетической сложности: переводчики интегрируют заключённую в скобках информацию в основной текст.

- Снижение степени ассоциативной связанности используемых в ручных переводах биграмм, свидетельствующее о том, что употребляемые в них коллокации менее устойчивы и менее типичны для переводящего языка.

- Увеличение степени ассоциативной связанности используемых в машинном переводе биграмм, свидетельствующее о том, что системы МП тяготеют к воспроизведению более предсказуемых, шаблонных коллокаций переводящего языка.

- Увеличение частоты употребления маркеров противопоставления в обоих подкорпусах ручных переводов (социологическом и медицинском).

- Увеличение частоты использования маркеров целевых отношений в подкорпусах ручных и в особенности машинных переводов медицинских статей.
- Увеличение частоты использования маркеров введения примеров в подкорпусах ручных и машинных переводов медицинских статей и небольшой рост частоты маркеров причинно-следственных отношений в подкорпусе ручных медицинских переводов.
- Снижение частоты употребления маркеров выводов и заключения и рост частоты использования маркеров порядка следования информации в подкорпусе МП по медицине.
- Рост частоты использования маркеров высокой и средней степени уверенности в подкорпусах ручных переводов (социологическом и медицинском).
- Рост доли неопределённых и определительных местоимений как в машинных, так и в ручных переводах, вероятно, обусловленный *интерференцией* английского языка, где необходимым является маркирование категории (не)определённости, для чего нередко используются местоимения типа *some, any, every, each* и т. д., которые при переводе могут передаваться схожими средствами русского языка (*некоторый, какой-то, любой, каждый* и т. д.).
- Рост доли подчинительных союзов и союзных слов во всех переводных подкорпусах. В частности увеличение частоты относительных местоимений, вводящих придаточные определительные предложения. Результат, вероятно, является как следствием калькирования английских сложноподчинённых конструкций, так и *экспликаций* имплицитно содержащейся в исходном тексте информации.
- Рост доли глаголов будущего времени в ручных переводах.
- Снижение частоты использования перечислений имён существительных через запятую (последовательность NOUN COMMA NOUN) во всех переводных подкорпусах.
- Снижение доли абстрактных имён существительных в подкорпусе ручных социологических переводов и увеличение их доли в подкорпусе ручных медицинских переводов.

- Увеличение доли кратких имён прилагательных в подкорпусах ручных переводов (социологическом и медицинском).
- Увеличение доли аналитических форм сравнительной степени в социологическом подкорпусе ручных переводов.
- Увеличение доли указательных местоимений и комбинаций с ними в подкорпусах ручных социологических и медицинских переводов. Данная тенденция может быть интерпретирована как проявление *экспликации*, поскольку в подкорпусе машинных переводов аналогичной тенденции не наблюдается.
- Рост частоты использования сослагательной частицы *by* в подкорпусах ручных переводов (социологическом и медицинском), вероятно, обусловленный интерференцией английских конструкций с модальными глаголами *would* и *could* в эпистемическом значении.

## ЗАКЛЮЧЕНИЕ

Настоящее диссертационное исследование посвящено изучению феномена переводческого языка на материале переводов научных статей с английского на русский язык. Целью работы являлось установление индикаторов феномена переводческого языка, влияющих на качество переводов научных статей с английского языка на русский. В ходе исследования был решен ряд задач, поставленных с целью достижения цели исследования.

При решении первой задачи были рассмотрены исторические предпосылки, обусловившие интерес к исследованиям феномена переводческого языка, а также изучены существующие подходы к его определению и оценке. Установлено, что возникновение научного интереса к данному феномену было обусловлено теоретико-методологическими сдвигами в переводоведении последней четверти XX века, в первую очередь переходом от прескриптивного к дескриптивному подходу, в рамках которого перевод стал рассматриваться как самостоятельный объект научного исследования. Дескриптивные исследования перевода стали играть в этом контексте главную роль, поскольку предполагают описание явлений, реально наблюдаемых в переводных текстах. Именно в рамках дескриптивного направления исследования феномена переводческого языка приобрели системный характер. Что касается подходов к пониманию феномена, то они варьируются от его оценки как нежелательного явления, снижающего качество переводных текстов, до нейтральных концепций языковых «отпечатков» исходного языка и промежуточного языка. При всём разнообразии трактовок все они признают объективное существование специфических характеристик, систематически проявляющихся в переводах и отличающих их от непереводаемых текстов на ПЯ.

В рамках второй задачи были рассмотрены концептуальные подходы к характеристикам переводческого языка как к переводческим универсалиям. Критический анализ концепции переводческих универсалий М. Бейкер, предпринятый Г. Тури и Э. Честерманом, показал, что утверждение об абсолютной универсальности характеристик переводных текстов некорректно ввиду зависимости их проявления от разнообразных факторов (языковой пары, жанрово-

стилистической специфики текста и т. д.). В связи с этим в настоящей работе был принят подход к характеристикам феномена переводческого языка как к гипотетически универсальным особенностям переводов, а в качестве методологической основы была выбрана классификация, включающая четыре характеристики переводных текстов: *упрощение*, *нормализацию*, *экспликацию* и *интерференцию*. Был принят подход «снизу вверх», позволяющий избежать преждевременных обобщений об универсальности изучаемого явления.

Третья задача предусматривала описание и сопоставление основных стилеобразующих характеристик научного стиля в английском и русском языках, что послужило основой для формирования дополнительной группы индикаторов, учитывающих нормы научного стиля русского языка, а также основой для адекватной интерпретации результатов корпусного анализа.

В рамках четвертой задачи был осуществлен отбор формальных индикаторов потенциальных универсальных характеристик переводов. Был сформирован комплекс из 46 индикаторов, распределённых по пяти группам: (1) 16 индикаторов характеристики *упрощение*, (2) 3 индикатора характеристики *нормализация*, (3) 7 индикаторов характеристики *экспликация*, (4) 1 индикатор характеристики *интерференция* и (5) 19 индикаторов, отражающих стилеобразующие черты научного стиля русского языка.

Пятая задача была связана с разработкой специализированного программного комплекса для автоматизированного анализа отобранных индикаторов. Разработанный на базе языка программирования Python программный комплекс обеспечил возможность автоматизированного сбора корпусов текстов, их предобработки (токенизации, лемматизации и морфологической разметки), вычисления показателей 46 индикаторов, а также сопоставления полученных результатов между подкорпусами.

В рамках шестой задачи были сформированы два сопоставимых исследовательских корпуса: корпус научных статей социологической тематики, включающий подкорпус непереводных текстов и подкорпус ручных переводов (по 59 текстов, ≈310 000 словоформ каждый), и корпус научных статей медицинской

тематики, включающий подкорпус непереводных текстов (70 текстов), подкорпус ручных переводов (70 текстов) и подкорпус машинных переводов (50 текстов), объём каждого из которых составил  $\approx 216\ 500$  словоформ. Общий объём проанализированного материала составил около 1 270 000 словоформ. Сформированные корпуса были проанализированы при помощи – разработанного программного комплекса, а полученные результаты подвергнуты статистическому анализу с применением двухвыборочного t-критерия Уэлча и U-критерия Манна–Уитни с поправкой на множественные сравнения. Для оценки практической значимости выявленных различий рассчитывалась величина эффекта  $\eta$ .

Результаты анализа подтвердили выдвинутую гипотезу: переводы научных статей с английского языка на русский демонстрируют статистически значимые отличия от непереводных русскоязычных научных статей по ряду индикаторов, соответствующих четырём основным характеристикам переводческого языка. При этом степень и характер проявления этих характеристик варьируются в зависимости от предметной области (социологическая / медицинская тематика) и способа выполнения перевода (ручной / машинный), причём машинные переводы демонстрируют более выраженные отклонения от норм русского научного стиля, чем ручные. Наиболее последовательно проявляющейся характеристикой во всех исследованных подкорпусах оказалась интерференция.

В рамках седьмой задачи на основании результатов анализа была произведена дифференциация выявленных проявлений переводческого языка на две группы. Первую группу составили проявления, потенциально влияющие на качество перевода, поскольку они связаны со стилистическими отклонениями от норм русского научного стиля. Для каждого из проявлений были сформулированы практические рекомендации для переводчиков и постредакторов. Вторую группу составили проявления, фиксирующие отличия переводных текстов от непереводных, но либо не влияющие на качество перевода, либо требующие дальнейших исследований для установления характера их влияния.

Таким образом, поставленная цель исследования – установление индикаторов феномена переводческого языка, влияющих на качество переводов

научных статей с английского языка на русский – достигнута. Определён набор конкретных индикаторов, фиксирующих стилистические отклонения от норм русского научного стиля и требующих учёта при переводе и постредктировании. Подтверждена гипотеза о наличии статистически значимых различий между переводными и непереводными научными статьями по ряду индикаторов, соответствующих четырём характеристикам переводческого языка. Разработанный программный комплекс обеспечивает возможность автоматизированного выявления рассматриваемых в настоящей работе индикаторов и может быть использован для установления аспектов перевода, требующих особого внимания при переводе и постредктировании.

Перспективы дальнейших исследований связаны с двумя основными направлениями. Во-первых, это расширение корпусной базы исследования за счёт включения в анализ параллельных корпусов текстов (оригинальных текстов на английском языке и их переводов), что позволит провести более детальный анализ переводческих характеристик, выявить механизмы их возникновения, проследить, каким образом конкретные особенности исходных текстов влияют на проявление переводческого языка. Во-вторых, это включение в исследование текстов научных статей других предметных областей, а также текстов иных стилей и жанров, что позволит проверить универсальность выявленных закономерностей на новом языковом материале.

## СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

В тексте приняты следующие условные обозначения:

- термины обозначаются курсивом;
- понятия приводятся в марровских кавычках ‘’.

→	графическое обозначение направления от ...к ..., указывающее на последовательность, преобразование одного варианта в другой
<b>ИЯ</b>	исходный язык
<b>ПЯ</b>	переводящий язык / язык перевода
<b>МП</b>	машинный перевод
<b>СОЦ-РП</b>	Социологический корпус (непереводные тексты и ручные переводы)
<b>МЕД-РП</b>	Медицинский корпус (непереводные тексты и ручные переводы)
<b>МЕД-МП</b>	Медицинский корпус (непереводные тексты и машинные переводы)

## СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1) Алексеева И. С. Введение в переводоведение: учебное пособие для студентов филологических и лингвистических факультетов высших учебных заведений / И. С. Алексеева. – Санкт-Петербург: Филологический факультет СПбГУ; Москва: Издательский центр «Академия», 2004. – 352 с.

2) Алексеева И. С. Профессиональное обучение переводчика: учебное пособие по устному и письменному переводу для переводчиков и преподавателей / И. С. Алексеева. – Санкт-Петербург: Издательство «Союз», 2001. – 288 с.

3) Арнольд И. В. Стилистика. Современный английский язык: учебник для вузов / И. В. Арнольд. – 10-е издание. – Москва: Флинта: Наука, 2010. – 384 с.

4) Арцишевская А. Н. Роль научного перевода в обеспечении качества и видимости публикаций / А. Н. Арцишевская // Научный редактор и издатель. – 2025. – Том 10, № 1. – С. 104-113.

5) Багдасарян А. Г. О "культурном повороте" в современных переводоведческих исследованиях / А. Г. Багдасарян // Филологические науки. Вопросы теории и практики. – 2018. – № 7-2 (85). – С. 295-298.

6) Балыгина Е. А. Номинализация как средство повышения коммуникативного статуса обстоятельства в англо-русском переводе / Е. А. Балыгина, О. А. Круковская // Язык и текст. – 2018. – Том 1, № 1. – С. 3-10.

7) Бархударов Л. С. Язык и перевод (вопросы общей и частной теории перевода) / Л. С. Бархударов. – Москва: Международные отношения, 1975. – 240 с.

8) Беляева Л. Н. Научная статья: подготовка к публикации в рецензируемых изданиях: учебное пособие / Л. Н. Беляева, Л. К. Гейхман, О. Н. Камшилова, Н. Л. Шубина. – Санкт-Петербург: ООО «Книжный дом», 2021. – 78 с.

9) Беляева Л. Н. Письменная научная коммуникация в русском языковом пространстве: монография / Л. Н. Беляева, Л. К. Гейхман, О. Н. Камшилова, Н. Л. Шубина. – Санкт-Петербург: ООО «Книжный дом», 2021. – 154 с.

10) Беляева Л. Н., Камшилова О. Н. Сетевое пространство и научный стиль: терминологические неологизмы / Л. Н. Беляева, О. Н. Камшилова // Общество. Коммуникация. Образование. – 2020. – Том 11, № 2. – С. 108-116.

11) Беляева Л. Н. Терминологические проблемы перевода / Л. Н. Беляева // Контрастивные исследования и прикладная лингвистика: материалы международной научной конференции, Минск, 29-30 октября 2014 года. – Минск: МГЛУ, 2015. – С. 8-12.

12) Беляева Л. Н. Технический перевод, терминология и информационные технологии / Л. Н. Беляева // Терминология и знание: материалы III международного симпозиума. – Москва: Вест-Консалтинг, 2013. – С. 209-219.

13) Боженкова Р. К. Русский язык и культура речи: учебник / Р. К. Боженкова, Н. А. Боженкова, В. М. Шаклеин. – 5-е издание, стереотипное. – Москва: ФЛИНТА, 2019. – 608 с.

14) Бойко Л. Б. Идиоматичность перевода как оценочная категория / Л. Б. Бойко // Вестник Балтийского федерального университета им. И. Канта. Серия: Филология, педагогика, психология. – 2023. – № 2. – С. 5-16.

15) Бойко Л. Б., Харитонова Е. В. В поисках утраченного и обретенного: заметки о переводе / Л. Б. Бойко, Е. В. Харитонова. – Москва: Языкознание, 2024. – 255 с.

16) Болсуновская Л. М. Виды дискурсивных маркеров в русскоязычных и англоязычных научных статьях по геологии и нефтегазовому делу / Л. М. Болсуновская, Ю. А. Зеремская, Н. В. Дубровская // Вестник ТГПУ. – 2015. – № 4 (157). – С. 117-123.

17) Бреус Е. В. Основы теории и практики перевода с русского языка на английский: учебное пособие / Е. В. Бреус. – 2-е издание, исправленное и дополненное. – Москва: Издательство УРАО, 2000. – 208 с.

18) Бутенко Ю. И., Авагян Н. А. Способы выражения модальности в параллельных текстах стандартов (на примере нормативной базы программной инженерии) / Ю. И. Бутенко, Н. А. Авагян // Вестник ВГУ. Серия: Лингвистика и межкультурная коммуникация. – 2021. – № 2. – С. 46-55.

19) Вагнер В. Н. Методика преподавания русского языка англоговорящим и франкоговорящим (фонетика, графика, части речи): автореферат диссертации на соискание ученой степени доктора педагогических наук: специальность 13.00.02: защищена 21.04.1997 / Вагнер В. Н.; Российский университет дружбы народов. – Москва, 1997. – 43 с.

20) Валеева Н. Г. Употребление дискурсивных маркеров в научных статьях по экологии на русском и английском языке / Н. Г. Валеева, М. А. Руднева // Вестник РУДН. Серия: Экология и безопасность жизнедеятельности. – 2016. – № 4. – С. 109-115.

21) Дейк, Т. А. ван., Кинч В. Стратегии понимания связного текста / Т. А. ван Дейк, В. Кинч // Новое в зарубежной лингвистике. Выпуск XXIII. Когнитивные аспекты языка. – Москва: Прогресс, 1988. – С. 153-211.

22) Ван Ли. Грамматическая специфика научных текстов русского языка / Ван Ли // Вестник Кемеровского государственного университета культуры и искусств. – 2011. – № 17. – С. 107-116.

23) Васильева А. Н. Курс лекций по стилистике русского языка: научный стиль речи / А. Н. Васильева. – Москва: Русский язык, 1976. – 192 с.

- 24) Вежбицкая А. Метатекст в тексте // А. Вежбицкая / Новое в зарубежной лингвистике. Выпуск VIII. Лингвистика текста. – Москва: Прогресс, 1978. – С. 402-421.
- 25) Викторова Е. Ю. Прагматическая полифункциональность дискурсивных слов / Е. Ю. Викторова // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2014. – № 2. – С. 18-22.
- 26) Викторова Е. Ю. Роль количественного подхода в изучении функционирования дискурсивных слов / Е. Ю. Викторова // Филология и человек. – 2013. – № 4. – С. 34-46.
- 27) Виноградов В. В. Итоги обсуждения вопросов стилистики / В. В. Виноградов // Вопросы языкознания. – 1955. – № 1. – С. 60-87.
- 28) Владимирова Т. Л. Язык и стиль научного текста: учебное пособие / Т. Л. Владимирова. – Томск: Издательство Томского политехнического университета. – 2010. – 80 с.
- 29) Власенко Н. И., Толмачёва И. А. Промежуточный язык как "третья система" в процессе взаимодействия языков / Н. И. Власенко, И. А. Толмачёва // Язык для специальных целей: система, функции, среда: сборник научных статей VI международной научно-практической конференции. – Курск: Юго-Западный государственный университет, 2016. – С. 96-101.
- 30) Волкова А. А. Функциональная противоречивость перевода в социокультурных практиках / А. А. Волкова // Вестник Бурятского государственного университета. Философия. – 2021. – Выпуск 4. – С. 62-69.
- 31) Гальперин И. Р. Очерки по стилистике английского языка / И. Р. Гальперин. – Москва: Издательство литературы на иностранных языках, 1958. – 459 с.
- 32) Гальперин И. Р. Текст как объект лингвистического исследования / И. Р. Гальперин. – 5-е издание, стереотипное. – Москва: КомКнига, 2007. – 144 с.

- 33) Гарбовский Н. К. Перевод и «Переводной дискурс» / Н. К. Гарбовский // Вестник Московского университета. Серия 22. Теория перевода. – 2011. – № 4. – С. 3-19.
- 34) Гарбовский Н. К. Русский переводной дискурс: миф или реальность / Н. К. Гарбовский // Русский язык и культура в зеркале перевода: материалы III Международной научно-практической конференции «Язык и культура в зеркале перевода», 25-29 апреля 2012 года. – С. 130-136.
- 35) Гарбовский Н. К., Костикова О. И. Перевод и общество / Н. К. Гарбовский, О. И. Костикова // Вестник Московского университета. Серия 22. Теория перевода. – 2018. – № 1. – С. 17-40.
- 36) Гринберг Дж. Меморандум о языковых универсалиях / Дж. Гринберг, Ч. Осгуд, Дж. Дженкинс // Новое в лингвистике. – Выпуск 5. – Москва: Прогресс, 1970. – С. 31-44.
- 37) Громова Л. Г. Формирование билингвизма при обучении переводу финских студентов в русле теории языковых контактов / Л. Г. Громова // Язык. Культура. Коммуникация: сборник научных трудов / под редакцией В. З. Демьянкова. – Москва: МГУ, 2015. – С. 221-224.
- 38) Гуревич В. В. English Stylistics. Стилистика английского языка: учебное пособие / В. В. Гуревич. – 10-е издание, стереотипное. – Москва: ФЛИНТА, 2024. – 72 с.
- 39) Залевская А. А. Вопросы теории двуязычия / А. А. Залевская. – Москва: Директ-Медиа, 2013. – 144 с.
- 40) Захаров В. П. Корпусная лингвистика: учебно-методическое пособие / В. П. Захаров. – Санкт-Петербург, 2005. – 48 с.
- 41) Захаров В. П., Хохлова М. В. Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке / В. П. Захаров, М. В. Хохлова // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной Международной конференции «Диалог». – 2010. – Выпуск 9 (16). – С. 137-143.

42) Захарова М. А. NORMS OF TRANSLATION (нормы перевода) / М. А. Захарова // Основные понятия англоязычного переводоведения: терминологический словарь-справочник / под редакцией М. Б. Раренко. – 2011. – № 2011. – С. 138-143.

43) Земская Е. А. Современный русский язык. Словообразование: учебное пособие / Е. А. Земская. – 3-е издание, исправленное и дополненное. – Москва: Флинта: Наука, 2011. – 328 с.

44) Иваницкая Е. В. Трансформация научного стиля в условиях меняющейся коммуникационной среды. Особенности современной научной статьи / Е. В. Иваницкая // Язык и текст langpsy.ru. – 2016. – Том 3, № 2. – С. 62-75.

45) Камышова Е. А. Лексико-грамматические особенности научного текста на английском языке / Е. А. Камышова, О. А. Колмакова // Язык, культура, наука и техника страны изучаемого языка: материалы Всероссийской научно-практической студенческой конференции, посвящённой 85-летию ИРНИТУ. – Иркутск: Издательство ИРНИТУ, 2015. – С. 196-206.

46) Кантышева Н. Г. Индикаторы переводного русскоязычного дискурса: текстовые и лексические параметры / Н. Г. Кантышева // Вестник Московского университета. Серия 22. Теория перевода. – 2021. – № 3. – С. 135-149.

47) Кашкин И. А. Для читателя-современника / И. А. Кашкин. – Москва: Советский писатель, 1977. – 560 с.

