

DOI: 10.33184/dokbsu-2026.1.17

Квантитативное исследование процессов симплификации в англоязычных переводах художественного текста

Е. А. Морозкина, А. Л. Смакова*, Г. В. Янгирова

Уфимский университет науки и технологий

Россия, Республика Башкортостан, 450076 Уфа, ул. Заки Валиди, 32.

**Email: smakova_alsu@mail.ru*

Цель статьи заключается в проведении корпусного сравнительно-сопоставительного исследования процессов симплификации, возникающих в англоязычных переводах художественного текста на лексическом уровне. В рамках исследования рассматриваются такие показатели, как уникальные лексемы, редуцированные лексемы, коэффициент лексического разнообразия и частотные списки лексем. Материалом для исследования послужил оригинал романа Ф. М. Достоевского «Преступление и наказание», а также два варианта его перевода на английский язык, выполненные К. Гарнетт и О. Рэди. Явление симплификации исследуется с помощью квантитативных методов корпусной лингвистики, а именно, методом поиска и анализа частотных списков лексем.

Ключевые слова: частотный список, симплификация, квантитативные методы, Достоевский, Преступление и наказание.

В современных науках о языке «наблюдается повышенный интерес ученых к применению прикладных методов компьютерной лингвистики» [4] к сопоставительному анализу оригинальных художественных текстов и их переводов. Впервые определение понятия «симплификация» было дано египетским ученым и пионером корпусной лингвистики М. Бейкер в статье 1993 г., в которой автор предложила концепцию переводческих универсалий – «особенностей, проявляющихся в переводных, но отсутствующих в непереводах текстах, и при этом не являющихся результатом взаимного влияния языковых систем» [6]. Согласно М. Бейкер, переводческие универсалии выступают как центральное направление корпусного переводоведения и включают упрощение (simplification), экспликацию (explicitation) и нормализацию (normalization). Симплификация (simplification, упрощение), по определению М. Бейкер, представляет собой процесс, при котором переводчики в своих работах сознательно представляют сложную информацию исходных текстов более простым способом. Нормализация (normalization) – это тенденция к приведению перевода в соответствие с типичными моделями и практиками целевого языка, даже на грани их преувеличения. Экспликация (explicitation, пояснение) – один из универсальных переводческих приемов, кото-

рый заключается в объяснении или прояснении скрытого смысла (выраженного имплицитно) в оригинальном тексте специальными лингвистическими средствами целевого языка [7]. В настоящем исследовании основное внимание уделено процессу симплификации.

Помимо М. Бейкер, проблемой симплификации в переводе занимались и другие отечественные и зарубежные ученые. Так, например, С. Лавьоза предположила, что упрощение в тексте перевода может наблюдаться как на лексическом, так и синтаксическом уровне, а также одной из первых сформулировала ряд гипотез о симплификации в своей статье 1998 г. Симплификация, по мнению С. Лавьозы, проявляется в том, что «переводной компонент сравнительного корпуса художественных текстов будет иметь более низкий показатель лексической плотности; средняя длина предложения переводного компонента будет меньше, чем в непереводном компоненте; переводной компонент сравнительного корпуса художественных текстов будет иметь более высокое процентное соотношение частотных слов к менее частотным; наиболее частотные слова, находящиеся во главе частотного списка (list head) в переводном компоненте будут также преобладать» [11].

В данной работе тенденция к упрощению в тексте перевода исследуется в аспекте четырех показателей: уникальные лексемы, редуцированные лексемы, коэффициент лексического разнообразия и частотность лексем. Таким образом, задача количественного анализа состоит в том, чтобы определить и пояснить разницу лексического наполнения текста оригинала и текстов перевода.

Частотные списки представляют собой наиболее эффективный и простой инструмент сравнения двух корпусов текстов, а также позволяют определить слова, наиболее характерные для каждого из них. В случае, если в одном корпусе наиболее частотные слова используются чаще, чем в другом, можно предполагать тенденцию к лексической симплификации и/или к редукации отдельных лексем.

