

УДК 81'42:811.112.2

<https://doi.org/10.25076/vpl.45.03>

А. И. Горожанов

Московский государственный лингвистический университет

И. А. Гусейнова

Московский государственный лингвистический университет

Д. В. Степанова

Минский государственный лингвистический университет

ИНСТРУМЕНТАРИЙ АВТОМАТИЗИРОВАННОГО АНАЛИЗА ПЕРЕВОДА ХУДОЖЕСТВЕННОГО ПРОИЗВЕДЕНИЯ

Рассматриваются особенности создания и апробации инструментария автоматизированного анализа оригинального художественного текста и его перевода для оценки адекватности передачи смыслов в рамках философской категории качества. Материалом исследования служит текст романа Ф. Кафки «Замок» на немецком языке и его перевод на русский язык. В работе применяется метод автоматизированного анализа текста, совмещающего в себе объективный программный и субъективный спекулятивный анализ материала, а также элементы корпусного подхода, в рамках которых используется библиотека Python обработки естественного языка spaCy, обладающая различными функциями для анализа размеченных текстовых массивов. Подробно рассматриваются этапы создания метода автоматизированного анализа качественной характеристики оригинального художественного текста и его экстраполяции на текст перевода с учетом формальных различий выбранных языков. Разработанный метод позволяет авторам предусмотреть нормализацию, токенизацию и частотный анализ текстового материала, визуализировать и интерпретировать полученные эмпирические данные, среди которых наиболее релевантными представляются коэффициенты отклонений. Авторы приходят к выводу о том, что использование современных программных инструментов обработки размеченных текстов может быть эффективным для выявления различий в восприятии качественной характеристики в тексте оригинала и тексте перевода.

Установленные различия могут быть следствием выбора неточных вариантов перевода. В качестве перспективного направления совершенствования инструментария предлагается работа над повышением точности выходных данных с помощью привлечения новейших программных решений, а также апробация инструментария на материале художественных произведений на других иностранных языках и их переводов в рамках философских категорий качества, пространства и времени.

Ключевые слова: инструментарий автоматизированного анализа, корпусный подход, оценка качества перевода, художественный перевод, интерпретация художественного произведения, роман Ф. Кафки «Замок», Python, библиотека spaCy

UDC 81'42:811.112.2

<https://doi.org/10.25076/vpl.45.03>

A. I. Gorozhanov

Moscow State Linguistic University

I. A. Guseynova

Moscow State Linguistic University

D. V. Stepanova

Minsk State Linguistic University

TOOLS FOR AUTOMATED ANALYSIS OF FICTION WORK TRANSLATION

The paper describes the features of the creation and approbation of tools for automated analysis of the original fiction text and its translation to assess the adequacy of the transfer of meanings for the philosophical category of quality. The material of the study is a text of the novel by F. Kafka "The Castle" in German and its translation into Russian. A method of automated text analysis is applied, which combines objective programmatic and subjective speculative analysis, as well as elements of the corpus approach, for which the spaCy natural language processing Python library is used, which has various functions for analyzing unmarked-up text arrays. The stages of creating a method for automated analysis of the qualitative characteristics of the original fiction text and its extrapolation to the translated text are considered in detail, taking into account the formal differences in the selected

languages. The developed method allows the authors to provide for the normalization, tokenization and frequency analysis of textual material, to visualize and interpret the obtained empirical data, among which the deviation coefficients seem to be the most relevant. The authors come to the conclusion that the use of modern software tools for processing unmarked-up texts can be effective in identifying differences in the perception of a qualitative characteristic in the original text and the translated text. The established differences may be the result of the choice of inaccurate translation options. As a promising direction for improving the tools, it is proposed to work on improving the accuracy of the output data by involving the latest software solutions, as well as approbation of the tools on the material of fiction works in other foreign languages and their translations within the philosophical categories of quality, space and time.

Keywords: tools for automated analysis, corpus approach, translation quality assessment, fiction work translation, fiction work interpretation, F. Kafka's novel "The Castle", Python, spaCy library

Введение

В настоящее время прикладная (компьютерная) лингвистика как никогда ранее имеет право претендовать на звание образца среди междисциплинарных областей знаний. Опираясь на математический аппарат и программирование, с одной стороны, и фундаментальные знания в области лингвистики, а также знание нескольких современных языков, с другой, прикладной лингвист призван решать сложные задачи, сопряженные с такими ключевыми, или сквозными, технологиями, как большие данные, искусственный интеллект, интернет вещей, машинное обучение и пр. (Auffarth, 2020; Carrier, 2017; Karsdorp, Kestemont, & Riddell, 2021, Малюга, 2009).

Однако не следует считать, что прикладная лингвистика служит сегодня исключительно утилитарным целям информационного общества. «Классическая» лингвистика также может использовать все достижения прикладной науки – даже в таких традиционно творческих областях, как интерпретация художественного текста и художественный перевод.

Целью настоящего исследования является создание и апробация инструментария автоматизированного компаративного анализа

оригинального художественного текста и его перевода, который бы помог оценить последний на предмет его соответствия оригиналу в рамках философской категории качества. Такая комплексная цель может быть достигнута только при решении ряда частных задач, а именно:

- создания метода автоматизированного анализа качественной характеристики оригинального художественного текста;
- экстраполяции этого метода на текст перевода, с учётом формальных различий исходного языка и языка перевода;
- визуализации полученных данных;
- интерпретации полученных данных с целью формулирования конкретных предложений для повышения уровня точности перевода.