48) Когут С. В. Дискурсивные маркеры в письменном научном дискурсе / С. В. Когут // Сибирский филологический журнал. – 2016. – № 2. – С. 157-163.

49) Когут С. В. Дискурсивные маркеры как отражение своеобразия естественнонаучной и научно-гуманитарной картин мира / С. В. Когут // Филологические науки. Вопросы теории и практики. – 2014. – № 10-1. – С. 101-106.

50) Кожина М. Н. Стилистика русского языка: учебник / М. Н. Кожина, Л. Р. Дускаева, В. А. Салимовский. – Москва: ФЛИНТА: Наука, 2008. – 464 с.

51) Козлова Л. А. Культурный код и проблема его ретрансляции в процессе перевода / Л. А. Козлова // Инновационные технологии и подходы в межкультурной коммуникации, лингвистике и лингводидактике: материалы II Всероссийской

научной конференции. – Барнаул: Алтайский государственный педагогический университет, 2024. – С. 267-277.

52) Комиссаров В. Н. Общая теория перевода / В. Н. Комиссаров. – Москва: ЧеРо, 1999. – 133 с.

53) Комиссаров В. Н. Теория перевода (лингвистические аспекты) / В. Н. Комиссаров. – Москва: Высшая школа, 1990. – 253 с.

54) Костикова О. И. Культуротворческая функция перевода: гуманитарная перспектива / О. И. Костикова // Новое и традиционное в переводоведении и преподавании русского языка как иностранного. – Баня-Лука: Panevropski univerzitet "APEIRON", 2024. – С. 179-187.

55) Костикова О. И. Переводческая критика: «прозрачность» vs «зеркальность» / О. И. Костикова // Вестник Московского университета. Серия 22. Теория перевода. – 2010. – № 3. – С. 41-54.

56) Котюрова М. П. Общенаучная лексика / М. П. Котюрова // Стилистический энциклопедический словарь русского языка / под редакцией М. Н. Кожинной. – 2-е издание, стереотипное. – Москва: Флинта: Наука, 2011. – 696 с.

57) Краснопеева Е. С. Характеристика особенностей русскоязычного переводного дискурса с позиций корпусного переводоведения / Е. С. Краснопеева // Вестник Тюменского государственного университета. Гуманитарные исследования. Humanitates. – 2015. – Том 1, № 4. – С. 65-73.

58) Краснопеева Е. С. Лексические особенности русскоязычного переводного дискурса: корпусное сравнительно-сопоставительное исследование на материале современной художественной прозы: автореферат диссертации на соискание ученой степени кандидата филологических наук: специальность 10.02.20: защищена 08.04.2016 / Краснопеева Е. С.; Тюменский государственный университет. – Тюмень, 2016. – 23 с.

59) Куниловская М. А. Переводческие универсалии как центральное направление корпусного переводоведения / М. А. Куниловская // Проблемы перевода региональных авторов, в том числе представителей коренных малочисленных народов Севера Западной Сибири, на иностранные языки:

региональная конференция, Ханты-Мансийск, 18 декабря 2015 года: материалы конференции. – Екатеринбург: Баско, 2015. – С. 13-19.

60) Куниловская М. А. Перспективы исследования лингвистических признаков русскоязычного переводного дискурса / М. А. Куниловская, Т. А. Ильющенья, М. А. Ковязина // Вестник ПНИПУ. Проблемы языкознания и педагогики. – 2017. – № 2. – С. 17-27.

61) Кушнина Л. В. Интерпретирующая функция переводческого понимания как отражение индивидуальной переводческой картины мира / Л. В. Кушнина, С. В. Литвинова, И. В. Перлова, К. В. Пермякова // Лингвориторическая парадигма: теоретические и прикладные аспекты. – 2019. – № 24. – С. 48-56.

62) Латышев Л. К. Технология перевода: учебник и практикум для вузов / Л. К. Латышев, Н. Ю. Северова. – 4-е издание, переработанное и дополненное. – Москва: Издательство Юрайт, 2025. – 263 с.

63) Латышев Л. К., Семенов А. Л. Перевод: теория, практика и методика преподавания: учебное пособие для студентов переводческих факультетов высших учебных заведений / Л. К. Латышев, А. Л. Семенов. – Москва: Издательский центр «Академия», 2003. – 192 с.

64) Лекомцева И. А. О «Третьем языке» в переводе / И. А. Лекомцева // Многоязычие в образовательном пространстве. – 2016. – № 8. – С. 99-104.

65) Лекомцева И. А. Возможности поиска языковых средств выражения семантических категорий при переводе с помощью корпусных методов исследования / И. А. Лекомцева, А. Х. Абдульманова, М. Н. Куликова // Дискурс. – 2022. – Том 8, № 1. – С. 158-167.

66) Леонтович О. А. Проблема ретрансляции и адаптации культурных смыслов / О. А. Леонтович // Вестник Московского университета. Серия 19. Лингвистика и межкультурная коммуникация. – 2008. – № 2. – С. 18-24.

67) Ляшевская О. Н. Новый частотный словарь русской лексики / О. Н. Ляшевская, С. А. Шаров. – Москва: Азбуковник, 2009. – URL: <http://dict.ruslang.ru/freq.php> (дата обращения: 13.01.2024). – Режим доступа: свободный.

68) Мельник В. И. Переводческие корпусы - что они могут сказать нам о переводе и его преподавании? / В. И. Мельник // Вестник Пермского национального исследовательского политехнического университета. Проблемы языкознания и педагогики. – 2015. – № 2 (12). – С. 32-36.

69) Минакова Д. В., Краснова Е. В. Использование приёма перестановки при переводе англоязычных научно-технических текстов / Д. В. Минакова, Е. В. Краснова // Молодой исследователь Дона. – 2017. – № 4 (7). – С. 141-146.

70) Михайлова А. Г. Корпусный подход к переводу профессиональных текстов / А. Г. Михайлова // Цифровая гуманитаристика и технологии в образовании (ДНТЕ 2023): материалы конференции. – Москва: Московский государственный психолого-педагогический университет, 2023. – С. 367-380.

71) Оборнева И. В. Автоматизированная оценка сложности учебных текстов на основе статистических параметров: диссертация на соискание ученой степени кандидата педагогических наук: специальность 13.00.02: защищена 30.03.2006 / Оборнева И. В.; Московский городской педагогический университет. – Москва, 2006. – 165 л.

72) Паршин А. Теория и практика перевода: учебное пособие для студентов лингвистических факультетов вузов / А. Паршин. – Саратов: Издательство СГУ: Русский язык, 2000. – 202 с.

73) Плещенко Т. П. Основы стилистики и культуры речи: учебное пособие для студентов вузов / Т. П. Плещенко, Н. В. Федотова, Р. Г. Чечет; под редакцией П. П. Шубы. – Минск: НТООО «ТетраСистемс», 1999. – 240 с.

74) Пилипчук С. В. Промежуточный язык как некоторый допустимый этап на пути овладения вторым языком / С. В. Пилипчук // Слово и текст, психолингвистический подход / под редакцией А. А. Залевской. – Тверь: Тверской государственный университет, 2003. – С. 122-126.

75) Поспелова Т. Б. Лексико-грамматические особенности научного стиля в английском языке / Т. Б. Поспелова // Homo Loquens: Актуальные вопросы лингвистики и методики преподавания иностранных языков. – Выпуск 5. – Санкт-Петербург: Астерион, 2013. – С. 277-285.

76) Прунч Э. Пути развития западного переводоведения. От языковой асимметрии к политической / Э. Прунч; перевод с немецкого. – Москва: Р. Валент, 2015. – 512 с.

77) Пумпянский А. Л. Введение в практику перевода научной и технической литературы на английский язык / А. Л. Пумпянский. – 2-е издание, дополненное. – Москва: Наука, 1981. – 344 с.

78) Ревзин И. И., Розенцвейг В. Ю. Основы общего и машинного перевода / И. И. Ревзин, В. Ю. Розенцвейг. – Москва: Высшая школа, 1963. – 243 с.

79) Рецкер Я. И. Теория перевода и переводческая практика. Очерки лингвистической теории перевода / Я. И. Рецкер; дополнения и комментарии Д. И. Ермоловича. – 5-е издание, исправленное и дополненное. – Москва: Аудитория, 2016. – 244 с.

80) Савко М. В. Языковое взаимодействие при аудиовизуальном переводе и корпусные исследования в переводоведении / М. В. Савко // Мова і культура. Науковий журнал. – Киев, 2010. – С. 353-360.

81) Савчук С. О. Национальный корпус русского языка 2.0: новые возможности и перспективы развития / С. О. Савчук, Т. А. Архангельский, А. А. Бонч-Осмоловская, О. В. Дониная, Ю. Н. Кузнецова, О. Н. Ляшевская, Б. В. Орехов, М. В. Подрядчикова // Вопросы языкознания. – 2024. – № 2. – С. 7-34.

82) Сдобников В. В. Переводоведение сегодня: вечные проблемы и новые вызовы / В. В. Сдобников // Russian Journal of Linguistics. – 2019. – Том 23, № 2. – С. 295-327.

83) Сдобников В. В. Теория перевода (коммуникативно-функциональный подход): учебник для студентов лингвистических вузов и факультетов иностранных языков / В. В. Сдобников, К. Е. Калинин, О. В. Петрова. – 2-е издание, переработанное. – Москва: Издательство ВКН, 2019. – 512 с.

84) Тестелец Я. Г. Введение в общий синтаксис / Я. Г. Тестелец. – Москва: РГГУ, 2001. – 797 с.

85) Тынянов Ю. Н. Поэтика. История литературы. Кино / Ю. Н. Тынянов. – Москва: Наука, 1977. – 574 с.

86) Умерова М. В. Переводческие трансформации, отражающие особенности языка оригинала в тексте перевода / М. В. Умерова // Актуальные проблемы английской лингвистики и лингводидактики: сборник научных трудов. – Выпуск 9. – Москва: Национальный книжный центр, 2010. – С. 138-148.

87) Умерова М. В. Полноценность и адекватность перевода в современном переводоведении / М. В. Умерова // Университетское переводоведение. Выпуск 8: материалы III юбилейной международной научной конференции по переводоведению «Федоровские чтения», посвященной 100-летию со дня рождения А. В. Федорова, 19-21 октября 2006 года. – Санкт-Петербург: Филологический факультет СПбГУ, 2007. – С. 468-474.

88) Успенский Б. А. Проблема универсалий в языкознании / Б. А. Успенский // Новое в лингвистике. – Выпуск 5. – Москва: Прогресс, 1970. – С. 5-30.

89) Федоров А. В. Основы общей теории перевода (лингвистические проблемы): для институтов и факультетов иностранных языков: учебное пособие / А. В. Федоров. – 5-е издание. – Санкт-Петербург: Филологический факультет СПбГУ; Москва: ООО «Издательский Дом «ФИЛОЛОГИЯ ТРИ», 2002. – 416 с.

90) Чернявская В. Е. Интерпретация научного текста: учебное пособие / В. Е. Чернявская. – 4-е издание. – Москва: Издательство ЛКИ, 2007. – 128 с.

91) Чуковский К. И. Высокое искусство / К. И. Чуковский. – Москва: Художественная литература, 1941. – 259 с.

92) Шаклеин В. М., Цяньминь Ч. Интерязык: в поисках уточнения термина / В. М. Шаклеин, Ч. Цяньминь // Международный научно-исследовательский журнал. – 2017. – № 5-1 (59). – С. 170-175.

93) Шакирова Н. Р. У. Фроли: "Предварительные рассуждения о теории перевода", или перевод как самостоятельная кодовая система / Н. Р. Шакирова // Электронный научный журнал. – 2016. – № 8 (11). – С. 82-87.

94) Швейцер А. Д. Теория перевода: статус, проблемы, аспекты / А. Д. Швейцер. – Москва: Наука, 1988. – 215 с.

95) Шелестюк Е. В., Гриценко Э. Д. О форенизации и доместикации в переводе и возможностях их лингвистической оценки / Е. В. Шелестюк, Э. Д. Гриценко // Вестник ЧелГУ. – 2016. – № 4 (386). – С. 202-207.

96) Щемелева И. Ю. Местоимения как средство передачи авторской позиции в английском и русском научных текстах / И. Ю. Щемелева // Вестник СПбГУ. Серия 9. – 2013. – Выпуск 4. – С. 143-153.

97) Baker M. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research / M. Baker // Target. – 1995. – Volume 7, № 2. – P. 223-243.

98) Baker M. Corpus linguistics and translation studies: Implications and applications / M. Baker // Text and technology: In Honour of John Sinclair / edited by M. Baker, G. Francis, E. Tognini-Bonelli. – Amsterdam; Philadelphia: John Benjamins, 1993. – P. 223-250.

99) Baker M. Corpus-based translation studies: the challenges that lie ahead / M. Baker // Terminology, LSP and Translation. Studies in language engineering in honour of Juan C. Sager / edited by H. Somers. – Amsterdam; Philadelphia: John Benjamins, 1996. – P. 175-186.

100) Banks D. The extent to which the passive voice is used in the scientific journal article, 1985-2015 / D. Banks // Functional Linguist. – 2017. – Volume 4, № 12.

101) Baroni M. A New Approach to the Study of Translationese: Machine-learning the Difference between Original and Translated Text / M. Baroni, S. Bernardini // Literary and Linguistic Computing. – 2006. – Volume 21, № 3. – P. 259-274.

102) Baroni M. A Preliminary Analysis of Collocational Differences in Monolingual Comparable Corpora / M. Baroni, S. Bernardini // Proceedings of Corpus Linguistics 2003, Lancaster, UK, March. – Lancaster, 2003.

103) Ben-Ari N. The Ambivalent Case of Repetitions in Literary Translation. Avoiding Repetitions: a "Universal" of Translation? / N. Ben-Ari // Meta. – 1998. – Volume 43, № 1. – P. 68-78.

104) Bello I. Cognitive implications of nominalizations in the advancement of scientific discourse / I. Bello // *International Journal of English Studies*. – 2016. – Volume 16, № 2. – P. 1-23.

105) Bernardini S., Baroni M. Spotting translationese: A corpus-driven approach using support vector machines / S. Bernardini, M. Baroni // *Proceedings of Corpus Linguistics Conference Series 2005*. – Volume 1. – Birmingham: University of Birmingham, 2005. – P. 1-12.

106) Bernardini S., Zanettin F. When is a universal not a universal?: Some limits of current corpus-based methodologies for the investigation of translation universals / S. Bernardini, F. Zanettin // *Translation Universals: Do they exist?* / edited by A. Mauranen, P. Kujamäki. – Amsterdam; Philadelphia: John Benjamins Publishing Company, 2004. – P. 51-62.

107) Biber D. *University Language: A corpus-based study of spoken and written registers* / D. Biber. – Amsterdam; Philadelphia: John Benjamins Publishing Company, 2006. – 261 p.

108) *Grammar of Spoken and Written English* / D. Biber, S. Johansson, G. N. Leech, S. Conrad, E. Finegan. – Amsterdam; Philadelphia: John Benjamins, 2021. – 1220 p.

109) How human is machine translationese? Comparing human and machine translations of text and speech / Y. Bizzoni, T. S. Juzek, C. España-Bonet, C. D. Chowdhury, J. Genabith, E. Teich // *Proceedings of the 17th International Conference on Spoken Language Translation*, Online, July 2020. – Association for Computational Linguistics, 2020. – P. 280-290.

110) Blum-Kulka S. Shifts of cohesion and coherence in translation / S. Blum-Kulka // *The Translation Studies Reader* / edited by L. Venuti, M. Baker. – Volume 216. – London: Routledge, 2000. – P. 298-313.

111) Blum-Kulka S., Levenston E. A. Universals of lexical simplification / S. Blum-Kulka, E. A. Levenston // *Strategies in Interlanguage Communication* / edited by C. Faerch, G. Kasper. – London; New York: Longman, 1983. – P. 119-139.

112) Bloor T., Bloor M. The functional analysis of English: A Hallidayan approach / T. Bloor, M. Bloor. – 3rd edition. – London: Routledge, 2013. – 320 p.

113) Corral Á., Serra I. The Brevity Law as a Scaling Law, and a Possible Origin of Zipf's Law for Word Frequencies / Á. Corral, I. Serra // Entropy. – 2020. – Volume 22, № 2. – P. 1-14.

114) Chesterman A. Hypotheses about translation universals / A. Chesterman // Claims, Changes and Challenges in Translation Studies: Selected contributions from the EST Congress, Copenhagen 2001 / edited by G. Hansen, K. Malmkjær, D. Gile. – Amsterdam; Philadelphia: John Benjamins, 2004a. – P. 1-13.

115) Chesterman A. Beyond the particular / A. Chesterman // Translation Universals: Do they exist? / edited by A. Mauranen, P. Kujamäki. – Amsterdam; Philadelphia: John Benjamins, 2004b. – P. 33-49.

116) Chesterman A. Why study translation universals / A. Chesterman // Acta Translatologica Helsingiensia. – 2010. – Volume 1. – P. 38-48.

117) Church K., Hanks P. Word association norms, mutual information, and lexicography / K. Church, P. Hanks // Computational Linguistics. – 1990. – Volume 16, № 1. – P. 22-29.

118) Comparing feature-engineering and feature-learning approaches for multilingual translationese classification / D. Pylypenko, K. Amponsah-Kaakyire, K. Dutta Chowdhury, J. van Genabith, C. España-Bonet // Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Punta Cana. – 2021. – P. 8596-8611.

119) Ding D. Rationality reborn: Historical roots of the passive voice in scientific discourse / D. Ding // Essays in the study of scientific discourse: Methods, practice, and pedagogy / edited by J. T. Battalio. – Stamford, CT: Ablex, 1998. – P. 117-135.

120) Duff A. The Third Language: Recurrent Problems of Translation into English / A. Duff. – Oxford: Pergamon Press, 1981. – 138 p.

121) Even-Zohar I. Polysystem theory / I. Even-Zohar // Poetics today. – 1979. – Volume 1, № 1/2. – P. 287-310.

122) Even-Zohar I. The Position of Translated Literature within the Literary Polysystem / I. Even-Zohar // *Translation Across Cultures* / edited by G. Toury. – New Delhi: Bahri, 1987. – P. 107-115.

123) Ferreira F. The misinterpretation of noncanonical sentences / F. Ferreira // *Cognitive Psychology*. – 2003. – Volume 47, № 2. – P. 164-203.

124) François T., Lefer M.-A. Revisiting simplification in corpus-based translation studies: Insights from readability research / T. François, M.-A. Lefer // *Meta*. – 2022. – Volume 67, № 1. – P. 50-70.

125) Frawley W. Prolegomenon to a theory of translation / W. Frawley // *The Translation Studies Reader* / edited by L. Venuti, M. Baker. – London: Routledge, 2000. – P. 252-263.

126) Gellerstam M. Translationese in Swedish novels translated from English / M. Gellerstam // *Translation studies in Scandinavia*. – 1986. – P. 88-95.

127) Gellerstam M. Fingerprints in Translation / M. Gellerstam // *In and Out of English: For Better, For Worse?* / edited by G. Anderman, M. Rogers. – Clevedon: Multilingual Matters, 2005. – P. 201-213.

128) Gellerstam M. Translations as a Source for Cross-linguistic Studies / M. Gellerstam // *Languages in Contrast* / edited by B. A. K. Aijmer, M. Johansson. – Lund: CWK Gleerup, 1996. – P. 53-62.

129) Graham Y. Translationese in machine translation evaluation / Y. Graham, B. Haddow, P. Koehn // arXiv preprint arXiv:1906.09833. – 2019.

130) Graham Y. Statistical power and translationese in machine translation evaluation / Y. Graham, B. Haddow, P. Koehn // *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. – 2020. – P. 72-81.

131) Granger S. Comparable and translation corpora in cross-linguistic research. Design, analysis and applications / S. Granger // *Journal of Shanghai Jiaotong University*. – 2010. – Volume 2, № 1. – P. 14-21.

132) Granger S. Tracking the third code: A cross-linguistic corpus-driven approach to metadiscursive markers / S. Granger // *The Corpus Linguistics Discourse* / edited by A. Cermakova, M. Mahlberg. – John Benjamins, 2018. – P. 185-204.