Для исследования процессов симплификации был создан русско-английский параллельный корпус, включающий оригинал романа Ф.М. Достоевского «Преступление и наказание» [2], а также два варианта его перевода на английский язык, выполненные К. Гарнетт [8] и О. Рэди [9]. В данном исследовании проведено сравнение первых десяти высокочастотных единиц текста при помощи функции Word List автоматической программы AntConc. Как верно было замечено Е. А. Морозкиной, В. В. Воробьевым, А. Д. Корниловой, «современные технологии позволяют быстро и автоматизировано провести предварительную обработку художественного текста» [3]. Русскоязычный текст был предварительно лемматизирован, в случае англоязычных текстов был использован готовый список лемм и их словоформ. Для повышения надежности исследования, а также преодоления разрыва между языком перевода и текстом источника,

автоматическим путем при помощи функции Stoplist программы AntConc были исключены все стоп-слова (stop-words), которые, как правило, являются самыми частотными в любом тексте, но не несут смысловой нагрузки [13]. К стоп-словам относятся служебные части речи (предлоги, союзы, частицы), междометия, местоимения и любые другие слова, не несущие дополнительного смысла в тексте. Мы также исключили формы вспомогательных глаголов *be*, *have* и *do*, а также формы глаголов *быть* и *становиться* из русскоязычного частотного списка лексем. Таким образом, нами были рассмотрены первые десять наиболее частотных слов в оригинале и переводах исследуемого романа. Результаты приведены в *табл. 1*.

Таблица 1

Топ-10 наиболее частотных слов

	Оригинал романа «Преступление и наказание»		Перевод К. Гарнетт		Перевод О. Рэди	
	слово	частота %	слово	частота %	слово	частота %
1	мочь	0,99	go	1,28	can	1,15
2	знать	0,82	can	1,20	say	1,08
3	сказать	0,71	say	1,08	go	1,01
4	говорить	0,70	know	0,98	know	0,86
5	человек	0,65	come	0,95	see	0,80
6	рука	0,48	look	0,91	come	0,76
7	дело	0,47	see	0,90	look	0,70
8	хотеть	0,44	man	0,73	take	0,69
9	видеть	0,36	take	0,63	get	0,67
10	минута	0,36	make	0,58	time	0,59
доля %	5,97		9,26		8,31	

Полученные данные показывают, что процентное соотношение суммы количества употреблений словоформ ранга 1–10 выше в обоих вариантах перевода романа по сравнению с оригинальным текстом, что позволяет говорить о более явной тенденции к накоплению редуцированных элементов в текстах англоязычных переводов. В переводе К. Гарнетт первые десять наиболее частотных слов составляют 9,26% массива текста, что на 0,95% выше по сравнению с переводом О. Рэди. Может показаться, что разница незначительна, однако, как верно заметили в своем исследовании Д. Бузаджи и В. Ланчиков, даже самые небольшие отличия в переводе играют значительную роль, поскольку отражают разницу в подходах переводчиков к передаче одного и того же смысла [1].

Из приведенной таблицы очевидно, что лексемы *мочь*, *знать*, *сказать* и *видеть* являются наиболее частотными и встречаются как в тексте оригинала, так и в обоих текстах перевода. Общая частотность данных лексем в тексте оригинала составляет 2,88%, в переводе К. Гарнетт – 4,16%, в переводе О. Рэди – 3,89%. При рассмотрении отдельных лексем мы также можем говорить о том, что в текстах перевода данные лексемы являются гораздо более частотными по сравнению с текстом оригинала.

Для подтверждения результатов были расширены рамки исследования и также рассмотрены топ-20, 30, 40 и 50 частотных слов. Результаты приведены в *табл. 2*.