Необходимость создания подобного инструментария объясняется очевидной потребностью в создании переводов высокого качества, поскольку качество перевода «является фундаментальной категорией теории перевода, ибо качество перевода фактически составляет ее истинный и единственный предмет» (Семко, 2005, с. 36). Традиционный метод анализа качества перевода путем сопоставления перевода и оригинала экспертом не признается сегодня всеми исследователями как достаточный; также в предметно-специальной литературе отмечается, что критерии оценки качества художественного перевода представляют особую сложность (Сдобников, 2015, с. 31, 59–62). Далее, исследователи отмечают, что «каждый [художественный] перевод – индивидуальное прочтение, индивидуальная интерпретация, то есть продукт субъективного познания оригинала конкретным переводчиком, и в этом смысле он субъективен» (Рябко, 2020, с. 137).

В целом, проблема качества перевода активно рассматривалась и продолжает рассматриваться в России и за рубежом (Waard & Nida, 1986; Швейцер, 1988; Комиссаров, 1990; Goriée, 1994; Fawcett, 2003; Гак 2003; Гарбовский, 2004; Welnitzova, Jakubickova, & Králik, 2021; Полякова, 2022; Томалин & Малюга, 2016). Не углубляясь в детали отдельных теоретических концепций – поскольку это потребовало бы отдельного обстоятельного исследования, – заключим, что отсутствие какого-либо универсального

объективного метода оценки качества перевода художественного произведения подкрепляет актуальность нашей работы.

Новизну работы подчеркивает тот факт, что в нем применяются новейшие программные средства: язык программирования Python 3 с использованием популярной библиотеки обработки естественного языка spaCy, что является на сегодня стандартом для исследований такого рода (Ayre et al., 2021; Jugran et al., 2021; Okhar'kin et al., 2021).

Материал и методы исследования

В качестве материала исследования был выбран оригинальный текст романа Ф. Кафки «Замок» и текст одного из его переводов на русский язык. Это обусловлено тем, что ранее нами была подготовлена определенная теоретико-практическая база в ходе прикладного исследования по автоматизированной интерпретации оригинального текста художественного произведения, материалом для которого и послужил указанный роман писателя-экспрессиониста (Горожанов, Гусейнова, 2021; Gorozhanov & Guseynova, 2020). Таким образом, настоящее исследование вписывается в более широкий контекст ряда научно-исследовательских работ, осуществление которых планируется продолжать и далее.

Методами настоящего исследования послужили автоматический (частотный) анализ текста, интерпретационный (мысленный) анализ текста, что в совокупности можно назвать автоматизированным анализом, т. е. анализом, совмещающим в себе объективный программный и субъективный спекулятивный анализ материала. Далее, в текущем исследовании широко применяется корпусный подход. Это находит свое отражение в том, что текст художественного произведения и его перевод рассматриваются как одноязычные сбалансированные неразмеченные корпуса в отдельности, и как параллельный двуязычный сбалансированный неразмеченный корпус в совокупности (Горожанов, Степанова, 2022, с. 205). В рамки корпусного подхода вписывается также и использование библиотеки spaCy, которая позволяет применять указанный подход для неразмеченных текстовых массивов, что, по нашему мнению, дает право говорить об использовании в исследовании spaCy-метода, что нисколько не противоречит общему определению

метода как совокупности приемов для достижения цели, поскольку, во-первых, технически любая программная библиотека содержит в себе различные функции (директивы, команды) для исполнения конкретных задач, а во-вторых, дифференцирует применение различных программных библиотек, которые представлены сегодня набором из минимум нескольких десятков вариантов, так что потенциально возможен не только spaCy-метод, но и NLTK-метод, Stanza-метод, Apache OpenNLP-метод и др.

Ход исследования и его результаты

Для решения первой задачи исследования были привлечены эмпирические данные, полученные в ходе предшествующей работы по определению пространственно-качественных характеристик оригинального текста романа Ф. Кафки «Замок». Это исследование, которое в значительной мере сформировало контекст настоящей работы, позволило выделить качественные характеристики локаций повествования как в целом для всего произведения, так и по каждой из его 20-ти глав.

Решение первой задачи (создания метода автоматизированного анализа качественной характеристики оригинального художественного текста) заключалось в следующем. Текст романа был нормализован, т. е. программно трансформирован в перечень отдельных словоформ, приведенных к написанию с маленькой буквы и разделенных пробелом. При этом все знаки препинания были удалены. Например, текст «Es war spät abends, als K. ankam. Das Dorf lag in tiefem Schnee.» превратился в текст «es war spät abends als k. ankam das dorf lag in tiefem schnee». Таким способом была получена 109191 словоформа. Далее из них программно были отобраны уникальные словоформы, что дало в результате 10939 единиц (токенов). Важно отметить, что речь здесь идет о формальном анализе и под уникальной словоформой подразумевается не начальная форма слова (например, инфинитив для глагола), а отличающаяся от других последовательность символов, поэтому «komm», «kam» и «kommt» считались тремя уникальными словоформами.

Затем уникальные единицы были отсортированы по частотности их употребления. Как и предполагалось, в начале списка оказались артикли, местоимения, частицы и предлоги:
und : 2927

die : 2443
sie : 2106
nicht : 1967
der : 1963
es : 1856
er : 1812
zu : 1674
ich : 1349
das : 1235
aber : 1211 и т. д.