- 133) Grieve J. Quantitative authorship attribution: An evaluation of techniques / J. Grieve // *Literary and Linguistic Computing*. – 2007. – Volume 22, № 3. – P. 251-270.
- 134) Halliday M. A. K., Hasan R. *Cohesion in English* / M. A. K. Halliday, R. Hasan. – London: Longman, 1976. – 374 p.
- 135) Halliday M. A. K., Martin J. R. *Writing Science: Literacy and Discursive Power* / M. A. K. Halliday, J. R. Martin. – London; Washington, D.C.: The Falmer Press, 2003. – 309 p.
- 136) Halliday M. A. K., Matthiessen C. M. I. M. *Halliday's Introduction to Functional Grammar* / M. A. K. Halliday, C. M. I. M. Matthiessen. – 4th edition. – London: Routledge, 2014. – 786 p.
- 137) Hoek J. Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study / J. Hoek, S. Zufferey, J. Evers-Vermeul, T. J. M. Sanders // *Journal of Pragmatics*. – 2017. – Volume 121. – P. 113-131.
- 138) Honoré A. Some simple measures of richness of vocabulary / A. Honoré // *Association of Literary and Linguistic Computing Bulletin*. – 1979. – Volume 7, № 2. – P. 172-177.
- 139) Hu H., Kübler S. Investigating translated Chinese and its variants using machine learning / H. Hu, S. Kübler // *Natural Language Engineering*. – 2021. – Volume 27, № 3. – P. 339-372.
- 140) Hvelplund K. T. *Allocation of Cognitive Resources in Translation: An Eye-tracking and Key-logging Study: doctoral dissertation* / Hvelplund K. T.; Copenhagen Business School. – Copenhagen: Samfundslitteratur, 2011. – URL: <https://research.cbs.dk/en/publications/allocation-of-cognitive-resources-in-translation-an-eye-tracking-/> (дата обращения: 10.02.2025). – Режим доступа: свободный.
- 141) Hyland K. *English for Academic Purposes: An Advanced Resource Book* / K. Hyland. – London: Routledge, 2006. – 356 p.
- 142) Hyland K. *Hedging in scientific research articles* / K. Hyland. – Amsterdam; Philadelphia: John Benjamins Publishing Company, 1998. – 307 p.

143) Hyland K., Jiang F. Academic naming: Changing patterns of noun use in research writing / K. Hyland, F. Jiang // *Journal of English Linguistics*. – 2021. – Volume 49, № 3. – P. 255-282.

144) Ilisei I.-N. A Machine Learning Approach to the Identification of Translational Language: An Inquiry into Translationese Learning Models: PhD thesis / I.-N. Ilisei. – Wolverhampton: University of Wolverhampton, 2012.

145) Ilisei I. Identification of translationese: A machine learning approach / I. Ilisei, D. Inkpen, G. Corpas Pastor, R. Mitkov // *International conference on intelligent text processing and computational linguistics*. – Berlin; Heidelberg: Springer Berlin Heidelberg, 2010. – P. 503-511.

146) Jiang Y., Niu J. A corpus-based search for machine translationese in terms of discourse coherence / Y. Jiang, J. Niu // *Across Languages and Cultures*. – 2022. – Volume 23, № 2. – P. 148-166.

147) Kayumova A. Cross-linguistic Influence on Punctuation: The English and Russian Dash in Contact / A. Kayumova // *Journal of International Scientific publications*. – 2015. – Volume 9. – P.154-166.

148) Kazemi F. Nominalization in applied linguistics and medical research articles: comparing native and non-native academic Writers / F. Kazemi // *Journal of English Language Pedagogy and Practice*. – 2025. – Volume 17, № 2. – P. 29-55.

149) Kazemian B. Ideational grammatical metaphor in scientific texts: A Hallidayan perspective / B. Kazemian, B. Behnam, N. Ghafoori // *International Journal of Linguistics*. – 2013. – Volume 5, № 4. – P. 146-168.

150) Kenny D. *Lexis and Creativity in Translation: A Corpus Based Approach* / D. Kenny. – 1st edition. – London: Routledge, 2001.

151) Kish L. *Survey Sampling* / L. Kish. – New York; London: John Wiley & Sons, Inc., 1965. – 643 p.

152) Koppel M., Ordan N. Translationese and Its Dialects / M. Koppel, N. Ordan // *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, 2011*. – P. 1318-1326.

153) Kotait R. Richness Lost in Machine Translationese / R. Kotait // The Egyptian Journal of Language Engineering. – 2024. – Volume 11, № 1. – P. 66-85.

154) Kunilovskaya M. Frequency distribution and usage of discourse markers in English to Russian translation: learner vs. professional translational behavior / M. Kunilovskaya // Corpus Linguistics Conference 2017. – Birmingham: University of Birmingham, 2017. – URL: <https://www.birmingham.ac.uk/research/centres-institutes/centre-for-corpus-research/corpus-linguistics-conference-2017> (дата обращения: 10.07.2024). – Режим доступа: свободный.

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

156) Kunilovskaya M., Corpas Pastor G. Translationese and register variation in English-to-Russian professional translation / M. Kunilovskaya, G. Corpas Pastor // New Perspectives on Corpus Translation Studies / edited by V. X. Wang, D. Li, L. Lim. – Singapore: Springer Nature Singapore Pte Ltd, 2021. – P. 133-180.

157) Kunilovskaya M., Kutuzov A. A quantitative study of translational Russian (based on a translational learner corpus) / M. Kunilovskaya, A. Kutuzov // Corpus Linguistics 2015: Proceedings of 7th International Conference. – Saint Petersburg: Saint Petersburg State University, 2015. – P. 33-40.

158) Kunilovskaya M., Kutuzov A. Testing target text fluency: A machine learning approach to detecting syntactic translationese in English-Russian translation / M. Kunilovskaya, A. Kutuzov // New perspectives on cohesion and coherence / edited by K. Menzel, E. Lapshinova-Koltunski, K. Kunz. – Berlin: Language Science Press, 2017. – P. 69-94.

159) Kunilovskaya M., Kutuzov A. Universal Dependencies-based syntactic features in detecting human translation varieties / M. Kunilovskaya, A. Kutuzov // Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories (TLT16). – 2017. – С. 27–36.

160) Kunilovskaya M., Lapshinova-Koltunski E. Translationese Features as Indicators of Quality in English-Russian Human Translation / M. Kunilovskaya, E. Lapshinova-Koltunski // Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019), Varna, Bulgaria. – Shoumen, Bulgaria: Incoma Ltd., 2019. – P. 47-56.

161) Kunilovskaya M. Translationese in Russian Literary Texts / M. Kunilovskaya, E. Lapshinova-Koltunski, R. Mitkov // Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature. – 2021. – P. 101-112.

162) Kurokawa D. Automatic detection of translated text and its impact on machine translation / D. Kurokawa, C. Goutte, P. Isabelle // Proceedings of MT-Summit XII. – 2009. – P. 81-88.

163) Kutuzov A., Kuzmenko E. WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models / A. Kutuzov, E. Kuzmenko // Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science / edited by D. Ignatov et al. – Volume 661. – Cham: Springer, 2017. – P. 155-161.

164) Lapshinova-Koltunski E. Variation in translation: evidence from corpora / E. Lapshinova-Koltunski // New directions in corpus-based translation studies / edited by C. Fantinuoli, F. Zanettin. – Berlin: Language Science Press, 2015. – P. 93-114.

165) Laviosa S. Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose / S. Laviosa // Meta. – 1998. – Volume 43, № 4. – P. 557-570.

166) Laviosa S. How comparable can 'comparable corpora' be? / S. Laviosa // Target. – 1997. – Volume 9, № 2. – P. 289-320.

167) Laviosa S. Universals / S. Laviosa // Routledge Encyclopedia of translation studies / edited by M. Baker and G. Saldanha. – 2nd edition. – London: Routledge, 2009. – P. 306-311.

168) Laviosa-Braithwaite S. Investigating Simplification in an English Comparable Corpus of Newspaper Articles / S. Laviosa-Braithwaite // Transferring necesses

est. Proceedings of the Second International Conference on Current Trends in Studies of Translation and Interpreting. – Budapest: Scholastica, 1997. – P. 531-540.

169) Leech G. N. Meaning and the English Verb / G. N. Leech. – 3rd edition. – Abingdon; New York: Routledge, 2013. – 167 p.

170) Leong P. A. The passive voice in scholarly writing: A diachronic look at science and history / P. A. Leong // Finnish Journal of Linguistics. – 2021. – Volume 34. – P. 77-102.

171) Liu Y. Measuring the difficulty of text translation: the combination of text-focused and translator-oriented approaches / Y. Liu, B. Zheng, H. Zhou // Target. – 2019. – Volume 31, № 1. – P. 125-149.

172) Liu K., Afzaal M. Syntactic complexity in translated and non-translated texts: A corpus-based study of simplification / K. Liu, M. Afzaal // PLoS One. – 2021. – Volume 16, № 6. – P. 1-20

173) Malmkjaer K. Punctuation in Hans Christian Andersen's Stories and in their Translations into English / K. Malmkjaer // Nonverbal Communication and Translation: New Perspectives and Challenges in Literature, Interpretation and the Media. – Philadelphia: John Benjamins Publishing Company, 1997. – P. 151-162.

174) Manakhimova S. Fine-Grained Evaluation of English-Russian MT in 2025: Linguistic Challenges Mirroring Human Translator Training / S. Manakhimova, M. Kunilovskaya, E. Lapshinova-Koltunski, E. Avramidis // Proceedings of the Tenth Conference on Machine Translation. – Suzhou, China: Association for Computational Linguistics, 2025. – P. 866-877.

175) Mann H. B., Whitney D. R. On a test of whether one of two random variables is stochastically larger than the other / H. B. Mann, D. R. Whitney // Annals of Mathematical Statistics. – 1947. – Volume 18. – P. 50-60.

176) Manning C. D., Schütze H. Foundations of Statistical Natural Language Processing / C. D. Manning, H. Schütze. – Cambridge, Massachusetts: The MIT Press, 1999. – 680 p.

177) Mauranen A. Corpora, universals and interference / A. Mauranen // Translation Universals: Do they exist? / edited by A. Mauranen, P. Kujamäki. – Amsterdam; Philadelphia: John Benjamins Publishing Company, 2004. – P. 65-82.

178) Mauranen A. Strange Strings in Translated Language: A Study on Corpora / A. Mauranen // Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects / edited by M. Olohan. – Manchester: St. Jerome Publishing, 2000. – P. 119-141.

179) May R. Sensible Elocution / R. May // The Translator. – 1997. – Volume 3, № 1. – P. 1-20.

180) McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice / T. McEnery, A. Hardie. – Cambridge: Cambridge University Press, 2012. – XV, 294 p.

181) Munday J. A computer-assisted approach to the analysis of translation shifts / J. Munday // Meta. – 1998. – Volume 43, № 4. – P. 542-556.

182) Nisioi S. Identifying Source-Language Dialects in Translation / S. Nisioi, A. S. Uban, L. P. Dinu // Mathematics. – 2022. – Volume 10. – P. 1-14.

183) Newmark P. About translation / P. Newmark. – United Kingdom: Multilingual Matters, 1991.

184) Nida E. A., Taber C. R. The Theory and Practice of Translation / E. A. Nida, C. R. Taber. – Leiden: E. J. Brill, 1969. – 220 p.

185) Oakes M. Describing a translational corpus / M. Oakes // Quantitative Methods in Corpus-Based Translation Studies: A practical guide to descriptive translation research / edited by M. P. Oakes, M. Ji. – Amsterdam; Philadelphia: John Benjamins, 2012. – P. 115-148.

186) Olohan M. Corpus Linguistics and Translation Studies: Interaction and Reaction / M. Olohan // Linguistica Antverpiensia. – 2002. – P. 419-429.

187) Olohan M. Spelling out the optionals in translation: A corpus study / M. Olohan // UCREL Technical Papers. – 2001. – Volume 13. – P. 423-432.

188) Ozyumenko V. Modality in English and Russian Legislative Discourse: A Comparative Study / V. Ozyumenko // Respectus Philologicus. – 2012. – Volume 22, № 27. – P. 37-44.

189) Øverås L. In search of the third code: An investigation of norms in literary translation: [Unpublished cand. of philol. thesis] / L. Øverås. – Oslo: Department of British and American Studies, University of Oslo, 1996. – 182 p.

190) Pápai V. Explicitation: A universal of translated text? / V. Pápai // Translation universals: Do they exist? – Amsterdam; Philadelphia: John Benjamins Publishing Company, 2004. – P. 143-164.

191) Popescu M. Studying Translationese at the Character Level / M. Popescu // Proceedings of the International Conference Recent Advances in Natural Language Processing 2011, Hissar, Bulgaria. – Association for Computational Linguistics, 2011. – P. 634-639.

192) Quirk R. A comprehensive grammar of the English language / R. Quirk, S. Greenbaum, G. Leech, J. Svartvik. – London; New York: Longman, 1985. – 1779 p.

193) Translation universals: do they exist? A corpus-based NLP study of convergence and simplification / G. Corpas Pastor, R. Mitkov, N. Afzal, V. Pekar // Proceedings of the 8th Conference of the Association for Machine Translation in the Americas: Research Papers. – 2008. – P. 75-81.

194) Puurtinen T. Explication of clausal relations: A corpus-based analysis of clause connectives in translated and non-translated Finnish children's literature / T. Puurtinen // Translation Universals: Do they exist? / edited by A. Mauranen, P. Kujamäki. – Amsterdam; Philadelphia: John Benjamins Publishing Company, 2004. – P. 165-176.

195) Quantitative analysis of translation revision: contrastive corpus research on native English and Chinese translationese / P. Rayson, X. Xu, J. Xiao, A. Wong, Q. Yuan // Proceedings of XVIII FIT World Congress, Shanghai, China, August 4-7, 2008. – Shanghai, 2008.

196) Riley P. Translationese as a language in "multilingual" NMT / P. Riley, I. Caswell, M. Freitag, D. Grangier // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. – 2020. – P. 7737-7746.

197) Shuttleworth M., Cowie M, Dictionary of Translation Studies / M. Shuttleworth, M. Cowie. – London: Routledge, 2014. – 233 p.

198) Rodríguez-Castro M. Translationese and punctuation: An empirical study of translated and non-translated international newspaper articles (English and Spanish) / M. Rodríguez-Castro // *Translation and Interpreting Studies*. – 2011.

199) Roxbury T. An fMRI study of concreteness effects in spoken word recognition / T. Roxbury, K. McMahon, D. A. Copland // *Behavioral and Brain Functions*. – 2014. – Vol. 10. – Art. 34. – 10-34.

200) -Meyer F. Hedges and textual communicative function in medical English written discourse / F. Salager-Meyer // *English for specific purposes*– 1994. – Volume 13, № 2. – P. 149-170.

201) Scott M. *The WordSmith Tools (v. 4.0)* / M. Scott. – Oxford: Oxford University Press, 2004.

202) Snell-Hornby M. *The Turns of Translation Studies. New Paradigms or Shifting Viewpoints?* / M. Snell-Hornby. – Amsterdam; Philadelphia: John Benjamins, 2006. – 205 p.

203) Toury G. *Interlanguage and its Manifestations in Translation* / G. Toury // *Meta*. – 1979. – Volume 24, № 2. – P. 223-231.

204) Toury G. *In Search of a Theory of Translation* / G. Toury. – Tel Aviv: The Porter Institute for Poetics and Semiotics, Tel Aviv University, 1980. – 159 p.

205) Toury G. *Descriptive translation studies and beyond: Revised edition* / G. Toury. – Volume 100. – Amsterdam; Philadelphia: John Benjamins Publishing, 2012. – 350 p.

206) Toury G. *A Rationale for Descriptive Translation Studies* / G. Toury // *The Manipulation of Literature: Studies in Literary Translation* / edited by T. Hermans. – London; Sydney: Croom Helm, 1985. – P. 16-41.

207) Toury G. *Probabilistic Explanations in Translation Studies: Welcome as They Are, Would They Qualify as Universals?* / G. Toury // *Translation Universals: Do they exist?* / edited by A. Mauranen, P. Kujamäki. – Amsterdam; Philadelphia: John Benjamins, 2004. – P. 15-32.

208) Tirkkonen-Condit S. *The Monitor Model Revisited: Evidence from Process Research* / S. Tirkkonen-Condit // *Meta*. – 2005. – Volume 50, № 2. – P. 405-414.

209) Selinker L. Interlanguage / L. Selinker // IRAL; International Review of Applied Linguistics in Language Teaching. – 1972. – Volume 10, № 3. – P. 209-231.

210) Ure J. Lexical density and register differentiation / J. Ure // Applications of Linguistics / edited by G. Perren, J. L. M. Trim. – London: Cambridge University Press, 1971. – P. 443-452.

211) Vanderauwera R. Dutch Novels Translated Into English: the Transformation of a Minority Literature / R. Vanderauwera. – Amsterdam: Rodopi, 1985. – 170 p.

212) Vanmassenhove E. Lost in Translation: Loss and Decay of Linguistic Richness in Machine Translation / E. Vanmassenhove, D. Shterionov, A. Way // Proceedings of the 17th Machine Translation Summit. – Volume 1. – 2019. – P. 222-232.

213) Vargha A., Delaney H. D. A Critique and Improvement of the CL Common Language Effect Size Statistics of McGraw and Wong / A. Vargha, H. D. Delaney // Journal of Educational and Behavioral Statistics. – 2000. – Volume 25, № 2. – P. 101-132.

214) Vinay J.-P., Darbelnet J. Comparative Stylistics of French and English: A Methodology for Translation / J.-P. Vinay, J. Darbelnet; translated and edited by J. C. Sager, M.-J. Hamel. – Amsterdam; Philadelphia: John Benjamins Publishing Company, 1958/1995. – 359 p.

215) Volansky V. On the features of translationese / V. Volansky, N. Ordan, S. Wintner // Digital Scholarship in the Humanities. – 2015. – Volume 30, № 1. – P. 98-118.

216) Wang Z. A corpus-based study of syntactic complexity of translated and non-translated chairman's statements / Z. Wang, K. Liu, R. Moratto // Translation & Interpreting: The International Journal of Translation and Interpreting Research. – 2023. – Volume 15, № 1. – P. 135-151.

217) Weinreich U. Languages in Contact / U. Weinreich. – The Hague: Mouton, 1953. – 172 p.

218) Welch B. L. The Generalization of 'Student's' Problem when Several Different Population Variances are Involved / B. L. Welch // Biometrika. – 1947. – Volume 34, № 1/2. – P. 28-35.

219) Yakhontova T. Conventions of English research discourse and the writing of non-Anglophone authors / T. Yakhontova // Journal of Korean Medical Science. – 2020. – Volume 35, № 40. – P. 1-4.

220) Zanettin F. Corpus methods for descriptive translation studies / F. Zanettin // Procedia - Social and Behavioral Sciences. – 2013. – Volume 95. – P. 20-32.

221) Zanina E. Strategic hedging: a comparative study of methods, results and discussion (and conclusion) sections of research articles in English and Russian / E. Zanina // Journal of Language and Education. – 2016. – Volume 2, № 2. – P. 52-60.

222) Zhang M. The effect of translationese in machine translation test sets / M. Zhang, A. Toral // Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers), Florence, Italy. – Association for Computational Linguistics, 2019. – P. 73-81.

223) Zhou F. The Comparison of Translationese in Machine Translation and Human Translation in terms of Translation Relations / F. Zhou // arXiv preprint arXiv:2404.08661. – 2024.

## **СПИСОК ИСТОЧНИКОВ**

### **Научные журналы, использованные при отборе статей по социологии (непереводные тексты и тексты ручных переводов)**

1) Журнал социологии и социальной антропологии : [сайт] / учредители ФНИСЦ РАН; Фонд «Интерсоцис» ; издатель Фонд «Интерсоцис». – Санкт-Петербург, 1998– . – Периодичность: ежеквартально. – URL: <https://www.jourssa.ru/> (дата обращения: 15.02.2025). – Текст : электронный.

2) Социологическое обозрение : [сайт] / учредитель Национальный исследовательский университет «Высшая школа экономики»; А. Ф. Филиппов. – Москва, 2001– . – Периодичность: ежеквартально. – URL: <https://sociologica.hse.ru/> (дата обращения: 15.02.2025). – Текст : электронный.

3) Экономическая социология : [сайт] / учредитель Национальный исследовательский университет «Высшая школа экономики». – Москва, 2000– . –

Периодичность: 5 раз в год. – URL: <https://ecsoc.hse.ru/> (дата обращения: 15.02.2025). – Текст : электронный.

**Научные журналы, использованные при отборе статей по медицине (непереводные тексты и тексты ручных переводов)**

4) Доказательная гастроэнтерология : [сайт] / учредитель и издатель издательство «Медиа Сфера». – Москва, 2011– . – Периодичность: ежеквартально. – URL: <https://www.mediasphera.ru/journal/dokazatel'naya-gastroenterologiya> (дата обращения: 25.04.2025). – Текст : электронный.

5) Кардиология : [сайт] / учредитель и издатель РОО «Кардиомаг». – Москва, 1961– . – Периодичность: 12 раз в год. – URL: <https://cardio.elpub.ru/> (дата обращения: 04.05.2025). – Текст : электронный.

6) Качественная клиническая практика : [сайт] / учредитель и издатель ООО «Издательство ОКИ». – Москва, 2001– . – Периодичность: ежеквартально. – URL: <https://www.clinvest.ru/> (дата обращения: 04.05.2025). – Текст : электронный.

7) Клиническая и экспериментальная тиреоидология : [сайт] / учредители: ФГБУ «НМИЦ эндокринологии» Минздрава России; ОО «Российская ассоциация эндокринологов»; издатель ФГБУ «НМИЦ эндокринологии» Минздрава России. – Москва, 2005– . – Периодичность: ежеквартально. – URL: <https://www.cet-endojournals.ru/> (дата обращения: 13.05.2025). – Текст : электронный.