Таблица 2

Топ-20, 30, 40, 50 частотных слов

	Оригинал романа «Преступление и наказание»	Перевод К. Гарнетт	Перевод О. Рэди
Топ-20	9,22	13,84	12,72
Топ-30	11,97	17,58	16,06
Топ-40	14,25	20,70	18,86
Топ-50	16,20	23,43	21,41

Как показали результаты, при рассмотрении топ-20, 30, 40 и 50 частотных слов наблюдается та же тенденция к симплификации в текстах переводов во всех исследуемых случаях.

Коэффициент лексического разнообразия (КЛР) является одной из фундаментальных метрик корпусной лингвистики для оценки сложности текста и представляет собой «диапазон и вариативность словарного запаса, который говорящий реализует в тексте» [12]. Самой простой, а также самой распространенной метрикой вычисления КЛР является TTR (англ. type/token ratio), представляющая собой соотношение уникальных лексических единиц (types) и всех словоформ (tokens). Для подсчета степени КЛР была использована следующая формула: $k = (n/N) * 100$, где k – коэффициент лексического разнообразия, n – количество разных слов в корпусе (лемм), N – количество токенов (всех словоформ) [10]. Данный показатель представлен в *табл. 3*.

Таблица 3

Коэффициент лексического разнообразия (КЛР)

Перевод	N	n	TTR
Оригинал	87098	10522	12,08
К. Гарнетт (1914 г.)	96106	6518	6,78
О. Рэди (2014 г.)	99360	7141	7,19

Основываясь на данных *табл. 3*, можно заключить, что TTR текстов перевода (6,78% и 7,19%) меньше почти в два раза по сравнению с оригиналом (12,08%), что говорит о межъязыковой симплификации при переводе. Данное явление можно объяснить меньшим количеством словоформ, а также большим количеством уникальных лексем в оригинальном тексте. Мы также можем сделать вывод о прямой взаимосвязи между частотой использования лексемы и КЛР текста. Как видим, чем выше показатель КЛР, тем меньше будет частота использования одних и тех лексем в тексте. Можно предположить, что количество редуцированных лексем несколько больше в переводе К. Гарнетт по сравнению с переводом О. Рэди, имеющим большее количество уникальных лексем.

Приведем несколько примеров перевода на английский язык К. Гарнетт и О. Рэди лексемы *волнение* для подтверждения наших предположений (*табл. 3*).

Таблица 3

Перевод лексемы «волнение»

	Оригинал романа «Преступление и наказание»	Перевод К. Гарнетт	Перевод О. Рэди
1	Он был в удивительном волнении .	He was in violent agitation .	He was extraordinarily agitated .
2	Он был в сильном волнении .	He was much excited .	He was extremely worked up .
3	Через минуту вошла со свечой и Соня...вся в невыразимом волнении ...	A minute later Sonia too, came in with the candle ... inexpressibly agitated ...	A minute later Sonya came in with the candle ... in utter confusion ...
4	Разумихин стоял в задумчиво- сти и в волнении ...	Razumihin stood lost in thought and excitement ...	Razumikhin stood there, pensive and uneasy ...
5	Он был тоже как будто в неко- тором волнении .	He, too, appeared to be in some agitation .	He, too, seemed rather restless .
6	... бормотал он в волнении he muttered in agitation he muttered nervously ...

При этом «основным ориентиром» для точной интерпретации данной лексемы становится «контекст и особенности взаимодействия между участниками» коммуникации [5]. Как видим, одна и та же лексема *волнение* в версии перевода К. Гарнетт передана с помощью лексем *agitation* и *excitement*, а также их вариативными словоформами *agitated* и *excited*. В версии перевода О. Рэди для передачи одной и той же лексемы *волнение*, употребленной в тексте оригинала в разных падежах, во всех семи фрагментах текста

переводчиком использованы семь разных лексем, обладающих разными значениями, например, *agitated, restless, uneasy* и т.д. Можно говорить о том, что переводы О. Рэди имеют меньшее число повторений при переводе, а также большее количество уникальных лексем и, следовательно, более разнообразный словарный запас по сравнению с переводом К. Гарнетт. Таким образом, можно сделать вывод о том, что частотные списки и КЛР являются эффективными инструментами для выявления тенденции к упрощению на лексическом уровне, а также сравнения нескольких вариантов перевода текста. Данное исследование является подспорьем для дальнейшего развития темы и изучения проблем симплификации на синтаксическом уровне.