После ручной правки в этом списке были оставлены только существительные, «лидерами» среди которых явились следующие:

k : 1205
frieda : 420
klamm : 366
barnabas : 246
wirtin : 204
schloß : 173
gehilfen : 170
herr : 150
amalia : 132
arbeit : 116
olga : 112
zimmer : 108
tür : 108
lehrer : 106
landvermesser : 102

Далее из полученного списка существительных были отобраны обозначения локаций романа, которые упоминаются более десяти раз, после чего фокус исследования был перенаправлен к словоформам (прилагательным и наречиям), наиболее обще характеризующим качества этих локаций. Был составлен следующий частотный список:

groß : 118
gut / besser : 81
jung : 53
alt : 52
still : 47

neu : 38
 offen : 34
 ruhig : 34
 ernst : 30
 richtig : 29
 frei : 22
 leise : 22
 voll : 21
 böse : 20
 gleichgültig : 20
 unrecht : 16
 wahr : 16
 klar : 16
 traurig : 14
 unglücklich : 14
 schwer : 13
 leer : 12
 unsicher : 11
 klein : 11
 dunkel : 11
 schlimm : 10
 kräftig : 10
 kurz : 10

Ключевым моментом является дальнейшее разделение этих словоформ на условно «положительные» и «отрицательные», которое было произведено вручную на основании экспертной интерпретации усредненных коннотаций этих лексических единиц в языке в целом (см. таблицу 1):

Таблица 1

+	–
groß : 118 gut / besser : 81 jung : 53 still : 47 neu : 38 offen : 34 ruhig : 34 ernst : 30	alt : 52 böse : 20 gleichgültig : 20 unrecht : 16 traurig : 14 unglücklich : 14 schwer : 13 leer : 12

richtig : 29 frei : 22 leise : 22 voll : 21 wahr : 16 klar : 16 kräftig : 10 Итого: 571	unsicher : 11 klein : 11 dunkel : 11 schlimm : 10 kurz : 10 Итого: 214
--	---

Отношение «положительных» и «отрицательных» качественных характеристик составило так называемый «коэффициент положительных характеристик» $N_{\text{pos/neg}}$. Здесь $N_{\text{pos/neg}} = 571 / 214 = 2,67$. Мы не будем приводить подробную интерпретацию этого показателя, поскольку в рамках настоящего исследования это не является релевантным. Тем не менее, то, что коэффициент положительных характеристик превышает значение «2», говорит скорее о позитивном восприятии пространства читателем, чем о негативном (Горожанов, Гусейнова, 2021, с. 31–32).

Затем текст романа был разделен на главы. Для каждой главы был рассчитан $N_{\text{pos/neg}}$. Например, для первой главы распределение «положительных» и «отрицательных» качественных характеристик принимает следующий вид (см. таблицу 2):

Таблица 2

+	–
jung : 14 gut / besser : 14 frei : 12 voll : 10 groß : 10 offen : 10 still : 9 wahr : 7 leise : 5 ruhig : 5 neu : 4 richtig : 2 klar : 2 ernst : 1 Итого: 105	alt : 24 klein : 9 leer : 3 kurz : 3 unsicher : 2 dunkel : 1 schwer : 1 traurig : 1 Итого: 44

При этом $N_{\text{pos/neg}1} = 2,386$.

Приведем подобные данные для всех глав романа (см. таблицу 3):

Таблица 3

Глава	$N_{\text{pos/neg}}$
1	2,386
2	1,302
3	1,25
4	1,244
5	1,698
6	1,684
7	1,172
8	1,655
9	2,533
10	1,55
11	2,2
12	1,444
13	1,957
14	1,815
15	1,693
16	1,2
17	2
18	1,622
19	2,026
20	1,741
Среднее арифметическое (μ)	1,7

Здесь важен показатель среднего арифметического μ из всех коэффициентов положительных характеристик. Он равен 1,7, и его не следует путать с $N_{\text{pos/neg}} = 2,67$ (для всего романа).

Заключительным шагом для решения первой задачи исследования является расчет еще одного коэффициента $N_{\text{dev}_{\text{pos/neg}}}$, который представляет собой частное от деления $N_{\text{pos/neg}}$ конкретной главы на среднее арифметическое μ . Коэффициент $N_{\text{dev}_{\text{pos/neg}}}$ призван показать, является ли глава более или менее положительной / отрицательной в качественном отношении на общем фоне восприятия качественных характеристик пространства романа. В результате получается следующая таблица (см. таблицу 4):

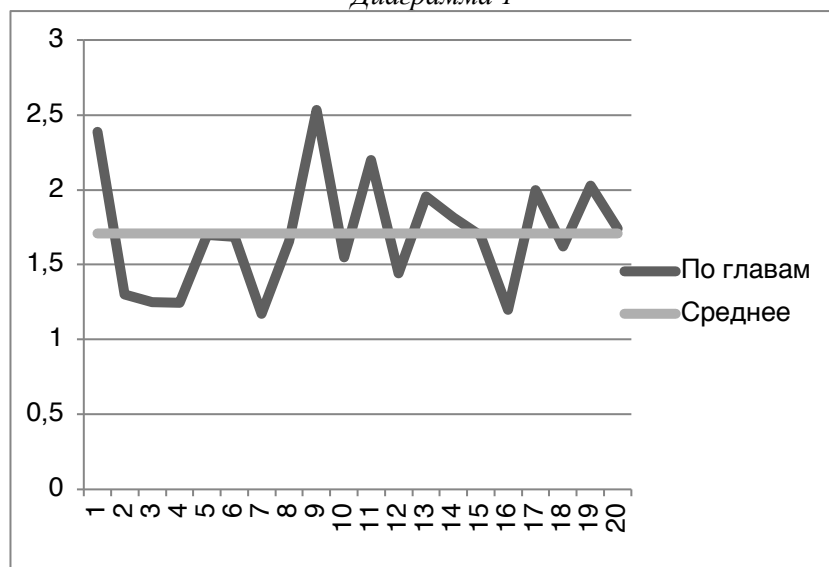
Таблица 4

Глава	$N_{dev_{pos/neg}}$
1	1,404
2	0,766
3	0,735
4	0,732
5	0,999
6	0,991
7	0,69
8	0,974
9	1,49
10	0,912
11	1,294
12	0,85
13	1,151
14	1,068
15	0,996
16	0,706
17	1,176
18	0,954
19	1,192
20	1,024

При коэффициенте $N_{dev_{pos/neg}} > 1$ восприятие представляется скорее позитивным, $N_{dev_{pos/neg}} < 1$ скорее негативным.