8) Нервно-мышечные болезни : [сайт] / учредитель и издатель ООО «ИД «АБВ-пресс». – Москва, 2011– . – Периодичность: ежеквартально. – URL: <https://nmb.abvpress.ru/> (дата обращения: 04.05.2025). – Текст : электронный.

9) Нефрология и диализ : [сайт] / учредители и издатели: Общероссийская общественная организация Нефрологов «Российское диализное общество»; Межрегиональная общественная организация «Творческое объединение детских нефрологов». – Москва, 1999– . – Периодичность: ежеквартально. – URL: <https://journal.nephro.ru/> (дата обращения: 25.04.2025). – Текст : электронный.

10) Остеопороз и остеопатии : [сайт] / учредители: ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России; ОО «Российская ассоциация эндокринологов»; ОО «Российская ассоциация по остеопорозу»;

издатель ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. – Москва, 1998– . – Периодичность: ежеквартально. – URL: <https://www.osteo-endojournals.ru/> (дата обращения: 04.05.2025). – Текст : электронный.

11) Патология кровообращения и кардиохирургия : [сайт] / учредитель и издатель ФГБУ «НМИЦ им. акад. Е. Н. Мешалкина» Минздрава России. – Новосибирск, 1997– . – Периодичность: ежеквартально. – URL: <https://journalmeshalkin.ru/> (дата обращения: 28.04.2025). – Текст : электронный.

12) Проблемы репродукции : [сайт] / учредитель и издатель Издательство «Медиа Сфера». – Москва, 1995– . – Периодичность: 6 раз в год. – URL: <https://www.mediasphera.ru/journal/problemy-reproduktcii> (дата обращения: 13.05.2025). – Текст : электронный.

13) Проблемы эндокринологии : [сайт] / учредители: ФГБУ «НМИЦ эндокринологии» Минздрава России; ООО «Российская ассоциация эндокринологов»; издатель ФГБУ «НМИЦ эндокринологии им. академика И.И. Дедова» Минздрава России. – Москва, 1966– . – Периодичность: 6 раз в год. – URL: <https://www.probl-endojournals.ru/> (дата обращения: 13.05.2025). – Текст : электронный.

14) Пульмонология : [сайт] / учредитель МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РСФСР; издатель ООО «Научно-практический журнал “Пульмонология“». – Москва, 1991– . – Периодичность: 6 номеров в год. – URL: <https://journal.pulmonology.ru/> (дата обращения: 12.05.2025). – Текст : электронный.

15) Рациональная фармакотерапия в кардиологии : [сайт] / учредитель Национальный медицинский исследовательский центр терапии и профилактической медицины; издатель ООО «СИЛИЦЕЯ-ПОЛИГРАФ». – 2005– . – Периодичность: 6 раз в год. – URL: <https://www.rpcardio.online/> (дата обращения: 25.04.2025). – Текст : электронный.

16) Российский кардиологический журнал : [сайт] / учредители: Российское кардиологическое общество; ООО «Фирма Силицея»; издатель ООО

«Фирма Силицея». – Москва, 1996– . – Периодичность: 12 раз в год. – URL: <https://russjcardiol.eipub.ru/> (дата обращения: 28.04.2025). – Текст : электронный.

17) Русский журнал детской неврологии : [сайт] / учредители: Петрухин А. С.; Мухин К. Ю.; издатель ООО «ИД «АБВ-пресс». – Москва, 2006– . – Периодичность: ежеквартально. – URL: <https://rjdn.abvpress.ru/> (дата обращения: 12.05.2025). – Текст : электронный.

18) The EYE ГЛАЗ : [сайт] / учредитель и издатель АНО «Академия медицинской оптики и оптометрии». – Москва, 1998– . – Периодичность: ежеквартально. – URL: <https://www.theeyeglaz.com/> (дата обращения: 25.04.2025). – Текст : электронный.

**Научные журналы, использованные при отборе статей по медицине (тексты для машинных ручных переводов)**

19) Biology of Reproduction : [website] / founder Society for the Study of Reproduction ; publisher Oxford University Press. – New York, 1969– . – Frequency: monthly. – URL: <https://academic.oup.com/biolreprod> (accessed: 22.05.2025). – Text : electronic.

20) European Journal of Pharmacology : [website] / founder and publisher Elsevier. – Amsterdam, 1967– . – Frequency: fortnightly. – URL: <https://www.sciencedirect.com/journal/european-journal-of-pharmacology> (accessed: 22.05.2025). – Text : electronic.

21) The Lancet : [website] / publisher Elsevier. – London, 1823– . – Frequency: weekly. – URL: <https://www.sciencedirect.com/journal/the-lancet> (accessed: 22.05.2025). – Text : electronic.

22) The New England Journal of Medicine : [website] / founder and publisher Massachusetts Medical Society. – Boston, 1928– . – Frequency: weekly. – URL: <https://www.nejm.org/> (accessed: 22.05.2025). – Text : electronic.

## ПРИЛОЖЕНИЕ А

### Классификация дискурсивных маркеров локального и глобального уровня связности текста

<b>Локальный уровень связности</b>	
<b>Функция</b>	<b>Примеры</b>
Введение новой или дополнительной информации	кроме того; также; помимо этого; более того; в том числе; вдобавок к; наряду с; вместе с тем; к тому же; с другой стороны; в дополнение к; притом; причем; при этом; в то же время
Пояснение, уточнение или перефразирование упомянутого ранее	другими словами; иными словами; иначе говоря; конкретнее говоря; то есть; а именно; точнее говоря; в том смысле, что; под ... понимается; если перефразировать сказанное; поясним сказанное; речь идёт о том, что; одним словом
Противопоставление или отход от основной линии изложения	однако; хотя; тем не менее; в то же время; между тем; при всем этом; вопреки этому; тогда как; ...идет вразрез с; ...вступает в противоречие с; при всей очевидности; на поверку оказывается; в действительности же; на деле же; на деле все обстоит иначе; в противоположность этому; напротив; наоборот
Введение примеров	например; такие...; как; к примеру; для иллюстрации; приведем пример; для примера рассмотрим; проиллюстрируем; примером служит; рассмотрим следующий пример; взять хотя бы
Выражение причинно-следственных отношений	поэтому; потому, что; по этой причине; причиной служит; в силу того, что; вследствие этого, из этого следует; причина кроется в; на основании того, что; ввиду того, что
Выражение целевых отношений	для того, чтобы; затем, чтобы; с расчетом на то, что; ради того, чтобы; в интересах; с целью; целью... является
<b>Глобальный уровень связности</b>	
<b>Функция</b>	<b>Примеры</b>
Введение читателя в тему статьи / исследования	в данной работе / статье речь пойдет о; в данной статье мы рассмотрим; в этой статье мы обратимся; настоящая работа посвящена; исследование посвящено; в данной статье / работе предпринята попытка
Обеспечение порядка следования информации	во-первых; во-вторых; прежде всего; в первую очередь; вначале / сперва / для начала рассмотрим; следует начать с; предварительно отметим; далее; вслед за этим; теперь обратимся / перейдем к; наконец; в завершение; в заключение
Описание расположения материала в тексте	как говорилось ранее / выше; как было показано ранее / выше; как уже отмечалось; о чем будет сказано дальше / ниже; будет представлено в...; как уже было отмечено; как было показано выше; описанные / перечисленные выше; будет продемонстрировано далее; как было отмечено ранее
Выводы и заключение	итак; таким образом; отсюда следует вывод; следовательно; можно сделать вывод; в результате; можно заключить, что

## ПРИЛОЖЕНИЕ Б

### Список тегов, использованных в анализе частеречных последовательностей (включая специальные теги)

Тег	Пояснение	Примеры
<b>NOUN</b>	имя существительное	измерение, анализ
<b>ADJF</b>	имя прилагательное (полное)	хороший, известный
<b>ADJS</b>	имя прилагательное (краткое)	хорош, известен
<b>VERB</b>	глагол (личная форма)	говорили, считают, рассмотрим
<b>INFN</b>	инфинитив	сказать, упомянуть
<b>GRND</b>	деепричастие	прочитав, говоря
<b>PRTF</b>	причастие (полное)	прочитавший, проанализированная
<b>PRTS</b>	причастие (краткое)	прочитана, рассмотрены
<b>ADVB</b>	наречие	длительно, значительно
<b>NUM</b>	числительное	сорок два, 42, 3.14, третий
<b>PREP</b>	предлог	в, на, под
<b>PREP_PRT</b>	часть составного предлога	«несмотря» в «несмотря на», «течение» в «в течение»
<b>CCONJ</b>	сочинительный союз	и, а, либо, или
<b>CCONJ_PRT</b>	вторая часть составного сочинит. союза	«есть» в «то есть», «именно» в «а именно»
<b>SCONJ</b>	подчинительный союз	что, как, чтобы, если
<b>SCONJ_PRT</b>	вторая часть составного подчинит. союза	«что» в «потому что»
<b>PRN_QW</b>	вопросит. местоимение в начале предложения	кто, какой
<b>ADV_QW</b>	вопросит. наречие в начале предложения	где, когда, почему
<b>PRN_CONJ_W</b>	местоименное союзное слово	который, чей (после запятой)
<b>ADV_CONJ_W</b>	наречное союзное слово	где, почему (после запятой)
<b>PRCL</b>	частица	же, лишь
<b>SBJ_PRCL</b>	сослагательная частица	бы, б
<b>NEG_PART</b>	отрицательная частица	не, ни, нет
<b>PARENTH</b>	вводное слово	конечно, например, действительно
<b>MDL_WRD</b>	модальное слово	нужно, можно, необходимо
<b>MDL_VRB</b>	модальный глагол/ его эквивалент	мочь
<b>COPULA</b>	глагол-связка (в составных именных сказуемых)	являться, быть, стать
<b>AUX</b>	вспомогательный глагол	«был» в «был построен», «будет» в «будет рассматриваться»
<b>PER_PRN</b>	личное местоимение функции подлежащего	я, ты, он, она, они, оно
<b>OBJ_PRN</b>	личное местоимение в функции дополнения	его, её, ему, (о) нём
<b>POSSV_PRN</b>	притяжательное местоимение	мой, твой, свой
<b>DMSTR_PRN</b>	указательное местоимение	этот, тот, данный
<b>RFLX_PRN</b>	возвратное местоимение	себя

<b>REC_PRN</b>	первая часть взаимного местоимения	«друг» в «друг друга», «друг с другом»
<b>REC_PRN_PART</b>	вторая часть взаимного местоимения	«друга» в «друг друга», «другом» в «друг с другом»
<b>DF_ADJ_N_PRN</b>	определятельное местоименное сущ./прил.	иной сказал бы, что...(мест.-сущ.), иной подход (мест.-прил.)
<b>DF_ADV_PRN</b>	определит. местоим.-наречие	везде, всегда, тогда
<b>INDF_NOUN_PRN</b>	неопр. мест.-сущ.	некто, кое-кто, кто-то
<b>INDF_ADJ_PRN</b>	неопр. мест.-прил.	некоторый, кое-какой, какой-то
<b>INDF_ADV_PRN</b>	неопр. мест.-нареч.	как-то, где-то, когда-то
<b>INDF_NUM_PRN</b>	неопр. мест.-числит.	несколько, сколько-то
<b>NEG_N_PRN</b>	отрицат. мест.-сущ.	никто, ничто
<b>NEG_ADJ_PRN</b>	отрицат. мест.-прил.	никакой, ничей
<b>NEG_ADV_PRN</b>	отрицат. мест.-нареч.	никогда, нигде, никак
<b>NEG_NUM_PRN</b>	отрицат. мест.-числит.	нисколько
<b>COMP</b>	синтетич. форма сравнительной степени	лучше, получше, выше
<b>SPRL</b>	синтетич. форма превосходной степени	наилучший, высочайший
<b>COMP_MOD</b>	модификатор аналитической формы сравнительной степени	более, менее
<b>SPRL_MOD</b>	модификатор аналитической формы превосходной степени	самый, наиболее, наименее
<b>S_START</b>	начало предложения	Пример выделен подчеркиванием
<b>S_END</b>	конец предложения	.?! ...
<b>COMMA</b>	запятая	,
<b>CLN</b>	двоеточие	:
<b>SMCLN</b>	точка с запятой	;
<b>DASH</b>	тире	—
<b>SLASH</b>	слэш	/
<b>RN_OPN_BR</b>	круглая скобка открыв.	(
<b>RN_CLS_BR</b>	круглая скобка закрыв.	)
<b>SQ_OPN_BR</b>	квадратная скобка открыв.	[
<b>SQ_CLS_BR</b>	квадратная скобка закрыв.	]
<b>FG_OPN_BR</b>	фигурная скобка открыв.	{
<b>FG_CLS_BR</b>	фигурная скобка закрыв.	}
<b>QUOTE</b>	кавычки	« » „ “ ”
<b>PUNCT</b>	прочие пунктуационные знаки	
<b>INTJ</b>	междометие	ой
<b>FRN_W</b>	иностранное слово	NLP, case
<b>N/A</b>	не определено	

## ПРИЛОЖЕНИЕ В

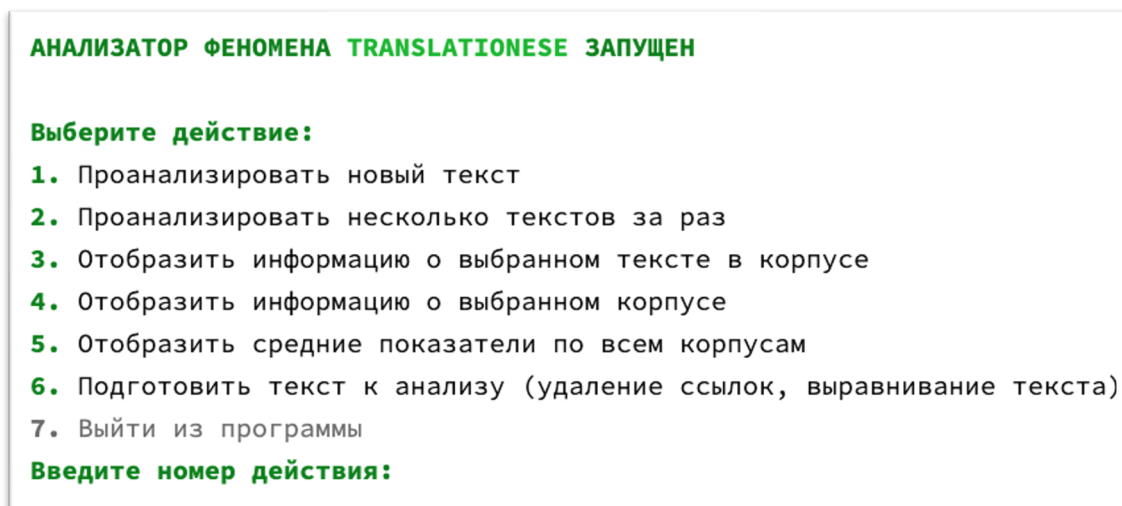
### Классификация дискурсивных маркеров авторского начала и взаимодействия автора с читателем

<b>Маркеры авторского начала</b>	
<b>Функция</b>	<b>Примеры</b>
Выражение мнения автора	с нашей точки зрения; на наш / мой взгляд; по нашему мнению; смеем утверждать; мы склонны полагать; мы считаем; я считаю
Выражение оценки предмета обсуждения / отношения к нему	к сожалению; к счастью; любопытно; справедливо; неоднозначно; удивительно; убедительно; примечательно; вызывает сомнения; приходится с сожалением констатировать
Выражение высокой степени уверенности	безусловно; очевидно; несомненно; без сомнения; определено; непременно; разумеется; бесспорно; нет никаких сомнений; можно с уверенностью утверждать; не подлежит сомнению; не вызывает возражений; достоверно; должен / должна / должно / должны + глагол ( <i>рассматриваться</i> и т. д.)
Выражение средней степени уверенности	возможно; вероятно; похоже; видимо; по-видимому; скорее всего; может быть; наверное; не исключено, как нам кажется; вполне очевидно; существует вероятность; можно + глагол ( <i>предположить, допустить</i> и т. д.)
Выражение низкой степени уверенности	маловероятно; (весьма / крайне) сомнительно; спорно; вряд ли; едва ли; трудно сказать; сложно утверждать; нельзя с уверенностью сказать; вызывает серьезные сомнения
<b>Маркеры взаимодействия автора с читателем</b>	
<b>Функция</b>	<b>Примеры</b>
Призыв читателя к действию	см. (смотрите); ср. (сравните); обратите внимание; сопоставьте; соотнесите
Призыв читателя к совместному с автором действию	рассмотрим; обозначим; обратимся к; допустим; проследим; поговорим о; проанализируем
Акцентирование внимания читателя на важной или дополнительной информации	следует / важно / необходимо / уместно отметить; примечательно, что...; подчеркнем; важным представляется подчеркнуть; следует / стоит обратить (особое) внимание на; мы хотели бы подчеркнуть / отметить; заметим, что; нельзя обойти стороной вопрос о
Апелляция к фоновым знаниям читателя	не секрет, что; общеизвестно; как известно; принято считать; принято полагать; ...является общепризнанным; ...является общеизвестным фактом; не вызывает споров; признается научным сообществом; ...является общепринятым в научной среде

## ПРИЛОЖЕНИЕ Г

### Описание интерфейса и функциональности разработанного программного комплекса для анализа индикаторов феномена переводческого языка

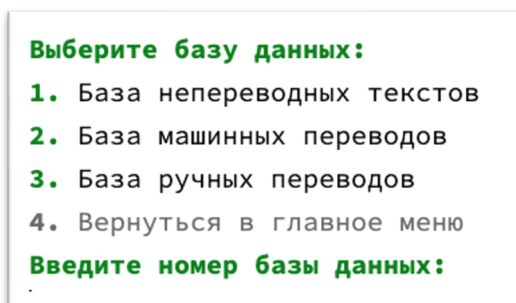
Ниже приведено краткое описание интерфейса и функциональности программного комплекса для анализа индикаторов феномена переводческого языка<sup>14</sup>. Программный комплекс реализован на базе языка программирования Python. На рисунке 1 представлен интерфейс главного меню программы.



*Рисунок 1.* Главное меню программы для анализа индикаторов феномена переводческого языка

#### Функционал программы:

При выборе опции «Проанализировать новый текст» пользователю предоставляется возможность выбрать корпус, к которому относится анализируемый текст (см. Рис. 2).



*Рисунок 2.* Варианты выбора корпуса для сохранения анализируемого текста

<sup>14</sup> Анализатор феномена переводческого языка. URL: <https://pypi.org/project/translationese-analyzer/>

Каждый корпус реализован в формате SQLite-базы данных, в которой каждая таблица связана с одной из переводческих характеристик (*упрощением, нормализацией, экспликацией, интерференцией, остальными индикаторами, не вошедшими в предыдущие характеристики*) и содержит соответствующие ей индикаторы. Строки таблиц содержат данные по каждому отдельному анализируемому программой тексту, столбцы – данные по анализируемым программой индикаторам.

Ввод текста осуществляется либо вручную через клавиатуру, либо через буфер обмена (*Ctrl+V*), либо загружается из файла (в данном случае пользователь указывает название директории с файлом). По завершению ввода пользователю предлагается заполнить паспорт текста, включающий информацию о названии статьи, предметной области, ключевых словах и авторах, а также другие метаданные (см. Рис. 2), которые могут быть полезны при анализе.

Поле	Значение
Название статьи	Особенности системно-теоретической проблематики власти...
Предметная область	Социология
Ключевые слова	власть, социальный порядок, ...
Год публикации	2008
Место публикации (журнал)	Социологическое обозрение
Автор(ы)	Владимир Попов
Пол автора(ов)	м
Год рождения автора(ов)	1982

Рисунок 3. Пример паспорта текста

После того, как паспорт текста создан, начинается анализ индикаторов характеристик феномена переводческого языка. По выбору пользователя данные с результатами анализа индикаторов каждой переводческой характеристики выводятся в консоль. При желании пользователь может проанализировать сразу много текстов за раз, выбрав в главном меню опцию «Проанализировать несколько текстов за раз» (см. Рис. 1).