Литература

1. Бузаджи Д. М., Ланчиков В. К. Буквализм и языковое разнообразие. Об использовании одного метода корпусной лингвистики в переводоведении // Мосты. Журнал переводчиков. 2011. №4(32). С. 12–27.
2. Достоевский Ф. М. Преступление и наказание. М.: изд-во АСТ, 2017. 672 с.
3. Морозкина Е. А., Воробьев В. В., Корнилова А. Д. Корпусное исследование семантики глагола в структуре ремарок драматургического текста (на материале оригинала и англоязычного перевода пьесы А. П. Чехова «Вишневый сад») // Доклады Башкирского университета. 2024. Т. 9. №4. С. 8–15.
4. Морозкина Е. А., Хабибуллина Е. П. Статистическое исследование текста программными средствами (на материале оригинала и перевода новеллы Э. А. По «Падение дома Ашеров») // Доклады Башкирского университета. 2025. Т. 10. №2. С. 48–54.
5. Янгирова Г. В., Нафикова Л. К. Речевые акты неодобрения в романе У. Фолкнера «Шум и ярость» // Доклады Башкирского университета. 2025. №2. С. 55–63
6. Baker M. Corpus Linguistics and Translation Studies: Implications and Applications. In: Text and Technology: In Honour of John Sinclair. Amsterdam & Philadelphia: John Benjamins, 1993. P. 233–250.
7. Baker M. Corpus-based translation studies: The challenges that lie ahead. In: Terminology, LSP and Translation: Studies in Language Engineering, in Honor of Juan C. Sager / ed by H. Somers. Amsterdam, Philadelphia: J. Benjamins Pub. Co., 1996. P. 175–186.
8. Dostoyevsky F. Crime and punishment / transl. by C. Garnett. Mineola (N.Y.): Dover Publications, 2001. 430 p.
9. Dostoyevsky F. Crime and punishment / transl. by O. Ready. Limited ed. Penguin Books, 2024. 720 p.
10. Goncalves L. L., Goncalves L. B. Fractal power laws in literary English // Physica A. 2006. Vol. 360(2). P. 557–575.
11. Laviosa S. Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose // Meta: Translators' Journal. 1998. Vol. 43. No. 4. P. 557–570.
12. McCarthy P. M. Voc-D: A theoretical and empirical evaluation // Language Testing. 2007. Vol. 24. No. 4. P. 459–488.
13. Rajaraman A., Leskovec J., Ullman J. D. Data Mining. In: Mining of Massive Datasets. Cambridge University Press, 2010. P. 1–17.

Quantitative study of simplification processes in English translations of a literary text

E. A. Morozkina, A. L. Smakova *, G. V. Yangirova

Ufa University of Science and Technology

32 Zaki Validi st., 450076 Ufa, Republic of Bashkortostan, Russia.

**Email: smakova_alsu@mail.ru.*

The aim of this article is to conduct a corpus-based comparative study of simplification processes occurring in English translations of a literary text at the lexical level. The study examines such indicators as unique lexemes, reduplicated lexemes, the lexical diversity coefficient, and lexeme frequency lists. The novel by F. M. Dostoevsky «Crime and Punishment» as well as two versions of its English translation by K. Garnett and O. Ready, serve as the material for the study. The phenomenon of simplification is examined using quantitative methods of corpus linguistics, namely, the method of searching and analyzing lexeme frequency lists.

Keywords: frequency list, simplification, quantitative methods, Dostoevsky, Crime and punishment.