Полученные результаты удобно представить в виде диаграммы, на которой по оси X отложены главы романа, а по оси Y значения $N_{pos/neg}$ (см. диаграмму 1):

Диаграмма 1



Мысленная субъективная интерпретация содержания глав романа подкрепляет достоверность полученных данных. Так, первая и девятая главы показаны как наиболее позитивные. В самом деле, в первой главе главный герой только приходит в деревню, он чувствует себя внутренне свободным и уверен в том, что его работа будет востребована. В девятой, самой «позитивной» главе романа, К. находится на господском постоялом дворе, что уже говорит об «элитности» локации. Он проявляет твердость характера, уклоняется от допроса, покидает локацию по собственной воле. Самыми частотными качественными характеристиками здесь являются «*richtig*» и «*frei*».

Своеобразное «дно» образуют главы 7 и 16. В седьмой главе К. беседует с деревенским учителем. Тот предлагает ему работу при школе. К. уже живет с Фридой, у него образовалось некое подобие семьи, и хочется приемлемо обустроить совместный быт. Ради ночлега К. соглашается на предложение учителя, ведь устроиться землемером никак не получается. Главный герой унижается перед всеми: учителем, старостой, хозяйкой постоялого двора «У моста».

В главе 16 Фрида уходит от К. к его бывшему подмастерью. Несмотря на отсутствие экспрессии в репликах главного героя, глава производит крайне гнетущее впечатление. К. остается один.

Интересны и нейтральные показатели – например, в главах 5, 6 и 18. Пятая глава разворачивается в доме старосты деревни. В ней К. как будто «зависает» в неопределенности: он не может быть принят на работу в качестве землемера, но в то же время ему и не отказано полностью в занятости. Фактически нет повода воспринимать что-то хорошо или плохо. В следующей главе происходит беседа К. с хозяйкой постоянного двора «У моста». События не имеют непосредственного отношения к персонажу, глава является скорее «вставочной», что и объясняет нейтральность характеристик. Восемнадцатая глава протекает на господском постоялом дворе. К. видится с Фридой, которая снова работает за барной стойкой. После разговора с Фридой, в котором она снова упрекает К., тот случайно встречает секретаря Бюргеля. Их беседа касается бюрократических порядков в деревне и замке. Эти события мало влияют как на судьбу главного героя, так и на сюжет романа в целом.

Заклучим, что решение первой задачи исследования состоялось, поскольку был найден и апробирован метод качественной характеристики художественного текста, предусматривающий нормализацию, токенизацию и частотный анализ текстового материала.

Для решения второй задачи работы (экстраполяции созданного метода на текст перевода, с учётом формальных различий исходного языка и языка перевода) мы обратились к переводу романа «Замок» на русский язык. Сразу оговоримся, что здесь и далее, мы ни в коей мере не преследуем цель критики профессионально выполненного перевода романа. Русский текст используется нами с нейтральных позиций ученого и исключительно в качестве материала исследования. В этой связи мы не упоминаем имени переводчика. Любые наши оценки и мнения относительно улучшения качества конкретно взятого перевода несколько не умаляют его достоинств.

Нам предстояло разбить текст на главы, определить набор качественных характеристик (прилагательные и наречия), которые бы соответствовали качественным характеристикам оригинального

немецкоязычного текста и на основании этого рассчитать коэффициент положительных характеристик для каждой главы текста перевода.

Для каждой из «положительных» и «отрицательных» характеристик на немецком языке (см. таблицу 1) мы подобрали наиболее частотные переводческие соответствия из русского языка, опираясь на основные прямые значения лексических единиц. Параллельно была произведена нормализация, токенизация и частотный анализ русского текста, и если какая-то лексическая единица из предлагаемых в качестве соответствий для немецких лексических единиц не встречалась в этом частотном списке, то она не включалась в перечень. Например, в русском тексте не встретилась форма «ново», поэтому она не была зафиксирована в группе «neu». В результате получились следующие соответствия:

groß : большой, большой
gut / besser : хороший, хорошо
jung : молодой
still : спокойный, спокойно
neu : новый
offen : открытый, открыто
ruhig : спокойный, спокойно
ernst : серьёзный
richtig : правильный, правый, правильно
frei : свободный, свободно
leise : тихий, тихо
voll : полный
wahr : явный, настоящий, явно
klar : ясный, понятный, ясно
kräftig : сильный, крепкий, сильно, крепко

и

alt : старый
böse : злой, злобно
gleichgültig : равнодушный, безразличный, равнодушно, безразлично
unrecht : несправедливый, неправильный, несправедливо, неправильно
traurig : грустный, грустно
unglücklich : несчастный

schwer : тяжёлый, трудный, тяжело
 leer : пустой, пусто
 unsicher : –
 klein : маленький, малый, мелкий
 dunkel : тёмный
 schlimm : плохой, плохо
 kurz : короткий, коротко

Поскольку русские прилагательные проявляют более богатую парадигму, чем немецкие, а наречия отличаются по форме от прилагательных, простой поиск по корню слова не представлялся удачным решением. В этой связи мы были вынуждены прибегнуть к использованию библиотеки обработки естественного языка spaCy. Так, частотный список русских словоформ был сгенерирован с применением фильтра по прилагательным и наречиям. Поскольку spaCy официально имеет погрешность на уровне нескольких процентов (spaCy, 2022), в русский список помимо прилагательных и наречий попали некоторые причастия (например, «открыто»).

Для прилагательного «unsicher» прямые соответствия в русском тексте (например, «неуверенный», «неуверенно») не были обнаружены.