Программа предоставляет возможности анализа следующих индикаторов (ниже приведено краткое описание анализируемых индикаторов):

### **Индикаторы характеристики упрощение:**

Лексическая плотность (*lexical density*) – отношение знаменательных слов (а именно существительных, прилагательных, глаголов и наречий) к общему количеству словарных токенов в тексте;

Лексическая вариативность (*lexical variety*) – отношение уникальных слов (типов) ко всем словарным токенам в тексте. Анализируются четыре версии коэффициента: *TTR*, *Log TTR*, *Modified TTR* и *Standardized TTR (STTR)*;

Средний ранг слов (*mean word rank*) – среднее значение частотного ранга используемой в тексте лексики, рассчитанное на основе частотного словаря русского языка [Ляшевская, Шаров, 2009: <http://dict.ruslang.ru/freq.php>];

Средняя доля наиболее частотных знаменательных слов подкорпуса (*list head ratio*);

Средняя длина слов (*mean word length*) в символах и в слогах;

Средняя длина предложений (*mean sentence length*) в символах и в токенах;

Средняя длина клауз в словарных токенах (*mean clause length*);

Среднее число клауз на предложение (*mean clauses per sentence*);

Средняя доля простых предложений (*simple sentences ratio*);

Индекс удобочитаемости Флеша (*Flesch reading ease score*);

Нормализованные частоты знаков препинания (*punctuation marks frequency*) относительно общего числа токенов;

### **Индикаторы характеристики нормализация:**

Повторяемость знаменательной лексики (*lexical repetition*) – отношение числа знаменательных слов (а именно глаголов, имен существительных, прилагательных и наречий), встречающихся в тексте более 1 раза (т. е. знаменательных слов с частотой  $\geq 2$ ) к общему количеству токенов в тексте;

Среднее значение поточечной взаимной информации (*average PMI*) — среднее значение *PMI / Modified PMI* по всем словарным биграммам подкорпуса. Поточечная взаимная информация биграмм слов (*pointwise mutual information, PMI*) – показатель силы ассоциативной связи между двумя соседствующими словами, то

есть то, насколько чаще два слова встречаются совместно, чем можно было бы ожидать, если бы их появления в тексте были бы независимы друг от друга;

Доля биграмм с  $PMI > 0$  (*positive PMI bigrams ratio*) — доля словарных биграмм, значение PMI / Modified PMI которых превышает 0.

### **Индикаторы характеристики экспликация**

#### ***Лексическая когезия:***

Среднее косинусное сходство пар знаменательных слов (*mean pairwise cosine similarity*);

Среднее косинусное сходство пар соседних предложений (*mean adjacent sentence cosine similarity*) – среднее косинусное сходство между векторными представлениями последовательных предложений текста;

Эксплицитное называние (*explicit naming*) – отношение числа личных и притяжательных местоимений ко всем именованным сущностям в тексте;

Единичное именование (*single naming*) – отношение числа именованных сущностей из одного токена ко всем именованным сущностям;

Средняя длина именованных сущностей в токенах (*mean named entity length*).

#### ***Грамматическая (референциальная) когезия:***

Нормализованные частоты местоимений (*pronoun frequency*) личных, притяжательных, указательных, относительных (относительно числа словарных токенов);

Эксплицитное называние (*explicit naming*) – отношение числа личных и притяжательных местоимений к общему числу именованных сущностей в тексте;

Единичное именование (*single naming*) – отношение числа именованных сущностей из одного токена к общему числу сущностей;

Средняя длина именованных сущностей (*mean named entities length*) в токенах.

#### ***Грамматическая (конъюнктивная) когезия:***

Нормализованные частоты дискурсивных маркеров локального уровня связности текста (*local cohesion discourse markers frequency*);

Нормализованные частоты дискурсивных маркеров глобального уровня связности текста (*global cohesion discourse markers frequency*).

### **Индикаторы характеристики интерференция**

Нормализованные частоты частеречных n-грамм (*POS n-grams frequency*), а именно униграмм, биграмм и триграмм. Нормализация осуществляется путём деления абсолютной частоты конкретной n-граммы на совокупное число всех n-грамм в тексте. Полный перечень используемых в анализе тегов представлен в ПРИЛОЖЕНИИ Б.

### **Дополнительные индикаторы, отражающие особенности научного стиля русского языка**

#### ***Дискурсивные маркеры:***

Нормализованные частоты маркеров авторского начала в тексте и маркеров взаимодействия автора с читателем (*authorial stance and reader engagement markers frequency*);

#### ***Морфологические индикаторы:***

Доля существительных в тексте (*nouns ratio*);

Доля существительных среднего рода (*neuter nouns ratio*);

Доля существительных в форме единственного (*singular nouns ratio*);

Доля абстрактных существительных (*abstract nouns ratio*);

Доля прилагательных в тексте (*adjectives ratio*);

Доля кратких прилагательных (*short-form adjectives ratio*);

Доля прилагательных в формах сравнительной/ превосходной степени (*comparative / superlative adjectives ratio*);

Доля глаголов в тексте (*verbs ratio*);

Доля глаголов в форме страдательного залога (*passive voice ratio*);

Доля причастий относительно всех глаголов (*participles ratio*);

Доля кратких причастий (*short-form participles ratio*);

Доля причастий в препозиции перед существительным (*pre-nominal participles ratio*);

Доля причастий, входящих в состав причастных оборотов (*participial phrases ratio*);

Доля деепричастий относительно всех глаголов (*gerunds ratio*);

Доля глаголов несовершенного вида среди всех глаголов и среди личных форм глаголов (*imperfective aspect verbs ratio*);

Доля глаголов в формах прошедшего, настоящего и будущего времени (*verb tense distribution*).

Помимо анализа индикаторов программа осуществляет морфологическую разметку (с помощью библиотеки *rumorphy*<sup>15</sup>) и синтаксическую разметку (с помощью библиотеки *Natasha*<sup>16</sup>) текста, последняя выполняется в соответствии со стандартом *Universal Dependencies*<sup>17</sup>.

По завершению этапа анализа текста все полученные результаты сохраняются в выбранную базу данных (т. е. выбранный подкорпус). При желании пользователь может обратиться к результатам анализа отдельного текста, выбрав в главном меню опцию «Отобразить информацию о выбранном тексте в корпусе» (см. Рис. 1 выше). При выборе данной опции пользователю становится доступна информация, представленная на рисунке 4.

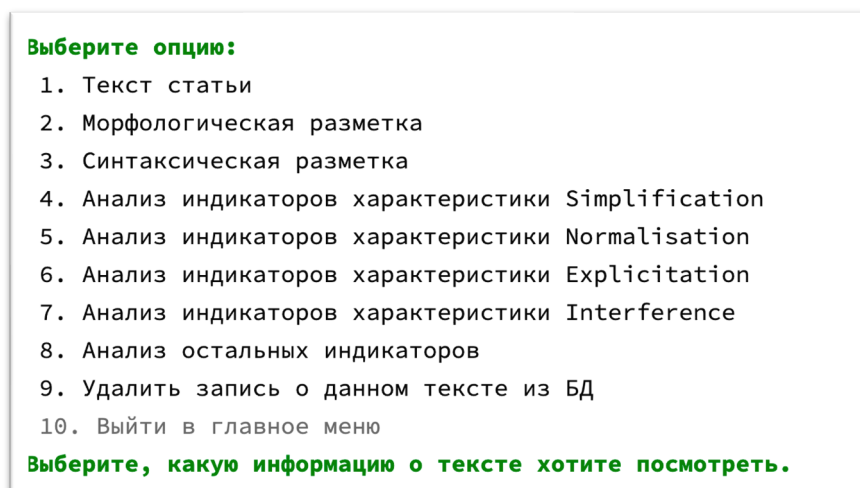


Рисунок 4. Информация, доступная для просмотра об отдельном тексте

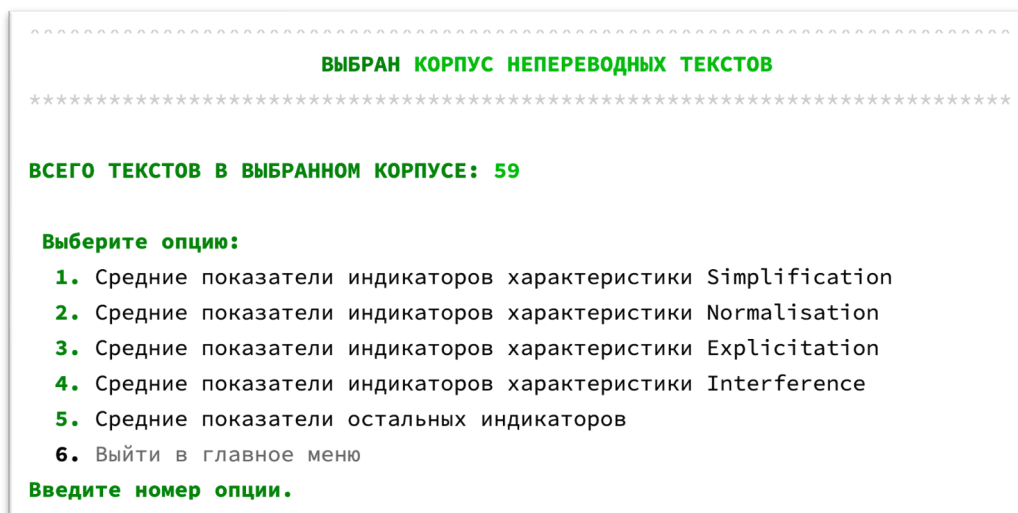
<sup>15</sup> Морфологический анализатор для русского языка *rumorphy2*: <https://pymorphy2.readthedocs.io/en/stable/>

<sup>16</sup> Библиотека для обработки естественного русского языка *Natasha*: <https://pypi.org/project/natasha/>

<sup>17</sup> *Universal Dependencies for Russian*: <https://universaldependencies.org/ru/index.html>

Ключевыми являются функция просмотра информации о всех текстах в корпусе (опция «Отобразить информацию о выбранном корпусе») и функция просмотра и сопоставления средних показателей между отдельными корпусами (опция «Отобразить средние показатели по всем корпусам») (Рис. 1).

При выборе опции «Отобразить информацию о выбранном корпусе» пользователю доступен просмотр данных по средним показателям каждой из характеристик для всего выбранного корпуса (см. Рис. 5).



*Рисунок 5.* Информация, доступная для просмотра об отдельном тексте

При выборе опции «Отобразить средние показатели по всем корпусам» пользователю доступен просмотр данных по средним показателям каждой из характеристик в виде сводной таблицы для всех созданных корпусов.

Дополнительно программа предоставляет возможность подготовить текст к дальнейшему анализу (опция «Подготовить текст к анализу (удаление ссылок, выравнивание текста)» (см. Рис. 1)), автоматически удалив из него внутритекстовые ссылки на литературу разных форматов. Также при желании пользователь может выравнивать текст по длине, выбрав желаемое количество предложений. Удаление предложений осуществляется двумя способами: (1) по выбору пользователя и (2) случайным образом.

## ПРИЛОЖЕНИЕ Д

## Результаты анализа индикаторов упрощения по корпусам

СОЦ-ПП						
Индикатор	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	Welch t (df) r	p	95% CI $\Delta$
Lexical density	64.79 (SD=2.45, CV=3.8%)	62.54 (SD=2.93, CV=4.7%)	+2.25	4.18 (93.44) r $\approx$ 0.4	<0.001 (***)	[1.18; 3.32]
TTR	28.18 (SD=5.03, CV=17.9%)	27.46 (SD=5.57, CV=20.3%)	+0.72	0.68 (96.43) r $\approx$ 0.07	0.497 (n.s.)	[-1.38; 2.82]
Log TTR	85.04 (SD=1.82, CV=2.1%)	84.94 (SD=1.76, CV=2.1%)	+0.10	0.30 (99.46) r $\approx$ 0.03	0.768 (n.s.)	[-0.6; 0.8]
Modified TTR	1954.77 (SD=176.8, CV=9%)	1975.27 (SD=170.8, CV=8.7%)	-20.5	-0.59 (99.48) r $\approx$ 0.06	0.554 (n.s.)	[-88.9; 47.9]
STTR	47.64 (SD=4.56, CV=9.6%)	47.99 (SD=3.76, CV=7.8%)	-0.35	-0.42 (98.55) r $\approx$ 0.04	0.673 (n.s.)	[-1.99; 1.29]
MWR1	1827.00 (SD=149.93, CV=8.2%)	1689.31 (SD=172.27, CV=10.2%)	+137.69	4.28 (95.07) r $\approx$ 0.4	<0.001 (***)	[74.1; 201.3]
MWR2	892.54 (SD=62.00, CV=7%)	863.36 (SD=63.49, CV=7.4%)	+29.18	2.34 (98.58) r $\approx$ 0.23	<0.05 (*)	[4.5; 53.8]
List head ratio	9.18 (SD=2.358, CV=25.7%)	8.60 (SD=2.213, CV=25.7%)	+0.58	1.27 (99.7) r $\approx$ 0.13	0.206 (n.s.)	[-0.32; 1.48]
Mean word length (chars)	6.83 (SD=0.27, CV=4%)	6.63 (SD=0.3, CV=4.5%)	+0.20	3.36 (96.33) r $\approx$ 0.32	0.001 (**)	[0.08; 0.31]
Mean word length (syllables)	2.92 (SD=0.13, CV=4.5%)	2.84 (SD=0.13, CV=4.6%)	+0.08	3.20 (99.39) r $\approx$ 0.31	<0.01 (**)	[0.03; 0.13]
Mean sentence length (chars)	175.28 (SD=44.31, CV=25.3%)	145.76 (SD=19.92, CV=13.7%)	+29.52	4.39 (73.46) r $\approx$ 0.46	<0.001 (***)	[16.1; 42.9]
Mean sentence length (tokens)	30.38 (SD=7.86, CV=25.9%)	25.58 (SD=3.27, CV=12.8%)	+4.8	4.08 (70.67) r $\approx$ 0.44	<0.001 (***)	[2.45; 7.14]
Mean sentence length (clauses)	2.2 (SD=0.38, CV=17.63%)	2.04 (SD=0.23, CV=11.22%)	+0.20	1.872 (86.40) r $\approx$ 0.20	0.065 (n.s.)	[-0.01; 0.24]
Mean clause length (tokens)	11.5 (SD=1.23, CV=10.73%)	10.5 (SD=0.93, CV=8.84%)	+1	4.641 (96.03) r $\approx$ 0.43	<0.001 (***)	[0.57; 1.43]
Simple sentences ratio	58.63 (SD=9.29, CV=15.84%)	56.63 (SD=7.63, CV=13.47%)	+2	1.190 (98.49) r $\approx$ 0.12	0.237 (n.s.)	[-1.33; 5.33]
Flesch reading ease score *	-0.5 (SD=10.98, CV=-)	8.91 (SD=8.52, CV=-)	-9.41	-4.85 (97.0) r $\approx$ 0.44	<0.001 (***)	[-13.26; -5.56]

Знаки препинания						
.	3.38 (SD=0.74, CV=21.8%)	3.83 (SD=0.52, CV=13.6%)	-0.45	-3.60 (93.5) $r \approx 0.35$	<0.001 (***)	[-0.70; -0.20]
–	0.70 (SD=0.38, CV=54.5%)	0.55 (SD=0.25, CV=45.0%)	+0.16	2.45 (89.5) $r \approx 0.25$	<0.05 (*)	[0.03; 0.28]
)	0.88 (SD=0.51, CV=58.4%)	0.67 (SD=0.44, CV=65.4%)	+0.21	2.24 (99.1) $r \approx 0.22$	<0.05 (*)	[0.02; 0.40]
(	0.85 (SD=0.51, CV=60.1%)	0.66 (SD=0.44, CV=66.8%)	+0.19	2.05 (99.3) $r \approx 0.20$	<0.05 (*)	[0.01; 0.38]
«	1.28 (SD=0.77, CV=60.6%)	0.71 (SD=0.60, CV=84.4%)	+0.56	4.12 (97.1) $r \approx 0.39$	<0.001 (***)	[0.29; 0.84]
»	1.28 (SD=0.77, CV=60.5%)	0.71 (SD=0.60, CV=84.4%)	+0.56	4.11 (97.2) $r \approx 0.39$	<0.001 (***)	[0.29; 0.83]
МЕД-ПИ						
Индикатор	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	Welch t (df) r	p	95% CI $\Delta$
Lexical density	66.75 (SD=3.16, CV=4.7%)	64.84 (SD=2.8, CV=4.3%)	+1.91	+3.53 (119.5) $r \approx 0.31$	<0.001 (***)	[0.84; 2.98]
TTR	29.12 (SD=4.70, CV=16.1%)	26.27 (SD=6.41, CV=24.4%)	+2.85	+2.79 (104.9) $r \approx 0.26$	<0.01 (**)	[0.83; 4.89]
Log TTR	84.56 (SD=1.83, CV=2.2%)	83.27 (SD=2.41, CV=2.9%)	+1.28	+3.29 (107.0) $r \approx 0.30$	<0.01 (**)	[0.51; 2.06]
Modified TTR	1760.88 (SD=141.09, CV=8.0%)	1676.61 (SD=165.72, CV=9.9%)	+84.27	+3.01 (113.4) $r \approx 0.27$	<0.01 (**)	[28.79; 139.76]
STTR	43.17 (SD=4.58, CV=10.6%)	40.41 (SD=5.20, CV=12.9%)	+2.75	+3.09 (115.0) $r \approx 0.28$	<0.01 (**)	[0.99; 4.52]
MWR1	2631.42 (SD=266.39, CV=10.1%)	2357.54 (SD=262.06, CV=11.1%)	+273.88	+5.71 (119.2) $r \approx 0.46$	<0.001 (***)	[179; 368.8]
MWR2	1041.56 (SD=82.82, CV=8.0%)	1027.10 (SD=77.78, CV=7.6%)	+14.47	+0.99 (119.7) $r \approx 0.09$	0.3224 (n.s.)	[-14.36; 43.3]
List head ratio	15.12 (SD=3.910, CV=25.9%)	16.25 (SD=3.757, CV=23.1%)	-1.13	-1.62 (119.5) $r \approx 0.15$	0.108 (n.s.)	[-2.5; 0.25]
Mean word length (chars)	6.88 (SD=0.32, CV=4.7%)	6.80 (SD=0.27, CV=4%)	+0.08	+1.53 (118.2) $r \approx 0.14$	0.130 (n.s.)	[-0.02; 0.19]
Mean word length (syllables)	2.96 (SD=0.16, CV=5.4%)	2.92 (SD=0.13, CV=4.5%)	+0.04	+1.54 (118.6) $r \approx 0.14$	0.127 (n.s.)	[-0.01; 0.09]
Mean sentence length	155.56	157.69	-2.13	-0.46 (112.7)	0.643 (n.s.)	[-11.2; 6.94]

(chars)	(SD=22.88, CV=14.7%)	(SD=27.23, CV=17.3%)		r≈0.04		
Mean sentence length (tokens)	26.45 (SD=3.85, CV=14.6%)	26.88 (SD=5.10, CV=19.0%)	-0.42	-0.51 (106.7) r≈0.05	0.608 (n.s.)	[-2.06; 1.21]
Mean sentence length (clauses)	1.7 (SD=0.214, CV=12.64%)	1.83 (SD=0.256, CV=14.04%)	-0.13	-3.095 (112.37) r≈0.28	<0.01 (**)	[-0.22; -0.05]
Mean clause length (tokens)	13.5 (SD=1.68, CV=12.45%)	12.8 (SD=1.88, CV=14.7%)	+0.70	2.171 (115.54) r≈0.20	0.032 (n.s.)	[0.06; 1.35]
Simple sentences ratio	73.06 (SD=8.07, CV=11.04%)	68.64 (SD=8.91, CV=12.98%)	+4.42	2.861 (116.1) r≈0.26	<0.01 (**)	[1.36; 7.48]
Flesch reading ease score	2.16 (SD=10.09, CV=466.7%)	3.32 (SD=7.62, CV=229.7%)	-1.16	-0.72 (115.17) r≈0.07	0.475 (n.s.)	[-4.35; 2.04]
<b>Знаки препинания</b>						
-	0.55 (SD=0.337, CV=61.6%)	0.27 (SD=0.249, CV=93.1%)	+0.28	5.25 (114.09) r≈0.44	<0.001 (***)	[0.18; 0.39]
)	1.81 (SD=0.68, CV=37.7%)	1.29 (SD=0.64, CV=49.3%)	+0.52	4.34 (119.69) r≈0.37	<0.001 (***)	[0.28; 0.76]
(	1.78 (SD=0.70, CV=39.3%)	1.25 (SD=0.60, CV=48.3%)	+0.53	4.52 (119.09) r≈0.38	<0.001 (***)	[0.3; 0.77]
<b>МЕД-МП</b>						
Индикатор	Non-translated (weighted mean)	Translated (weighted mean)	Δ (NT-T)	Welch t (df) r	p	95% CI Δ
Lexical density	66.75 (SD=3.16, CV=4.7%)	62.74 (SD=3.64, CV=5.8%)	+4.01	5.99 (88.6) r≈0.54	<0.001 (***)	[2.68; 5.33]
TTR	29.12 (SD=4.70, CV=16.1%)	24.01 (SD=3.47, CV=14.4%)	+5.11	6.55 (107.1) r≈0.53	<0.001 (***)	[3.57; 6.67]
Log TTR	84.56 (SD=1.83, CV=2.2%)	82.75 (SD=1.57, CV=1.9%)	+1.81	5.52 (104.0) r≈0.47	<0.001 (***)	[1.16; 2.45]
Modified TTR	1760.88 (SD=141.09, CV=8.0%)	1729.73 (SD=126.25, CV=7.3%)	+31.15	1.21 (102.5) r≈0.12	0.229 (n.s.)	[-19.9; 82.2]
STTR	43.17 (SD=4.58, CV=10.6%)	38.55 (SD=3.63, CV=9.4%)	+4.62	5.87 (106.3) r≈0.49	<0.001 (***)	[3.06; 6.18]
MWR1	2631.42 (SD=266.39, CV=10.1%)	2529.52 (SD=390.2, CV=15.4%)	+101.9	+1.53 (74.3) r≈0.17	0.13 (n.s.)	[-30.83; 234.63]
MWR2	1041.56 (SD=82.82, CV=8.0%)	988.49 (SD=89.33, CV=9.04%)	+53.07	+3.16 (92.5) r≈0.31	<0.01 (**)	[19.71; 86.45]
List head ratio	15.12 (SD=3.910, CV=25.9%)	15.37 (SD=3.002, CV=19.5%)	-0.25	-0.38 (106.8) r≈0.04	0.706 (n.s.)	[-1.56; 1.06]