Общий коэффициент $N_{\text{pos/neg}}$ русского текста составил 781 / 405 = 1,93 (см. таблицу 5):

Таблица 5

+	–
хороший, хорошо : 142	маленький, малый, мелкий : 89
большой, большой : 126	тяжёлый, трудный, тяжело : 75
явный, настоящий, явно : 70	плохой, плохо : 55
молодой : 56	короткий, коротко : 31
ясный, понятный, ясно : 54	старый : 30
правильный, правый,	равнодушный, безразличный,
правильно : 51	равнодушно, безразлично : 23
новый : 46	несправедливый,
спокойный, спокойно : 46	неправильный, несправедливо,
сильный, крепкий, сильно,	неправильно : 23
крепко : 44	пустой, пусто : 21
полный : 42	несчастный : 17
тихий, тихо : 37	тёмный : 15

свободный, свободно : 29 открытый, открыто : 20 серьёзный : 18	грустный, грустно : 14 злой, злобно : 12
Итого: 781	Итого: 405

Далее были определены значения коэффициентов $N_{pos/neg}$ для каждой главы русского текста и среднее арифметическое μ (см. Таблицу 6):

Таблица 6

Глава	$N_{pos/neg}$
1	2,261
2	1,522
3	1,769
4	2,133
5	1,783
6	1,385
7	1,625
8	1,867
9	4,714
10	1,364
11	3,75
12	1,167
13	1,641
14	3,222
15	2,355
16	1,222
17	1,4
18	1,8
19	1,778
20	1,973
Среднее арифметическое (μ)	2,04

Были определены коэффициенты отклонения $N_{dev_{pos/neg}}$ (см. Таблицу 7):

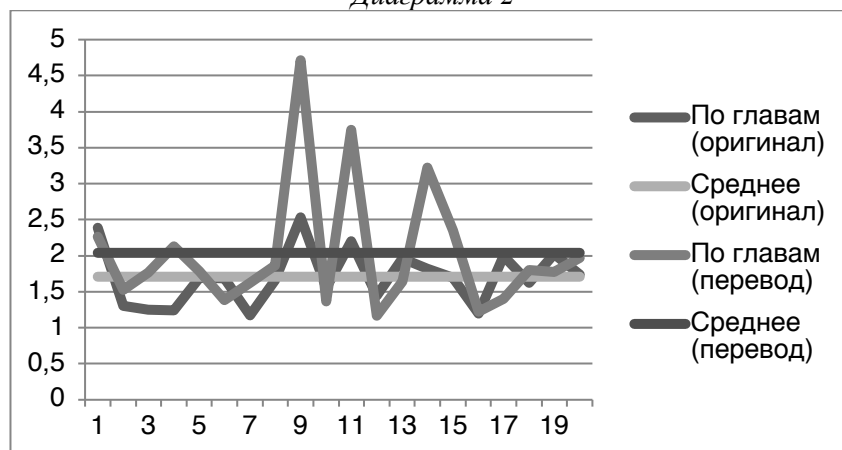
Таблица 7

Глава	$N_{dev_{pos/neg}}$
1	1,11
2	0,747
3	0,869

4	1,047
5	0,875
6	0,68
7	0,798
8	0,916
9	2,314
10	0,669
11	1,841
12	0,573
13	0,806
14	1,582
15	1,156
16	0,6
17	0,687
18	0,884
19	0,873
20	0,969

Решение третьей задачи исследования (визуализации полученных данных) уже было частично исполнено выше (см. диаграмму 1). Наложим данные таблицы 6 на диаграмму 1 (см. диаграмму 2):

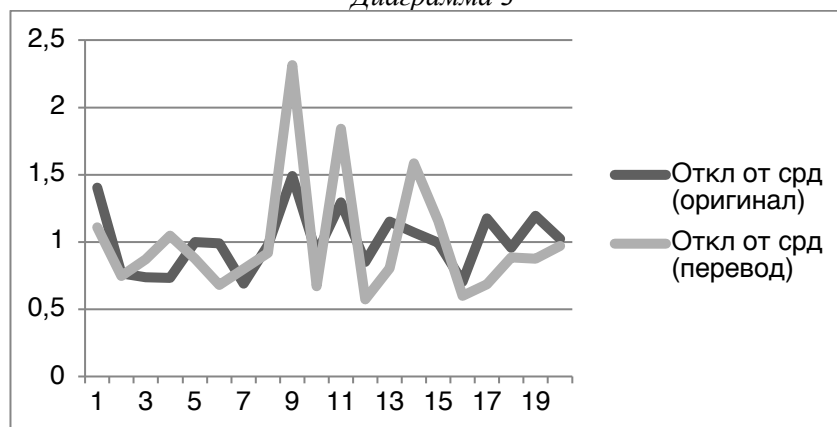
Диаграмма 2



На представленной диаграмме видно, что среднее арифметическое текста перевода лежит выше μ текста оригинала, а кривые коэффициентов положительных характеристик хотя и имеют приблизительно сходный характер, но практически не совпадают. Исключения представляют собой главы 1, 5 и 16.

Еще одним вариантом визуализации полученных данных является наложение друг на друга кривых коэффициентов отклонений $N_{dev_{pos/neg}}$ оригинала и перевода (см. Диаграмму 3):

Диаграмма 3



Мы подошли к решению последней задачи исследования (интерпретации полученных данных с целью формулирования конкретных предложений для повышения уровня точности перевода). Благодаря реализации трех предыдущих задач мы получили эмпирические данные, из которых наиболее релевантными представляются кривые коэффициентов отклонений $N_{dev_{pos/neg}}$ оригинала и перевода.

В целом, в рамках нашего исследования любые численные показатели в отдельности не являются достаточно информативными, но приобретают важность при сопоставлении друг с другом. Этот тезис ярко демонстрируют данные диаграмм 2 и 3. Коэффициенты $N_{pos/neg}$ для первых глав оригинального текста и его перевода имеют значения 2,386 и 2,261, соответственно, и практически совпадают. Однако принимая во внимание тот факт, что средние арифметические μ для оригинала и перевода

отличаются в 1,2 раза, факт совпадения восприятия качественных характеристик пространства при прочтении оригинального и переводного текстов подлежит сомнению. Другими словами, среднее арифметическое μ устанавливает для читателя некую нейтраль, от которой и производится эмоциональный отсчет «позитивное / негативное», поэтому для вычисления совпадений корректно будет использовать данные диаграммы 4, согласно которой минимальные различия демонстрируют главы 2, 8, 18 и 20. Пики различий приходятся на главы 9, 11, 14 и 17.