Mean word length (chars)	6.88 (SD=0.32, CV=4.7%)	6.57 (SD=0.29, CV=4.4%)	+0.31	5.28 (102.5) $r \approx 0.46$	<0.001 (***)	[0.19; 0.43]
Mean word length (syllables)	2.96 (SD=0.16, CV=5.3%)	2.90 (SD=0.11, CV=3.9%)	+0.06	2.36 (107.1) $r \approx 0.22$	<0.05 (*)	[0.01; 0.11]
Mean sentence length (chars)	155.56 (SD=22.88, CV=14.7%)	167.43 (SD=24.26, CV=14.5%)	-11.87	-2.58 (93.5) $r \approx 0.26$	<0.05 (*)	[-21; -2.74]
Mean sentence length (tokens)	26.45 (SD=3.85, CV=14.6%)	30.08 (SD=4.35, CV=14.5%)	-3.63	-4.52 (89.7) $r \approx 0.43$	0.01 (**)	[-5.23; -2.03]
Mean sentence length (clauses)	1.7 (SD=0.214, CV=12.64%)	1.9 (SD=0.169, CV=8.96%)	-0.20	-5.555 (106.21) $r \approx 0.47$	<0.001 (***)	[-0.28; -0.13]
Mean clause length (tokens)	13.5 (SD=1.68, CV=12.45%)	13.76 (SD=1.87, CV=13.6%)	-0.26	-0.729 (90.54) $r \approx 0.08$	0.467 (n.s.)	[-0.94, 0.44]
Simple sentences ratio	73.06 (SD=8.07, CV=11.04%)	64.92 (SD=6.46, CV=9.95%)	+8.147	5.852 (106.06) $r \approx 0.49$	<0.001 (***)	[5.39; 10.91]
Flesch reading ease score	2.16 (SD=10.09, CV=466.7%)	4.27 (SD=8.88, CV=208%)	-2.11	-1.15 (103.13) $r \approx 0.11$	0.251 (n.s.)	[-5.72; 1.51]
<b>Знаки препинания</b>						
.	3.76 (SD=0.63, CV=16.8%)	3.37 (SD=0.46, CV=13.8%)	+0.39	3.71 (107.07) $r \approx 0.34$	<0.001 (***)	[0.18; 0.60]
,	6.40 (SD=1.07, CV=16.7%)	7.04 (SD=0.90, CV=12.8%)	-0.64	-3.37 (104.87) $r \approx 0.31$	<0.001 (***)	[-1.01; -0.26]
–	0.55 (SD=0.34, CV=61.6%)	0.07 (SD=0.08, CV=120.1%)	+0.48	10.96 (71.25) $r \approx 0.79$	<0.001 (***)	[0.4; 0.57]
)	1.81 (SD=0.68, CV=37.7%)	2.12 (SD=0.64, CV=29.9%)	-0.31	-2.44 (100.76) $r \approx 0.24$	<0.05 (*)	[-0.56; -0.06]
(	1.78 (SD=0.70, CV=39.3%)	2.12 (SD=0.64, CV=30.1%)	-0.34	-2.65 (101.60) $r \approx 0.25$	$p < 0.001$ (***)	[-0.60; -0.09]

### Используемые в таблице обозначения:

- **weighted mean** – взвешенное среднее значение анализируемого показателя;
- **SD** – стандартное отклонение (измеряется в тех же единицах, что и среднее значение);
- **CV** – коэффициент вариации;
- **NT** – непереводные тексты на русском языке (non-translated texts);
- **T** – переводные тексты на русском языке (translated texts);
- **$\Delta$  (NT–T)** – разность взвешенных средних непереводными текстами и переводами;
- **Welch t** и **df** - значение t-статистики и количество степеней свободы;
- **p-value** – достигнутый уровень значимости (n. s. – результат статистически незначим (т.е.  $p \geq 0.05$ ); \* –  $p < 0.05$ ; \*\* –  $p < 0.01$ ; \*\*\* –  $p < 0.001$ );

- **95% CI  $\Delta$**  – 95%-ый доверительный интервал для разности средних.

\* Коэффициент вариации (CV) не приводится по причине того, что в непереводном подкорпусе среднее значение показателя, близкое к нулю (weighted mean=-0.50), делает CV неинтерпретируемым. В связи с этим для большей интерпретируемости данных дополнительно применялся непараметрический U-теста Манна-Уитни, который подтвердил сдвиг распределений в пользу переводов:  $U=735.0$ ,  $Z=-5.412$ ,  $p<0.001$ ;  $HL=-9.96$  при 95% ДИ [-13.176; -6.781]; вероятность превосходства  $A_{12}\approx 0.21$  (т.е.  $\approx 79\%$  шансов, что случайный перевод имеет более высокий индекс удобочитаемости, чем случайный непереводной текст);  $r\approx 0.5$  (крупный эффект).

## ПРИЛОЖЕНИЕ Е

### Результаты анализа индикатора лексической повторяемости по корпусам

Lexical repetition						
Корпус	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	Welch t (df) r	<i>p</i>	95% CI $\Delta$
СОЦ-ПП	9.65 (SD=0.93, CV=9.58%)	9.3 (SD=0.76, CV=8.16%)	+0.35	2.121 (98.48) r $\approx$ 0.21	<i>&lt;0.05</i> (* )	[0.02; 0.69]
МЕД-ПП	9.49 (SD=1.34, CV=14.1%)	9.02 (SD=1.39, CV=15.37%)	+0.47	1.899 (118.17) r $\approx$ 0.17	<i>0.06</i> ( <i>n.s.</i> )	[-0.02; 0.96]
МЕД-МП	9.49 (SD=1.34, CV=14.1%)	7.73 (SD=0.75; CV=9.75%)	+1.76	8.752 (101.53) r $\approx$ 0.65	<i>&lt;0.001</i> (*** )	[1.37; 2.17]

#### Используемые в таблице обозначения:

- **weighted mean** – взвешенное среднее значение анализируемого показателя;
- **SD** – стандартное отклонение (измеряется в тех же единицах, что и среднее значение);
- **CV** – коэффициент вариации;
- **NT** – непереводные тексты на русском языке (non-translated texts);
- **T** – переводные тексты на русском языке (translated texts);
- **$\Delta$  (NT-T)** – разность взвешенных средних непереводными текстами и переводами;
- **Welch t** и **df** - значение t-статистики и количество степеней свободы;
- **p-value** – достигнутый уровень значимости (n. s. – результат статистически незначим (т.е.  $p \geq 0.05$ ); \* –  $p < 0.05$ ; \*\* –  $p < 0.01$ ; \*\*\* –  $p < 0.001$ );
- **r** – размер наблюдаемого эффекта (~0.1 маленький, ~0.3 средний, ~0.5 большой);
- **95% CI  $\Delta$**  – 95%-ый доверительный интервал для разности средних.

## ПРИЛОЖЕНИЕ Ё

## Результаты анализа индикаторов экспликации по корпусам

СОЦ-ПП						
Индикатор	Non-translated (median) [Q <sub>1</sub> -Q <sub>3</sub> ], IQR, MAD	Translated (median) [Q <sub>1</sub> -Q <sub>3</sub> ], IQR, MAD	Mann-Whitney U (Z)	Effect size: r, A <sub>12</sub>	p	HL 95% CI
Индикаторы, связанные с употреблением имен собственных						
Explicit naming	187.9 [121.72-298.28], IQR=176.56, MAD=82.22	300 [167.26-574.02], IQR=406.76, MAD=157.46	1257.5 (-2.6)	r=0.24, A <sub>12</sub> ≈0.36	<0.01 (**)	-81.09, [-160; -20.59]
Single naming	58.82 [35.12-73.56], IQR=38.44, MAD=17.92	66.67 [51.73-76.21], IQR=24.48, MAD=11.11	1390 (-1.89)	r≈0.17, A <sub>12</sub> ≈0.4	0.06 (n.s.)	-6.94, [-16.04; 0.36]
Mean named entities length	1.53 [1.309-1.784], IQR=0.48, MAD=0.243	1.45 [1.296-1.769], IQR=0.47, MAD=0.191	1558.5 (+0.98)	r=0.09, A <sub>12</sub> ≈0.55	0.33 (n.s.)	+0.052, [-0.06; 0.17]
Дискурсивные маркеры локального уровня связности текста						
Введение новой информации	0.94 [0.796-1.204], IQR=0.408, MAD=0.205	0.85 [0.682-0.963], IQR=0.281, MAD=0.142	1218.5 (2.81)	r=0.26, A <sub>12</sub> ≈0.65	<0.01 (**)	+0.14, [0.05; 0.24]
Противопоставление	0.84 [0.629-1.043], IQR=0.414, MAD=0.212	0.97 [0.810-1.102], IQR=0.292, MAD=0.153	1314 (-2.296)	r≈0.21, A <sub>12</sub> ≈0.38	<0.05 (*)	-0.12, [-0.21; -0.02]
Индикатор	Non-translated (weighted mean)	Translated (weighted mean)	Δ (NT-T)	Welch t (df) r	p	95% CI Δ
Лексическая когезия (косинусное сходство)						
Word pairwise cosine	0.172 (SD=0.016, CV=9.30%)	0.160 (SD=0.014, CV=8.75%)	+0.012	4.21 (99.68) r≈0.39	<0.001 (***)	[0.007; 0.018]
Sentence cosine (adjacent)	0.700 (SD=0.052, CV=7.43%)	0.655 (SD=0.048, CV=7.33%)	+0.045	4.46 (99.72) r≈0.41	<0.001 (***)	[0.025; 0.064]
МЕД-ПП						
Индикатор	Non-translated (median) [Q <sub>1</sub> -Q <sub>3</sub> ], IQR, MAD	Translated (median) [Q <sub>1</sub> -Q <sub>3</sub> ], IQR, MAD	Mann-Whitney U (Z)	Effect size: r, A <sub>12</sub>	p	HL 95% CI
Индикаторы, связанные с употреблением имен собственных						
Explicit naming	164.9 [103.07-309.38], IQR=206.31, MAD=83.25	157.9 [82.09-344.87], IQR=262.78, MAD=102.11	2395.5 (0.369)	r=0.031; A <sub>12</sub> ≈0.52	0.714 (n.s.)	+10.1, [-43.33; 56.56]
Single naming	50.00 [32.02-66.13], IQR=34.11, MAD=16.67	66.67 [54.93-83.33], IQR=28.41, MAD=16.67	1707 (-3.210)	r=0.27; A <sub>12</sub> ≈0.34	<0.01 (**)	-16.67, [-25.0; -7.14]
	1.91	1.56	1470	r≈0.35;		+0.44,

Mean named entities length	[1.60–2.44], IQR=0.84, MAD=0.45	[1.13–1.88], IQR=0.75, MAD=0.40	(4.188)	$A_{12} \approx 0.7$	$<0.001$ (***)	[0.24; 0.64]
<b>Дискурсивные маркеры локального уровня связности текста</b>						
Введение новой информации	0.85 [0.65–1.09], IQR=0.438, MAD=0.224	0.98 [0.74–1.38], IQR=0.643, MAD=0.312	1835 (-2.680)	$r \approx 0.23$ ; $A_{12} \approx 0.37$	$<0.01$ (**)	-0.15, [-0.27; -0.05]
Противопоставление	0.54 [0.4–0.64], IQR=0.245, MAD=0.137	0.7 [0.4905–0.93], IQR=0.44, MAD=0.225	1489 (-4.107)	$r \approx 0.35$ ; $A_{12} \approx 0.30$	$<0.001$ (***)	-0.188, [-0.27; -0.10]
Введение примеров	0.07 [0.032–0.129], IQR=0.1, MAD=0.053	0.16 [0.09–0.26], IQR=0.17, MAD=0.087	1403 (-4.469)	$r \approx 0.38$ ; $A_{12} \approx 0.28$	$<0.001$ (***)	-0.084, [-0.125; -0.05]
Причинно-следственные отношения	0.47 [0.29–0.59], IQR=0.3, MAD=0.162	0.53 [0.3785–0.7525], IQR=0.374, MAD=0.161	1997.5 (-2.01)	$r \approx 0.17$ ; $A_{12} \approx 0.4$	$<0.05$ (*)	-0.076, [-0.156; -0.003]
Целевые отношения	0.57 [0.4–0.89], IQR=0.49, MAD=0.21	0.86 [0.51–1.14], IQR=0.63, MAD=0.33	1686 (-3.295)	$r \approx 0.28$ ; $A_{12} \approx 0.34$	$0.001$ (**)	-0.22, [-0.35; -0.09]
<b>Индикатор</b>	<b>Non-translated (weighted mean)</b>	<b>Translated (weighted mean)</b>	<b><math>\Delta</math> (NT–T)</b>	<b>Welch t (df) r</b>	<b>p</b>	<b>95% CI <math>\Delta</math></b>
<b>Лексическая когезия (косинусное сходство)</b>						
Word pairwise cosine	0.174 (SD=0.009, CV=5.26%)	0.175 (SD=0.011, CV=6.08%)	-0.001	-0.46 (113.96) $r \approx 0.04$	$0.645$ (n.s.)	[-0.004; 0.003]
Sentence cosine (adjacent)	0.683 (SD=0.04, CV=5.88%)	0.690 (SD=0.047, CV=6.84%)	-0.007	-0.89 (113.36) $r \approx 0.08$	$0.373$ (n.s.)	[-0.023; 0.009]
<b>МЕД-МП</b>						
<b>Индикатор</b>	<b>Non-translated (median) [Q<sub>1</sub>–Q<sub>3</sub>], IQR, MAD</b>	<b>Translated (median) [Q<sub>1</sub>–Q<sub>3</sub>], IQR, MAD</b>	<b>Mann-Whitney U (Z)</b>	<b>Effect size: r, A<sub>12</sub></b>	<b>p</b>	<b>HL 95% CI</b>
<b>Индикаторы, связанные с употреблением имен собственных</b>						
Explicit naming	164.9 [103.07–309.38], IQR=206.31, MAD=83.25	99.59 [67.74–191.15], IQR=123.412, MAD=41.71	1248 (2.672)	$r \approx 0.24$ ; $A_{12} \approx 0.64$	$<0.01$ (**)	+51.87, [15.9; 91.08]
Single naming	50.00 [32.02–66.13], IQR=34.11, MAD=16.67	72.78 [65.129–77.419], IQR=12.29, MAD=7.006	789 (-5.117)	$r \approx 0.47$ ; $A_{12} \approx 0.23$	$<0.001$ (***)	-21.46, [-28.54; -13.76]
Mean named entities length	1.91 [1.60–2.44], IQR=0.84, MAD=0.45	1.562 [1.32–1.88], IQR=0.55, MAD=0.255	1057 (3.69)	$r \approx 0.34$ ; $A_{12} \approx 0.7$	$<0.001$ (***)	+0.35, [0.17; 0.52]
<b>Дискурсивные маркеры локального уровня связности текста</b>						
Введение примеров	0.07	0.16	1024.5 (-3.867)	$r \approx 0.35$ ; $A_{12} \approx 0.29$	$<0.001$ (***)	-0.073,

	[0.032–0.129], IQR=0.1, MAD=0.053	[0.092–0.253], IQR=0.16, MAD=0.083				[-0.12; -0.04]
Целевые отношения	0.57 [0.4–0.89], IQR=0.49, MAD=0.21	1.36 [1.1020–1.6418], IQR=0.54, MAD=0.28	303 (-7.703)	$r \approx 0.70$ ; $A_{12} \approx 0.09$	$< 0.001$ (***)	-0.73, [-0.88; -0.59]
Дискурсивные маркеры глобального уровня связности текста						
Обеспечение порядка следования информации	0.4 [0.21–0.68], IQR=0.47, MAD=0.237	0.55 [0.378–0.766], IQR=0.39, MAD=0.174	1304.5 (-2.371)	$r \approx 0.22$ ; $A_{12} \approx 0.37$	$< 0.05$ (*)	-0.134, [-0.25; -0.03]
Выводы и заключение	0.22 [0.119–0.319], IQR=0.2, MAD=0.103	0.18 [0.101–0.250], IQR=0.15, MAD=0.075	1323.5 (2.27)	$r \approx 0.21$ ; $A_{12} \approx 0.62$	$< 0.05$ (*)	+0.05, [0.01; 0.1]
<b>Индикатор</b>	<b>Non-translated (weighted mean)</b>	<b>Translated (weighted mean)</b>	<b><math>\Delta</math> (NT–T)</b>	<b>Welch t (df) r</b>	<b>p</b>	<b>95% CI <math>\Delta</math></b>
Лексическая когезия (косинусное сходство)						
Word pairwise cosine	0.174 (SD=0.009, CV=5.26%)	0.170 (SD=0.009, CV=5.48%)	+0.004	2.48 (96.11) $r \approx 0.24$	$< 0.05$ (*)	[0.001; 0.008]
Sentence cosine (adjacent)	0.683 (SD=0.040, CV=5.88%)	0.694 (SD=0.032, CV=4.55%)	-0.011	-1.64 (106.41) $r \approx 0.16$	0.104 (n.s)	[-0.025; 0.002]

### Используемые в таблице обозначения:

- **p-value** – достигнутый уровень значимости (n. s. – результат статистически незначим (т.е.  $p \geq 0.05$ ); \* –  $p < 0.05$ ; \*\* –  $p < 0.01$ ; \*\*\* –  $p < 0.001$ );
- **weighted mean** – взвешенное среднее значение анализируемого показателя;
- **SD** – стандартное отклонение (измеряется в тех же единицах, что и среднее значение);
- **CV** – коэффициент вариации;
- **NT** – непереводные тексты на русском языке (non-translated texts);
- **T** – переводные тексты на русском языке (translated texts);
- **$\Delta$  (NT–T)** – разность взвешенных средних непереводными текстами и переводами;
- **Welch t** и **df** - значение t-статистики и количество степеней свободы;
- **95% CI  $\Delta$**  – 95%-ый доверительный интервал для разности средних;
- **median** – медианное значение анализируемого показателя;
- **Q1–Q3** – границы интервала, в котором находится 50% значений показателя. Q1 соответствует 25-му перцентилю, Q3 – 75-му, разница между ними отражает степень разброса данных;
- **IQR (Interquartile Range)** – разность между третьим и первым квартилем (Q3–Q1);

- **MAD** (Median Absolute Deviation) – медиана модулей отклонений всех наблюдений от общей медианы, отражающая типичную величину разброса значений относительно центра распределения;
- **Mann-Whitney U и z** – значения U-статистики и z-оценки;
- **r** – размер наблюдаемого эффекта (~0.1 маленький, ~0.3 средний, ~0.5 большой);
- **A<sub>12</sub>** – показатель вероятности того, что при случайном выборе по 1-му тексту из обоих подкорпусов значение показателя в непереводном тексте будет выше, чем соответствующее значение в переводном;
- **HL, 95% CI** – сдвиг Ходжеса–Лемана HL с 95%-доверительным интервалом.

## ПРИЛОЖЕНИЕ Ж

Результаты анализа нормализованных частот *частеречных n-грамм* по корпусам

СОЦ-ПП						
N-грамма	Non-translated (weighted mean)	Translated (weighted mean)	$\Delta$ (NT-T)	Welch t (df) r	p	95% CI $\Delta$
Частеречные униграммы						
NOUN	29.44 (SD=2.35, CV=8%)	27.48 (SD=2.02, CV=7.3%)	+1.96	4.52 (99.2) r $\approx$ 0.41	<0.001 (***)	[1.10; 2.82]
ADJF	9.76 (SD=1.43, CV=14.7%)	8.96 (SD=1.63, CV=18.2%)	+0.80	+2.63 (95.3) r $\approx$ 0.26	0.01 (*)	[0.20; 1.41]
VERB	4.95 (SD=0.83, CV=16.7%)	5.71 (SD=0.57, CV=10%)	-0.76	-5.42 (92.7) r $\approx$ 0.49	<0.001 (***)	[-1.04; -0.48]
DMSTR_PRN	2.13 (SD=0.39, CV=18.6%)	2.5 (SD=0.51, CV=20.4%)	-0.37	-4.11 (89.7) r $\approx$ 0.40	<0.001 (***)	[-0.55; -0.19]
SCONJ	2.02 (SD=0.43, CV=21.3%)	2.61 (SD=0.45, CV=17.2%)	-0.59	-6.70 (98.3) r $\approx$ 0.56	<0.001 (***)	[-0.76; -0.41]
ADVB	1.91 (SD=0.5, CV=26.3%)	2.19 (SD=0.49, CV=22.4%)	-0.28	-2.88 (99.3) r $\approx$ 0.28	<0.01 (**)	[-0.48; -0.09]
INFN	1.58 (SD=0.41, CV=25.8%)	1.87 (SD=0.63, CV=33.7%)	-0.29	-2.74 (81.1) r $\approx$ 0.29	<0.01 (**)	[-0.50; -0.08]
PRN_CONJ_W	0.73 (SD=0.27, CV=37%)	0.93 (SD=0.27, CV=29.2%)	-0.20	-3.7 (98.7) r $\approx$ 0.35	<0.001 (***)	[-0.3; -0.09]
PER_PRN	0.6 (SD=0.279, CV=44%)	1.02 (SD=0.327, CV=31%)	-0.42	-6.70 (94.3) r $\approx$ 0.57	<0.001 (***)	[-0.53; -0.29]
MDL_VRB	0.33 (SD=0.13, CV=39.4%)	0.48 (SD=0.22, CV=45.8%)	-0.14	-3.85 (76.2) r $\approx$ 0.40	<0.001 (***)	[-0.2; -0.07]
MDL_WRD	0.27 (SD=0.1, CV=37%)	0.33 (SD=0.16, CV=48.4%)	-0.06	-2.24 (77.2) r $\approx$ 0.25	<0.05 (*)	[-0.11; -0.01]
COPULA	0.16 (SD=0.122, CV=74%)	0.26 (SD=0.136, CV=52%)	-0.10	-3.73 (96.1) r $\approx$ 0.36	<0.001 (***)	[-0.15; -0.045]
INDF_ADJ_PRN	0.11 (SD=0.08, CV=71.2%)	0.18 (SD=0.09, CV=48.8%)	-0.07	-4.43 (95.1) r $\approx$ 0.41	<0.001 (***)	[-0.11; -0.04]
SPRL	0.09 (SD=0.06, CV=67.3%)	0.13 (SD=0.068, CV=53.2%)	-0.04	-2.97 (95.9) r $\approx$ 0.29	<0.01 (**)	[-0.06; -0.013]
SBJ_PRCL	0.07 (SD=0.06, CV=85%)	0.15 (SD=0.11, CV=74%)	-0.08	-4.4 (72.0) r $\approx$ 0.46	<0.001 (***)	[-0.12; -0.04]

INDF_NUM_PRN	0.04 (SD=0.035, CV=90%)	0.06 (SD=0.042, CV=70.4%)	-0.02	-2.7 (93.2) $r \approx 0.27$	<0.01 (**)	[-0.04; -0.01]
<b>Частеречные биграммы</b>						
NOUN NOUN	5.80 (SD=1.219, CV=21.0%)	4.93 (SD=0.960, CV=19.5%)	+0.87	4.01 (97.44) $r \approx 0.38$	<0.001 (***)	[0.44; 1.30]
NOUN ADJF	2.33 (SD=0.648, CV=27.8%)	1.86 (SD=0.605, CV=32.5%)	+0.47	3.76 (99.70) $r \approx 0.35$	<0.001 (***)	[0.22; 0.71]
COMMA SCONJ	1.21 (SD=0.313, CV=25.8%)	1.74 (SD=0.342, CV=19.6%)	-0.53	-8.10 (96.88) $r \approx 0.63$	<0.001 (***)	[-0.66; -0.4]
PER_PRN VERB	0.29 (SD=0.155, CV=53.3%)	0.54 (SD=0.212, CV=39.5%)	-0.25	-6.66 (86.86) $r \approx 0.58$	<0.001 (***)	[-0.32; -0.17]
COMMA PRN CONJ W	0.44 (SD=0.189, CV=43.0%)	0.62 (SD=0.202, CV=32.5%)	-0.18	-4.67 (97.55) $r \approx 0.43$	<0.001 (***)	[-0.26; -0.10]
ADVB VERB	0.32 (SD=0.124, CV=38.7%)	0.43 (SD=0.131, CV=30.8%)	-0.11	-4.24 (97.66) $r \approx 0.39$	<0.001 (***)	[-0.16; -0.06]
MDL_VRB INFN	0.22 (SD=0.101, CV=46.3%)	0.32 (SD=0.161, CV=50.4%)	-0.10	-3.79 (78.84) $r \approx 0.39$	<0.001 (***)	[-0.16; -0.05]
S_START PER_PRN	0.07 (SD=0.065, CV=91.8%)	0.17 (SD=0.102, CV=59.1%)	-0.10	-5.96 (79.66) $r \approx 0.56$	<0.001 (***)	[-0.14; -0.07]
COMP_MOD ADJF	0.11 (SD=0.071, CV=64.5%)	0.18 (SD=0.083, CV=46.4%)	-0.07	-4.36 (94.86) $r \approx 0.41$	<0.001 (***)	[-0.10; -0.04]
<b>Частеречные триграммы</b>						
NOUN ADJF NOUN	1.94 (SD=0.578, CV=29.8%)	1.59 (SD=0.530, CV=33.3%)	+0.35	3.18 (99.72) $r \approx 0.30$	<0.01 (**)	[0.13; 0.57]
NOUN COMMA NOUN	0.94 (SD=0.427, CV=45.3%)	0.63 (SD=0.253, CV=40.5%)	+0.32	4.61 (85.61) $r \approx 0.44$	<0.001 (***)	[0.18; 0.46]
NOUN NOUN NOUN	1.03 (SD=0.402, CV=39.2%)	0.79 (SD=0.322, CV=40.9%)	+0.24	3.35 (97.86) $r \approx 0.32$	<0.01 (**)	[0.01; 0.38]
ADJF NOUN NOUN	1.22 (SD=0.295, CV=24.2%)	0.99 (SD=0.221, CV=22.4%)	+0.23	4.54 (95.85) $r \approx 0.42$	<0.001 (***)	[0.13; 0.34]
NOUN NOUN ADJF	0.41 (SD=0.192, CV=46.8%)	0.26 (SD=0.125, CV=48.6%)	+0.15	4.83 (89.97) $r \approx 0.45$	<0.001 (***)	[0.09; 0.22]
NOUN COMMA SCONJ	0.4 (SD=0.114, CV=28.7%)	0.53 (SD=0.154, CV=29.0%)	-0.13	-4.99 (87.55) $r \approx 0.47$	<0.001 (***)	[-0.19; -0.08]
NOUN COMMA PRN_CONJ_W	0.33 (SD=0.170, CV=51.2%)	0.48 (SD=0.171, CV=35.7%)	-0.15	-4.33 (98.93) $r \approx 0.40$	<0.001 (***)	[-0.21; -0.08]
VERB COMMA SCONJ	0.20 (SD=0.112, CV=55.7%)	0.32 (SD=0.144, CV=44.4%)	-0.12	-4.78 (90.08) $r \approx 0.45$	<0.001 (***)	[-0.17; -0.07]

COMMA SCONJ NOUN	0.3 (SD=0.108, CV=36.3%)	0.45 (SD=0.118, CV=26.3%)	-0.15	-6.82 (96.66) r≈0.57	<0.001 (***)	[-0.2; -0.11]
DMSTR_PRN COMMA SCONJ	0.24 (SD=0.125, CV=51.5%)	0.37 (SD=0.132, CV=35.7%)	-0.126	-4.91 (97.95) r≈0.44	<0.001 (***)	[-0.18; -0.08]
COMMA PRN_CONJ_W VERB	0.19 (SD=0.095, CV=50.3%)	0.28 (SD=0.120, CV=42.6%)	-0.09	-4.32 (90.89) r≈0.41	<0.001 (***)	[-0.14; -0.05]
<b>МЕД-ПН</b>						
<b>N-грамма</b>	<b>Non-translated (weighted mean)</b>	<b>Translated (weighted mean)</b>	<b>Δ (NT-T)</b>	<b>Welch t (df)</b>	<b>p</b>	<b>95% CI Δ</b>
<b>Частеречные униграммы</b>						
NOUN	34.83 (SD=1.95, CV=5.6%)	33.07 (SD=1.82, CV=5.5%)	+1.76	+5.15 (119.7) r≈0.43	<0.001 (***)	[1.08; 2.43]
PRTF	1.78 (SD=0.50, CV=28.0%)	2.05 (SD=0.46, CV=22.4%)	-0.27	-3.15 (119.7) r≈0.28	<0.001 (***)	[-0.45; -0.10]
ADVB	1.27 (SD=0.4, CV=31.5%)	1.62 (SD=0.51, CV=31.3%)	-0.36	-4.26 (109.0) r≈0.38	<0.001 (***)	[-0.52; -0.19]
DMSTR_PRN	1.23 (SD=0.39, CV=31.8%)	1.72 (SD=0.58, CV=33.8%)	-0.49	-5.38 (99.7) r≈0.47	<0.001 (***)	[-0.67; -0.31]
SCONJ	1.09 (SD=0.41, CV=37.1%)	1.54 (SD=0.50, CV=32.4%)	-0.45	-5.38 (111.0) r≈0.45	<0.001 (***)	[-0.61; -0.28]
INFN	0.7 (SD=0.40, CV=57.3%)	1.1 (SD=0.62, CV=55.7%)	-0.4	-4.25 (97.7) r≈0.37	=0.001 (**)	[-0.59; -0.22]
AUX	0.44 (SD=0.2482, CV=56.9%)	0.64 (SD=0.2877, CV=45.2%)	-0.20	-4.10 (114.0) r≈0.36	<0.001 (***)	[-0.30; -0.10]
DF_ADJ_N_PRN	0.4 (SD=0.1577, CV=39.5%)	0.6 (SD=0.2761, CV=46.0%)	-0.2	-4.89 (89.9) r≈0.46	<0.001 (***)	[-0.28; -0.12]
PRN_CONJ_W	0.27 (SD=0.145, CV=54.1%)	0.46 (SD=0.2235, CV=48.4%)	-0.19	-5.63 (97.4) r≈0.50	<0.001 (***)	[-0.26; -0.13]
MDL_VRB	0.27 (SD=0.1611, CV=60.3%)	0.48 (SD=0.3345, CV=69.4%)	-0.21	-4.45 (81.4) r≈0.44	<0.001 (***)	[-0.31; -0.12]
COPULA	0.15 (SD=0.0987, CV=67.1%)	0.22 (SD=0.1632, CV=71.1%)	-0.08	3.34 (93.2) r≈0.33	<0.01 (**)	[-0.13; -0.03]
MDL_WRD	0.15 (SD=0.1294, CV=88.9%)	0.26 (SD=0.2347, CV=90.1%)	-0.11	-3.31 (88.0) r≈0.33	<0.01 (**)	[-0.18; -0.05]
PER_PRN	0.14 (SD=0.101, CV=70.6%)	0.29 (SD=0.231, CV=80.0%)	-0.15	-4.45 (77.4) r≈0.45	<0.001 (***)	[-0.21; -0.08]
INDF_ADJ_PRN	0.05 (SD=0.0480, CV=100.2%)	0.09 (SD=0.0728, CV=84.2%)	-0.04	-3.43 (98.3) r≈0.33	<0.001 (***)	[-0.06; -0.02]

RFLX_PRN	0.04 (SD=0.0423, CV=117.5%)	0.06 (SD=0.0616, CV=109.8%)	-0.02	-2.08 (100.9) r≈0.20	<0.05 (*)	[-0.04; -0.001]
SBJ_PRCL	0.01 (SD=0.022, CV=199%)	0.03 (SD=0.037, CV=116.2%)	-0.02	-3.74 (91.9) r≈0.36	<0.001 (***)	[-0.03; -0.01]
Частеречные биграммы						
NOUN NOUN	9.51 (SD=1.279, CV=13.4%)	8.94 (SD=1.597, CV=17.9%)	+0.57	+2.17 (110.1) r≈0.20	<0.05 (*)	[0.05; 1.09]
NOUN ADJF	2.47 (SD=0.709, CV=28.7%)	2.15 (SD=0.613, CV=28.5%)	+0.32	+2.65 (119.2) r≈0.24	<0.01 (**)	[0.08; 0.56]
COMMA SCONJ	0.79 (SD=0.307, CV=38.8%)	1.04 (SD=0.368, CV=35.5%)	-0.25	-4.00 (112.2) r≈0.35	<0.001 (***)	[-0.37; -0.13]
DMSTR_PRN NOUN	0.6 (SD=0.218, CV=36.2%)	0.79 (SD=0.302, CV=38.4%)	-0.19	-3.84 (104.0) r≈0.35	<0.001 (***)	[-0.28; -0.09]
PREP DMSTR_PRN	0.47 (SD=0.150, CV=32.0%)	0.64 (SD=0.247, CV=38.9%)	-0.17	-4.42 (93.6) r≈0.42	<0.001 (***)	[-0.24; -0.09]
COMMA PRN_CONJ_W	0.14 (SD=0.095, CV=68.6%)	0.30 (SD=0.187, CV=62.1%)	-0.16	-5.99 (83.8) r≈0.55	<0.001 (***)	[-0.22 -0.11]
AUX PRTS	0.37 (SD=0.219, CV=60.2%)	0.51 (SD=0.247, CV=48.1%)	-0.15	-3.49 (115.3) r≈0.31	<0.001 (***)	[-0.23; -0.06]
MDL_VRB INFN	0.19 (SD=0.122, CV=65.4%)	0.32 (SD=0.262, CV=81.4%)	-0.13	-3.59 (80.0) r≈0.37	<0.001 (***)	[-0.21; -0.06]
PER_PRN VERB	0.05 (SD=0.054, CV=115.2%)	0.15 (SD=0.169, CV=111.7%)	-0.1	-4.51 (68.4) r≈0.48	<0.001 (***)	[-0.15; -0.06]
COMMA PRTF	0.68 (SD=0.262, CV=38.2%)	0.81 (SD=0.268, CV=33.0%)	-0.13	-2.61 (118.5) r≈0.23	<0.05 (*)	[-0.22; -0.03]
Частеречные триграммы						
NOUN COMMA NOUN	1.01 (SD=0.474, CV=47.1%)	0.62 (SD=0.253, CV=41.0%)	+0.39	+5.73 (96.5) r≈0.50	<0.001 (***)	[0.26; 0.56]
NOUN ADJF NOUN	2.13 (SD=0.635, CV=29.7%)	1.86 (SD=0.506, CV=27.3%)	+0.27	+2.69 (117.2) r≈0.24	<0.01 (**)	[0.07; 0.48]
ADJF NOUN NOUN	1.91 (SD=0.406, CV=21.3%)	1.72 (SD=0.419, CV=24.5%)	+0.19	+2.54 (118.3) r≈0.23	<0.05 (*)	[0.04; 0.34]
NOUN NOUN ADJF	0.56 (SD=0.271, CV=48.5%)	0.42 (SD=0.189, CV=45.0%)	+0.14	+3.30 (111.5) r≈0.30	<0.01 (**)	[0.06; 0.22]
NOUN COMMA PRN_CONJ_W	0.11 (SD=0.082, CV=72.8%)	0.26 (SD=0.166, CV=63.7%)	-0.15	-6.16 (82.6) r≈0.56	<0.001 (***)	[-0.2; -0.1]
COMMA PRTF PREP	0.26 (SD=0.140, CV=54.2%)	0.38 (SD=0.209, CV=55.0%)	-0.12	-3.73 (99.4) r≈0.35	<0.001 (***)	[-0.19; -0.06]

NOUN COMMA SCONJ	0.28 (SD=0.123, CV=43.9%)	0.37 (SD=0.167, CV=44.7%)	-0.09	-3.49 (105.2) r≈0.32	<0.001 (***)	[-0.15; -0.04]
PREP DMSTR_PRN NOUN	0.22 (SD=0.093, CV=43.4%)	0.3 (SD=0.133, CV=44.6%)	-0.08	-3.96 (102.3) r≈0.37	<0.001 (***)	[-0.13; -0.04]
PREP DMSTR_PRN COMMA	0.11 (SD=0.073, CV=69.2%)	0.19 (SD=0.105, CV=56.5%)	-0.08	-4.84 (102.0) r≈0.43	<0.001 (***)	[-0.11; -0.05]
COMMA PRN_CONJ_W VERB	0.07 (SD=0.076, CV=115.4%)	0.15 (SD=0.096, CV=66.6%)	-0.08	-4.97 (109.5) r≈0.43	<0.001 (***)	[-0.11; -0.05]
COMMA SCONJ NOUN	0.25 (SD=0.152, CV=61.4%)	0.32 (SD=0.158, CV=49.8%)	-0.07	-2.54 (117.9) r≈0.23	<0.05 (*)	[-0.13; -0.02]
NOUN MDL_VRB INFN	0.09 (SD=0.078, CV=84.1%)	0.16 (SD=0.140, CV=87.2%)	-0.07	-3.26 (88.6) r≈0.33	<0.01 (**)	[-0.11; -0.03]
NOUN COMMA PRTF	0.61 (SD=0.243, CV=40.0%)	0.74 (SD=0.256, CV=34.7%)	-0.13	-2.89 (117.7) r≈0.26	<0.01 (**)	[-0.22; -0.04]
<b>МЕД-МII</b>						
<b>N-грамма</b>	<b>Non-translated (weighted mean)</b>	<b>Translated (weighted mean)</b>	<b>Δ (NT-T)</b>	<b>Welch t (df)</b>	<b>p</b>	<b>95% CI Δ</b>
<b>Частеречные униграммы</b>						
NOUN	34.83 (SD=1.95, CV=5.6%)	31.89 (SD=1.55, CV=4.9%)	+2.94	+8.77 (106.1) r≈0.65	<0.001 (***)	[2.28; 3.60]
ADJF	8.92 (SD=1.62, CV=18.2%)	7.48 (SD=1.65, CV=22.0%)	+1.44	4.54 (96.1) r≈0.42	<0.001 (***)	[0.81; 2.07]
PRTF	1.78 (SD=0.498, CV=28.0%)	2.1 (SD=0.564, CV=26.9%)	-0.32	-3.06 (89.5) r≈0.31	<0.01 (**)	[-0.53; -0.11]
SCONJ	1.09 (SD=0.406, CV=37.1%)	1.46 (SD=0.386, CV=26.4%)	-0.37	-4.81 (99.6) r≈0.43	<0.001 (***)	[-0.52; -0.22]
PRTS	1.01 (SD=0.336, CV=33.3%)	1.14 (SD=0.316, CV=27.8%)	-0.13	-2.04 (100.3) r≈0.20	<0.05 (*)	[-0.254; -0.004]
INFN	0.7 (SD=0.401, CV=57.3%)	0.55 (SD=0.293, CV=53.0%)	+0.15	+2.21 (107.1) r≈0.21	<0.05 (*)	[0.02; 0.28]
AUX	0.44 (SD=0.248, CV=56.9%)	0.97 (SD=0.267, CV=27.6%)	-0.53	-10.59 (92.7) r≈0.74	<0.001 (***)	[-0.63; -0.43]
DF_ADJ_N_PRN	0.4 (SD=0.158, CV=39.5%)	0.55 (SD=0.225, CV=40.7%)	-0.15	-3.97 (75.9) r≈0.41	<0.001 (***)	[-0.23; -0.08]
PRN_CONJ_W	0.27 (SD=0.145, CV=54.1%)	0.53 (SD=0.28, CV=52.4%)	-0.26	-5.84 (62.7) r≈0.59	<0.001 (***)	[-0.35; -0.17]
COPULA	0.15 (SD=0.099, CV=67.1%)	0.33 (SD=0.181, CV=55.7%)	-0.18	-6.03 (64.3) r≈0.60	<0.001 (***)	[-0.24; -0.12]