Проанализируем в качестве примера главу 9 оригинального и переводного текста, для того чтобы разобраться в том, что вызвало разницу в значениях коэффициента отклонений $N_{dev_{pos/neg}}$ и тем самым верифицировать с помощью «классического» мысленного анализа результаты предлагаемого экспериментального метода. Для этого рассмотрим подробно оба протокола анализа главы 9 (см. листинг 1 и 2):

Листинг 1

```
chapter_9.txt
richtig : 5
frei : 5
                                alt : 5

gut : 4
groß : 3
neu : 3
offen : 3
ruhig : 3
voll : 3
wahr : 3
                                kurz : 3

besser : 2
ernst : 2
                                gleichgültig : 2
                                leer : 2
                                klein : 2

jung : 1
still : 1
                                schwer : 1
-----
```

38

$N_{\text{pos/neg}} = 2,533$
 $\mu = 1,7$
 $N_{\text{dev}_{\text{pos/neg}}} = 1,490$

15

Листинг 2

глава09рус.txt

молодой : 13
хороший : 6
большой : 4
спокойно : 2
ясный : 2

пустой : 2
малый : 2

спокойный : 1
правильный : 1
полный : 1
явный : 1
настоящий : 1
понятный : 1

равнодушный : 1
трудный : 1
короткий : 1

33

7

$N_{\text{pos/neg}} = 4,714$
 $\mu = 2,037$
 $N_{\text{dev}_{\text{pos/neg}}} = 2,314$

При сравнении данных листингов 1 и 2 фиксируется высокая концентрация прилагательного «молодой» в русском тексте при однократном употреблении этой лексемы в тексте оригинала.

Единственное употребление прилагательного «jung» является характеристикой молодого господина Мома (Momus), который является деревенским секретарем Кламма:

*Und er riß sich los und ging ins Haus zurück, diesmal nicht an der Mauer entlang, sondern mitten durch den Schnee, traf im Flur den Wirt, der ihn stumm grüßte und auf die Tür des Ausschanks zeigte, folgte dem Wink, weil ihn fror und weil er Menschen sehen wollte, war aber sehr enttäuscht, als er dort an einem Tischchen, das wohl eigens hingestellt worden war, denn sonst begnügte man sich dort mit Fässern, den **jungen Herrn** sitzen und vor ihm – ein für K. bedrückender Anblick – die Wirtin aus dem Brückengasthaus stehen sah.*

В русском варианте это предложение имеет следующий вид:

*И он сорвался с места и пошел обратно в дом, но уже не прижимаясь к стенке, а прямо посередине двора, по снегу, встретил в коридоре хозяина, который молча поздоровался с ним и показал ему на вход в буфет; К. туда и пошел, потому что промерз и хотел видеть людей, но он был очень разочарован, когда увидел у специально поставленного столика (обычно все довольствовались бочонками) того самого **молодого человека**, а перед ним – это особенно расстроило К. – стояла хозяйка постоянного двора «У моста».*

Дальнейшие употребления прилагательного «молодой» относятся также к Мому: 11 раз «молодой человек» и два раза «молодой господин». В оригинальном тексте мы находим один раз «der junge Herr», 12 раз «der Herr» (без определения).

Получается, что скачок кривой отклонений вызван только переводческим решением «der (junge) Herr» = «молодой человек / господин». Мы не в коем случае не критикуем это решение уважаемого переводчика, но считаем, что такой вариант в значительной степени «размывает» оппозицию «замок – простые люди». Заметим, что сочетание «der junge Mann» (букв. «молодой человек») используется в романе, но ни разу в отношении представителей замковой бюрократии. Указанная оппозиция находит свое отражение и в контексте начала девятой главы:

*Pepi aber schien K. überhaupt erst zu bemerken, als er zum Ausschankpult trat und einen Kognak bestellte. K. lehnte dort, drückte die Hand an die Augen und kümmerte sich um nichts. Dann nippte er von dem Kognak und schob ihn zurück, weil er ungenießbar sei. „Alle **Herren** trinken ihn“, sagte Pepi kurz, goß*

den Rest aus, wusch das Gläschen und stellte es ins Regal. „Die Herren haben auch besseren“, sagte K. „Möglich“, sagte Pepi, „ich aber nicht.“ Damit hatte sie K. erledigt und war wieder dem Herrn zu Diensten, der aber nichts benötigte und hinter dem sie nur im Bogen immerfort auf und ab ging, mit respektvollen Versuchen, über seine Schultern hinweg einen Blick auf die Papiere zu werfen...

В приведенном отрывке отчетливо показано, что есть два мира: господа (Herren) и все остальные. У господ лучше напитки. Если кто-то из господ в комнате, то обращать внимание нужно только на него. Пренебрежительный тон в разговоре о господах категорично осуждается.

Выдвинем предположение, что выявленное различие в восприятии качественной характеристики, выраженное скачком на диаграмме, является следствием выбранного варианта перевода «молодой человек» вместо, на наш взгляд, более подходящего эквивалента «господин».

Дискуссия и заключение

В целом задачи исследования были решены и цель работы достигнута. Как и в любой исследовательской деятельности, а тем более если она носит прикладной характер, важно определить уровень достоверности получаемых результатов. В нашем случае, поскольку мы используем процедуры формального анализа в отношении трудно формализуемых смыслов художественного текста, необходимо понимать величину неизбежно возникающей погрешности. В дальнейшем, в ходе развития предлагаемого метода исключение «слабых мест» позволит скорректировать метод в сторону повышения точности данных на выходе.