MDL_WRD	0.15 (SD=0.129, CV=88.9%)	0.06 (SD=0.05, CV=91.3%)	+0.09	5.05 (85.69) r≈0.48	<0.001 (***)	[0.05; 0.13]
PER_PRN	0.14 (SD=0.101, CV=70.6%)	0.43 (SD=0.239, CV=55.1%)	-0.29	-7.76 (56.6) r≈0.72	<0.001 (***)	[-0.37; -0.22]
<b>Частеречные биграммы</b>						
AUX PRTS	0.37 (SD=0.219, CV=60.2%)	0.79 (SD=0.242, CV=30.8%)	-0.42	-9.31 (91.2) r≈0.70	<0.001 (***)	[-0.51; -0.33]
PER_PRN VERB	0.05 (SD=0.054, CV=115.3%)	0.31 (SD=0.204, CV=66.0%)	-0.26	-8.50 (49.5) r≈0.77	<0.001 (***)	[-0.32; -0.20]
COMMA PRTF	0.68 (SD=0.262, CV=38.2%)	0.94 (SD=0.337, CV=36.0%)	-0.25	-4.24 (81.5) r≈0.42	<0.001 (***)	[-0.37; -0.13]
COMMA PRN_CONJ_W	0.14 (SD=0.095, CV=68.6%)	0.38 (SD=0.185, CV=48.8%)	-0.24	-8.09 (62.1) r≈0.71	<0.001 (***)	[-0.30; -0.18]
COMMA SCONJ	0.7 (SD=0.307, CV=38.8%)	0.96 (SD=0.297, CV=31.1%)	-0.17	-2.85 (98.6) r≈0.28	<0.01 (**)	[-0.28; -0.05]
NOUN NOUN	9.51 (SD=1.279, CV=13.4%)	7.81 (SD=1.089, CV=14.0%)	+1.71	+7.52 (104.4) r≈0.59	<0.001 (***)	[1.26; 2.16]
ADJF NOUN	7.38 (SD=1.387, CV=18.8%)	6.14 (SD=1.280, CV=20.8%)	+1.24	+4.82 (101.1) r≈0.43	<0.001 (***)	[0.73; 1.75]
NOUN ADJF	2.47 (SD=0.709, CV=28.7%)	1.40 (SD=0.434, CV=30.9%)	+1.07	+9.71 (104.4) r≈0.69	<0.001 (***)	[0.85; 1.29]
<b>Частеречные триграммы</b>						
NOUN ADJF NOUN	2.13 (SD=0.635, CV=29.7%)	1.20 (SD=0.378, CV=31.4%)	+0.93	+9.56 (103.5) r≈0.69	<0.001 (***)	[0.74; 1.12]
NOUN NOUN NOUN	2.32 (SD=0.597, CV=25.8%)	1.6 (SD=0.470, CV=29.2%)	+0.71	+6.94 (106.4) r≈0.56	<0.001 (***)	[0.51; 0.91]
ADJF NOUN NOUN	1.91 (SD=0.406, CV=21.3%)	1.51 (SD=0.377, CV=25.1%)	+0.40	+5.29 (100.8) r≈0.47	<0.001 (***)	[0.25; 0.55]
NOUN NOUN ADJF	0.56 (SD=0.271, CV=48.5%)	0.24 (SD=0.118, CV=49.9%)	+0.32	+8.42 (90.4) r≈0.66	<0.001 (***)	[0.25; 0.40]
NOUN COMMA NOUN	1.01 (SD=0.474, CV=47.1%)	0.71 (SD=0.289, CV=40.7%)	+0.3	+4.04 (104.4) r≈0.37	<0.001 (***)	[0.15; 0.44]
VERB COMMA SCONJ	0.12 (SD=0.085, CV=70.6%)	0.2 (SD=0.107, CV=53.9%)	-0.08	-4.12 (82.8) r≈0.41	<0.001 (***)	[-0.12; -0.04]
S_START PER_PRN VERB	0.012 (SD=0.018, CV=149%)	0.104 (SD=0.124, CV=119%)	-0.092	-5.00 (46.3) r≈0.59	<0.001 (***)	[-0.13; -0.06]
PER_PRN VERB NOUN	0.01 (SD=0.018, CV=150.2%)	0.10 (SD=0.070, CV=67.3%)	-0.09	-8.66 (49.5) r≈0.78	<0.001 (***)	[-0.11; -0.07]

COMMA PRN_CONJ_W VERB	0.07 (SD=0.076, CV=115.4%)	0.19 (SD=0.107, CV=57.1%)	-0.12	-6.57 (76.5) $r \approx 0.60$	<0.001 (***)	[-0.16; -0.09]
NOUN COMMA PRN_CONJ_W	0.11 (SD=0.082, CV=72.8%)	0.28 (SD=0.139, CV=49.4%)	-0.17	-7.35 (67.3) $r \approx 0.67$	<0.001 (***)	[-0.21; -0.12]
NOUN AUX PRTS	0.19 (SD=0.134, CV=69.8%)	0.37 (SD=0.151, CV=40.9%)	-0.18	-6.38 (89.6) $r \approx 0.56$	<0.001 (***)	[-0.23; -0.12]
COMMA PRTF PREP	0.26 (SD=0.140, CV=54.2%)	0.45 (SD=0.208, CV=46.2%)	-0.19	-5.43 (73.6) $r \approx 0.53$	<0.001 (***)	[-0.26; -0.12]
NOUN COMMA PRTF	0.61 (SD=0.243, CV=40.0%)	0.84 (SD=0.312, CV=37.1%)	-0.23	-4.22 (81.7) $r \approx 0.42$	<0.001 (***)	[-0.34; -0.12]
AUX PRTS PREP	0.11 (SD=0.085, CV=77.1%)	0.43 (SD=0.145, CV=33.8%)	-0.32	-13.38 (67.0) $r \approx 0.85$	<0.001 (***)	[-0.37; -0.27]
NOUN COMMA SCONJ	0.28 (SD=0.123, CV=43.9%)	0.35 (SD=0.116, CV=33.0%)	-0.07	-3.01 (100.3) $r \approx 0.29$	<0.01 (**)	[-0.12; -0.02]

### Используемые в таблице обозначения:

- **weighted mean** – взвешенное среднее значение анализируемого показателя;
- **SD** – стандартное отклонение (измеряется в тех же единицах, что и среднее значение);
- **CV** – коэффициент вариации;
- **NT** – непереводные тексты на русском языке (non-translated texts);
- **T** – переводные тексты на русском языке (translated texts);
- $\Delta$  (**NT-T**) – разность взвешенных средних непереводными текстами и переводами;
- **Welch t** и **df** - значение t-статистики и количество степеней свободы;
- **p-value** – достигнутый уровень значимости (n. s. – результат статистически незначим (т.е.  $p \geq 0.05$ ); \* –  $p < 0.05$ ; \*\* –  $p < 0.01$ ; \*\*\* –  $p < 0.001$ );
- **95% CI  $\Delta$**  – 95%-ый доверительный интервал для разности средних.

## ПРИЛОЖЕНИЕ 3

Результаты анализа индикаторов, отражающих особенности научного  
стиля русского языка, по корпусам

СОЦ-РП						
Индикатор	Non-translated (median) [Q <sub>1</sub> -Q <sub>3</sub> ], IQR, MAD	Translated (median) [Q <sub>1</sub> -Q <sub>3</sub> ], IQR, MAD	Mann-Whitney U (Z)	Effect size: r, A <sub>12</sub>	p	HL 95% CI
Маркеры авторского начала в тексте						
Высокая степень уверенности	0.26 [0.18-0.36], IQR=0.180, MAD=0.09	0.38 [0.27-0.47], IQR=0.205, MAD=0.1	1042.5 (-3.757)	r≈0.35, A <sub>12</sub> ≈0.30	<0.001 (***)	-0.12, [-0.18;-0.06]
Средняя степень уверенности	0.47 [0.38-0.59], IQR=0.212, MAD=0.109	0.66 [0.482-0.87], IQR=0.386, MAD=0.188	1014.5 (-3.908)	r=0.360, A <sub>12</sub> ≈0.29	<0.001 (***)	-0.19, [-0.28;-0.09]
Маркеры взаимодействия автора с читателем						
Акцентирование внимания	0.027 [0.000-0.072], IQR=0.072, MAD=0.027	0.019 [0.000-0.036], IQR=0.036, MAD=0.019	1303.5 (+2.4)	r=0.221; A <sub>12</sub> ≈0.63	<0.05 (*)	+0.013, [0.00; 0.03]
Индикатор	Non-translated (weighted mean)	Translated (weighted mean)	Δ (NT-T)	Welch t (df) r	p	95% CI Δ
Имена существительные						
Nouns to tokens ratio	36.12 (SD=2.81, CV=7.77%)	33.38 (SD=2.56, CV=7.67%)	+2.74	5.154 (99.72) r≈0.46	<0.001 (***)	[1.7; 3.8]
Singular number nouns to all nouns	71.30 (SD=5.86, CV=8.22%)	67.04 (SD=5.09, CV=7.59%)	+4.26	3.925 (99.39) r≈0.37	<0.001 (***)	[2.11; 6.42]
Abstract nouns to all nouns	42.38 (SD=5.93, CV=13.99%)	39.19 (SD=6.03, CV=15.40%)	+3.19	2.689 (98.70) r≈0.26	<0.01 (**)	[0.84; 5.55]
Имена прилагательные						
Adjectives to tokens ratio	15.75 (SD=1.73, CV=10.99%)	14.39 (SD=1.82, CV=12.66%)	+1.36	3.860 (97.94) r≈0.36	<0.001 (***)	[0.66; 2.06]
Short adjectives to all adjectives	4.66 (SD=1.33, CV=28.44%)	6.31 (SD=2.56, CV=40.58%)	-1.65	-4.017 (70.30) r≈0.43	<0.001 (***)	[-2.46; -0.83]
Глаголы						
Verbs to tokens ratio	11.86 (SD=1.28, CV=10.8%)	12.88 (SD=1.25, CV=9.7%)	-1.02	-4.04 (99.3) r≈0.38	<0.001 (***)	[-1.51; -0.52]
Passive verbs ratio	20.54 (SD=5.69, CV=27.7%)	17.74 (SD=4.21, CV=23.7%)	+2.79	2.83 (95.49) r≈0.28	<0.01 (**)	[0.83; 4.75]
Participles to all verbs ratio	22.20 (SD=6.33, CV=28.5%)	17.99 (SD=3.25, CV=18.1%)	+4.21	4.27 (79.0) r≈0.43	<0.001 (***)	[2.25; 6.18]
Future verbs to all finite verbs	2.92	4.30	-1.39	-3.32 (76.3)	<0.01 (**)	

	(SD=1.50, CV=51.5%)	(SD=2.53, CV=58.8%)		$r \approx 0.36$		[-2.22; -0.55]
<b>Местоимения</b>						
Personal / possessive pronouns	2.38 (SD=0.78, CV=32.6%)	3.06 (SD=0.75, CV=24.4%)	-0.68	-4.51 (99.49) $r \approx 0.41$	<0.001 (***)	[-0.98; -0.38]
Relative pronouns	2.04 (SD=0.51, CV=25.0%)	2.62 (SD=0.52, CV=19.8%)	-0.59	-5.72 (98.7) $r \approx 0.50$	<0.001 (***)	[-0.79; -0.38]
<b>МЕД-ПП</b>						
<b>Индикатор</b>	<b>Non-translated (median) [Q<sub>1</sub>-Q<sub>3</sub>], IQR, MAD</b>	<b>Translated (median) [Q<sub>1</sub>-Q<sub>3</sub>], IQR, MAD</b>	<b>Mann- Whitne y U (Z)</b>	<b>Effect size: r, A<sub>12</sub></b>	<b>p</b>	<b>HL 95% CI</b>
<b>Маркеры авторского начала в тексте</b>						
Высокая степень уверенности	0.15 [0.06-0.23], IQR=0.173, MAD=0.086	0.26 [0.13-0.41], IQR=0.285, MAD=0.144	1682 (-3.314)	$r=0.279$ ; A <sub>12</sub> ≈0.34	$p<0.001$ (***)	-0.10, [-0.16; -0.04]
Средняя степень уверенности	0.29 [0.169-0.475], IQR=0.31, MAD=0.144	0.47 [0.3-0.8], IQR=0.51, MAD=0.248	1501 (-4.058)	$r=0.342$ ; A <sub>12</sub> ≈0.3	$p<0.001$ (***)	-0.19, [-0.29; -0.09]
<b>Индикатор</b>	<b>Non-translated (weighted mean)</b>	<b>Translated (weighted mean)</b>	<b>Δ (NT-T)</b>	<b>Welch t (df) r</b>	<b>p</b>	<b>95% CI Δ</b>
<b>Имена существительные</b>						
Nouns to tokens ratio	41.27 (SD=2.92, CV=7.07%)	38.90 (SD=2.62, CV=6.74%)	+2.37	4.708 (119.60) $r \approx 0.40$	<0.001 (***)	[1.37; 3.36]
Abstract nouns to all nouns	36.80 (SD=4.97, CV=13.51%)	39.16 (SD=4.99, CV=12.74%)	-2.36	-2.607 (118.91) $r \approx 0.23$	0.01 (*)	[-4.14; -0.57]
<b>Имена прилагательные</b>						
Adjectives to tokens ratio	15.63 (SD=2.06, CV=13.19%)	14.62 (SD=2.01, CV=13.72%)	+1.01	2.735 (119.4) $r \approx 0.24$	<0.01 (**)	[0.28; 1.74]
Short adjectives to all adjectives	2.96 (SD=1.47, CV=49.74%)	4.36 (SD=2.60, CV=59.76%)	-1.39	-3.590 (89.44) $r \approx 0.36$	<0.001 (***)	[-2.16; -0.62]
<b>Глаголы</b>						
Verbs to tokens ratio	9.62 (SD=1.14, CV=11.9%)	10.60 (SD=1.53, CV=14.5%)	-0.98	-3.97 (105.8) $r \approx 0.36$	<0.001 (***)	[-1.47; -0.49]
Future verbs to all finite verbs	0.83 (SD=1.01, CV=122.01%)	1.42 (SD=1.57, CV=110.36%)	-0.60	-2.467 (96.82) $r \approx 0.24$	<0.05 (*)	[-1.07; -0.12]
<b>Местоимения</b>						
Personal / possessive pronouns	0.82 (SD=0.4, CV=48.1%)	1.02 (SD=0.6, CV=57.1%)	-0.19	-2.11 (100.63) $r \approx 0.21$	<0.05 (*)	[-0.37; -0.01]
Demonstrative pronouns	1.76 (SD=0.53, CV=30.0%)	2.42 (SD=0.8, CV=32.8%)	-0.66	-5.38 (98.68) $r \approx 0.48$	<0.001 (***)	[-0.91; -0.42]

Relative pronouns	1.24 (SD=0.42, CV=33.8%)	1.52 (SD=0.49, CV=32.1%)	-0.29	-3.46 (113.66) $r \approx 0.31$	<0.001 (***)	[-0.45; -0.12]
Defining pronouns	0.5 (SD=0.22, CV=43.2%)	0.7 (SD=0.32, CV=46.3%)	-0.2	-3.97 (99.1) $r \approx 0.37$	<0.001 (***)	[-0.3; -0.1]
Indefinite pronouns	0.13 (SD=0.09, CV=72.1%)	0.21 (SD=0.13, CV=65.4%)	-0.075	-3.56 (101.7) $r \approx 0.34$	<0.001 (***)	[-0.12; -0.03]
<b>МЕД-МП</b>						
<b>Индикатор</b>	<b>Non-translated (median) [Q<sub>1</sub>-Q<sub>3</sub>], IQR, MAD</b>	<b>Translated (median) [Q<sub>1</sub>-Q<sub>3</sub>], IQR, MAD</b>	<b>Mann-Whitney U (Z)</b>	<b>Effect size: r, A<sub>12</sub></b>	<b>p</b>	<b>HL 95% CI</b>
<b>Маркеры взаимодействия автора с читателем</b>						
Апелляция к фоновым знаниям	0.042 [0-0.06], IQR=0.06, MAD=0.042	0 [0-0.024], IQR=0.024, MAD=0	1004 (4.231)	$r=0.386$ ; A <sub>12</sub> $\approx 0.7$	<0.001 (***)	+0.028, [0.000; 0.043]
<b>Индикатор</b>	<b>Non-translated (weighted mean)</b>	<b>Translated (weighted mean)</b>	<b><math>\Delta</math> (NT-T)</b>	<b>Welch t (df) r</b>	<b>p</b>	<b>95% CI <math>\Delta</math></b>
<b>Имена существительные</b>						
Nouns to tokens ratio	41.27 (SD=2.92, CV=7.07%)	37.41 (SD=2.54, CV=6.79%)	+3.86	7.358 (103.58) $r \approx 0.59$	<0.001 (***)	[2.82; 4.9]
Singular number nouns to all nouns	72.49 (SD=4.09, CV=5.65%)	66.82 (SD=4.57, CV=6.83%)	+5.672	6.688 (90.44) $r \approx 0.58$	<0.001 (***)	[3.99; 7.36]
<b>Имена прилагательные</b>						
Adjectives to tokens ratio	15.63 (SD=2.06, CV=13.19%)	13.39 (SD=2.19, CV=16.35%)	+2.242	5.409 (93.47) $r \approx 0.49$	<0.001 (***)	[1.42; 3.06]
Superlative adjectives to all adjectives	1.11 (SD=0.63, CV=56.28%)	1.53 (SD=1.11, CV=72.62%)	-0.42	-2.295 (65.52) $r \approx 0.27$	<0.05 (*)	[-0.78; -0.054]
<b>Глаголы</b>						
Verbs to tokens ratio	9.62 (SD=1.14, CV=11.9%)	10.13 (SD=1.17, CV=11.6%)	-0.51	-2.28 (95.3) $r \approx 0.23$	<0.05 (*)	[-0.96; -0.07]
Passive verbs ratio	35.27 (SD=7.57, CV=21.5%)	39.39 (SD=7.73, CV=19.6%)	-4.12	-2.78 (95.81) $r \approx 0.28$	<0.01 (**)	[-7.07; -1.18]
Imperfective verbs to all verbs ratio	62.74 (SD=7.93, CV=12.64%)	54.01 (SD=5.54, CV=10.26%)	+8.73	6.77 (107.0) $r \approx 0.55$	<0.001 (***)	[6.17; 11.29]
Imperfective finite verbs to all finite verbs	83.37 (SD=7.31, CV=8.76%)	74.99 (SD=7.54, CV=10.06%)	+8.38	5.80 (95.1) $r \approx 0.51$	<0.001 (***)	[5.51; 11.24]
Present verbs to all finite verbs	57.02 (SD=20.13, CV=35.30%)	38.27 (SD=11.14, CV=29.10%)	+18.75	6.211 (100.85) $r \approx 0.53$	<0.001 (***)	[12.76; 24.74]
Past verbs to all finite verbs	41.84 (SD=20.29, CV=48.50%)	60.21 (SD=11.39, CV=18.91%)	-18.37	-6.012 (101.40) $r \approx 0.51$	<0.001 (***)	[-24.43; -12.31]

Participles before nouns	24.48 (SD=6.13, CV=25.06%)	21.06 (SD=4.98, CV=23.67%)	+3.42	3.210 (105.71) $r \approx 0.30$	<0.001 (***)	[1.31; 5.54]
<b>Местоимения</b>						
Personal / possessive pronouns	0.82 (SD=0.397, CV=48.1%)	1.04 (SD=0.505, CV=48.4%)	-0.22	-2.45 (82.27) $r \approx 0.26$	<0.05 (*)	[-0.4; -0.04]
Defining pronouns	0.5 (SD=0.22, CV=43.3%)	0.69 (SD=0.27, CV=39.2%)	-0.19	-3.92 (83.4) $r \approx 0.4$	<0.001 (***)	[-0.3; -0.1]
Relative pronouns	1.24 (SD=0.42, CV=33.8%)	1.65 (SD=0.51, CV=30.8%)	-0.41	-4.49 (85.12) $r \approx 0.44$	<0.001 (***)	[-0.6; -0.23]
Indefinite pronouns	0.13 (SD=0.093, CV=72.2%)	0.17 (SD=0.122, CV=70.5%)	-0.045	-2.06 (80.49) $r \approx 0.22$	<0.05 (*)	[-0.09; -0.002]

### Используемые в таблице обозначения:

- **p-value** – достигнутый уровень значимости (n. s. – результат статистически незначим (т.е.  $p \geq 0.05$ ); \* –  $p < 0.05$ ; \*\* –  $p < 0.01$ ; \*\*\* –  $p < 0.001$ );
- **weighted mean** – взвешенное среднее значение анализируемого показателя;
- **SD** – стандартное отклонение (измеряется в тех же единицах, что и среднее значение);
- **CV** – коэффициент вариации;
- **NT** – непереводные тексты на русском языке (non-translated texts);
- **T** – переводные тексты на русском языке (translated texts);
- $\Delta$  (**NT-T**) – разность взвешенных средних непереводными текстами и переводами;
- **Welch t** и **df** - значение t-статистики и количество степеней свободы;
- **95% CI  $\Delta$**  – 95%-ый доверительный интервал для разности средних;
- **median** – медианное значение анализируемого показателя;
- **Q1-Q3** – границы интервала, в котором находится 50% значений показателя. Q1 соответствует 25-му перцентиллю, Q3 – 75-му, разница между ними отражает степень разброса данных;
- **IQR** (Interquartile Range) – разность между третьим и первым квартилем (Q3–Q1);
- **MAD** (Median Absolute Deviation) – медиана модулей отклонений всех наблюдений от общей медианы, отражающая типичную величину разброса значений относительно центра распределения;
- **Mann-Whitney U** и **z** – значения U-статистики и z-оценки;
- **r** – размер наблюдаемого эффекта (~0.1 маленький, ~0.3 средний, ~0.5 большой);
- **$A_{12}$**  – показатель вероятности того, что при случайном выборе по 1-му тексту из обоих подкорпусов значение показателя в непереводном тексте будет выше, чем соответствующее значение в переводном;
- **HL, 95% CI** – сдвиг Ходжеса–Лемана HL с 95%-доверительным интервалом.