В рамках решения первой задачи исследования погрешности появляются на стадии поиска заданных словоформ. Для нас очевидно, что использование sраСу-метода позволит значительно увеличить точность результата. В частности, при простом поиске последовательности символов «frei» мы получим словоформы «**freilich**» или «**Dorfreise**», которые не имеют никакого отношения к семе «свобода». Библиотека sраСу способна осуществлять поиск по начальной форме, каким-либо другим грамматическим формам и сортировку по частям речи, пусть, как мы отмечали выше, и с некоторой погрешностью. Важно, что для каждого языка,

заложенного в базы данных библиотеки, имеется свой набор сортировок, свойственных этому языку. Также достоверность результата зависит от выбора единиц, характеризующих качественные характеристики, хотя если речь идет о том, чтобы сопоставить в дальнейшем эти единицы с их переводными вариантами, сам выбор уже уходит на второй план, уступая место точности выбора потенциальных переводных вариантов.

При экстраполяции метода на другие языки точность данных будет зависеть от профессионализма исследователя, от того, насколько глубоко он понимает языки, с которыми работает. В нашем случае переключение с немецкого языка на русский было достаточно простым, поскольку оба языка принадлежат одной языковой семье и поскольку мы опирались на части речи – прилагательные и наречия, – имеющиеся в обоих языках. Но если бы исследовательская задача заключалась в том, чтобы оценить адекватность перевода средств категории определенности / неопределенности (артиклей), то это потребовало бы несравнимо больших усилий.

Визуализация данных – самая простая, техническая, задача, в которой трудно совершить ошибку. Работа исследователя здесь будет заключаться в том, чтобы выбрать максимально наглядный способ представления данных.

Чрезвычайно трудной является последняя задача, т. к. она требует от исследователя помимо отличного знания касающихся языков и развитых переводческих умений еще и глубокого понимания оригинального текста художественного произведения. Соответственно, погрешности будут вызваны в этом случае субъективными факторами.

Заключим, что созданный в результате настоящего исследования инструментарий, при правильном его применении, может оказать специалисту существенную помощь при оценке своего собственного или чужого перевода. Безусловно, не следует считать предлагаемые программы частью программного обеспечения для осуществления автоматического перевода, хотя работа и вносит определенный вклад в теорию машинного перевода. Качественная характеристика, рассмотренная нами, не ограничивает возможности созданного метода. Представляется, что сходным образом можно работать практически с любой философской

категорией, если исследователь сумеет формализовать набор языковых средств, которые ее отражают.

В перспективе планируется апробировать инструментарий на материале англоязычных художественных произведений и их переводов на русский язык, комплексно рассмотреть категорию качества, а также категории пространства и времени.

Литература

1. Гак В. Г., Григорьев Б. Б. Теория и практика перевода. Французский язык. – М.: Интердиалект+, 2003.
2. Гарбовский Н. К. Теория перевода. – М.: МГУ, 2004.
3. Горожанов А. И., Гусейнова И. А. Прикладные аспекты анализа и интерпретации текстов (на материале немецкого и русского языков). – Казань: Бук, 2021.
4. Горожанов А. И., Степанова Д. В. Интерпретация художественного произведения: корпусный подход // Филологические науки. Вопросы теории и практики. Тамбов: Грамота, 2022. – Т. 15. – № 1. – С. 203–208. – DOI 10.30853/phil20220020.
5. Комиссаров В. Н. Теория перевода (лингвистические аспекты). – М.: Высшая школа, 1990.
6. Малюга Е.Н. К вопросу об обучении иностранным языкам с использованием новых информационных технологий // Вопросы прикладной лингвистики. – 2009. – № 1. – С. 91-95.
7. Полякова Н. В. Оценка качества письменного перевода: проблема поиска эффективных стандартов, критериев и параметров // Вестник ТГПУ. – 2022. – № 1(219). – С. 54- 64. – DOI 10.23951/1609-624X-2022-1-54-64.
8. Рябко Е. И. Оценка качества художественного перевода и его социальной значимости // Индустрия перевода. – Пермь: ПНИПУ, 2020. Т. 1. – С. 133–138.
9. Сдобников В. В. Оценка качества перевода (коммуникативно-функциональный подход): монография. – М.: ФЛИНТА; Наука, 2015.
10. Семко С. А. Лекции по теории перевода // Серия «Язык. Культура. Коммуникация». – Вып. 5. – Н. Новгород: НГЛУ им. Н. А. Добролюбова, 2005.

11. Томалин Б., Малюга Е. Н. Деловой язык в век глобализации // Вопросы прикладной лингвистики. – 2016. – № 24. – Pp.7-18.
12. Швейцер А. Д. Теория перевода: статус, проблемы, аспекты. – М.: Наука, 1988.
13. Auffarth B. Artificial Intelligence with Python Cookbook: Proven recipes for applying AI algorithms and deep learning techniques using TensorFlow 2.x and PyTorch 1.6. – Birmingham: Packt Publishing Ltd., 2020.
14. Ayre K., Bittar A., Kam J., Verma S., Howard L. M., Dutta R. Developing a natural language processing tool to identify perinatal self-harm in electronic healthcare records // PLoS ONE. – 2021. – No. 16(8). DOI: 10.1371/journal.pone.0253809.
15. Carrier M. Automated Speech Recognition in language learning: Potential models, benefits and impact // Training, Language and Culture. – 2017. – №1(1). – С. 46-61. doi: 10.29366/2017tlc.1.1.3
16. Fawcett P. Translation and Language. Linguistic Theories Explained. – Manchester, UK & Northampton, MA: St. Jerome Publishing, 2003.
17. Gorfée, D. Semiotics and the Problems of Translation (With Specific Reference to the Semiotics of Charles S. Peirce). – Amsterdam-Atlanta, 1994.
18. Gorozhanov A. I., Guseynova I. A. Programming for specific purposes in linguistics: A new challenge for the humanitarian curricula. Training, Language and Culture. – 2020. – №4(4). – Pp. 23-38. Doi: 10.22363/2521-442X-2020-4-4-23-38
19. Jugran S., Kumar A., Tyagi B. S., Anand V. Extractive automatic text summarization using SpaCy in python NLP // 2021 International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021. – Pp. 582–585. DOI: 10.1109/ICACITE51222.2021.9404712.
20. Karsdorp, F., Kestemont, M., Riddell, A. Humanities Data Analysis: Case Studies with Python. – Princeton University Press, 2021.
21. Okhapkin, V. P., Okhapkina, E. P., Iskhakova, A. O., Iskhakov A. Y. Constructing of Semantically Dependent Patterns Based on SpaCy and StanfordNLP Libraries // Communications in Computer and Information Science (in books). – 2021. – Vol. 1395 CCIS. – Pp. 500–512. DOI: 10.1007/978-981-16-1480-4_45.

22. spaCy // Russian Model. URL: <https://spacy.io/models/ru> (дата обращения: 15.03.2022).
23. Waard, J. de & Nida, E. A. *From One Language to Another (Functional Equivalence in Bible Translating)*. – Nashville: Thomas Nelson Publishers, 1986.
24. Welnitzova, K., Jakubickova, B., Králik, R. Human-Computer Interaction in Translation Activity: Fluency of Machine Translation // *RUDN Journal of Psychology and Pedagogics*. – 2021. – Vol. 18. – No 1. – P. 217-234. – DOI 10.22363/2313-1683-2021-18-1-217-234.

References

- Auffarth, B. (2020). *Artificial Intelligence with Python Cookbook: Proven recipes for applying AI algorithms and deep learning techniques using TensorFlow 2.x and PyTorch 1.6*. Birmingham: Packt Publishing Ltd.
- Ayre, K., Bittar, A., Kam, J., Verma, S., Howard, L. M., & Dutta R. (2021). Developing a natural language processing tool to identify perinatal self-harm in electronic healthcare records. *PLoS ONE*, 16(8). DOI: 10.1371/journal.pone.0253809.
- Carrier, M. (2017). Automated Speech Recognition in language learning: Potential models, benefits and impact. *Training, Language and Culture*, 1(1), 46-61. doi: 10.29366/2017tlc.1.1.3
- Fawcett, P. (2003). *Translation and Language. Linguistic Theories Explained*. Manchester, UK & Northampton, MA: St. Jerome Publishing.
- Gak, V. G. & Grigoriev, B. B. (2003). *Theory and practice of translation. French*. Moscow: Interdialect+.
- Garbovsky, N. K. (2004). *Theory of translation*. Moscow: MSU.
- Gorlée, D. (1994). *Semiotics and the Problems of Translation (With Specific Reference to the Semiotics of Charles S. Peirce)*. Amsterdam-Atlanta.
- Gorozhanov, A. I. & Guseynova, I. A. (2021). *Applied aspects of text analysis and interpretation (on the material of German and Russian languages)*. Kazan: Buk. 208 p. ISBN 978-5-00118-759-2.
- Gorozhanov, A. I. & Stepanova, D. V. (2022). Interpretation of a work of art: corpus approach. *Philological Sciences. Questions of theory and practice*, 15(1), 203–208. DOI 10.30853/phil20220020.

- Gorozhanov, A. I., & Guseynova, I. A. (2020). Programming for specific purposes in linguistics: A new challenge for the humanitarian curricula. *Training, Language and Culture*, 4(4), 23-38. Doi: 10.22363/2521-442X-2020-4-4-23-38
- Jugran, S., Kumar, A., Tyagi, B. S., & Anand, V. (2021). Extractive automatic text summarization using SpaCy in python NLP. In *2021 International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2021*, (pp. 582–585). DOI: 10.1109/ICACITE51222.2021.9404712.
- Karsdorp, F., Kestemont, M., & Riddell, A. (2021). *Humanities Data Analysis: Case Studies with Python*. Princeton University Press. ISBN 978-0-691-20033-0.
- Komissarov, V. N. (1990). *Theory of translation (linguistic aspects)*. Moscow: Vysshaja shkola.
- Malyuga, E.N. (2009). On the issue of teaching foreign languages using new information technologies. *Issues of Applied linguistics*, 1, 91-95.
- Okhapkin, V. P., Okhapkina, E. P., Iskhakova, A. O., & Iskhakov, A. Y. (2021). Constructing of Semantically Dependent Patterns Based on SpaCy and StanfordNLP Libraries. In *Communications in Computer and Information Science (in books)*. Vol. 1395 CCIS. Pp. 500–512. DOI: 10.1007/978-981-16-1480-4_45.
- Polyakova, N. V. (2022). Evaluation of the quality of written translation: the problem of finding effective standards, criteria and parameters. *Vestnik of TSPU*, 1(219), 54–64. DOI 10.23951/1609-624X-2022-1-54-64.
- Ryabko, E. I. (2020). Evaluation of the quality of literary translation and its social significance. *Translation Industry*, 1, 133–138.
- Schweitzer, A. D. (1988). *Translation theory: status, problems, aspects*. Moscow: Nauka.
- Sdobnikov, V. V. (2015). *Evaluation of translation quality (communicative-functional approach): monograph*. Moscow: FLINTA; Nauka.
- Semko, S. A. (2005). Lectures on the theory of translation. In *Series "Language. Culture. Communication"*. Iss. 5. Nizhny Novgorod: NSLU after N. A. Dobrolyubov.
- spaCy Russian Model. (2022). URL: <https://spacy.io/models/ru> (accessed on 15.03.2022).

- Tomalin, B., Malyuga, E.N. (2016). Business language in the global age. *Issues of Applied Linguistics*, 24, 7-18.
- Waard, J. de & Nida, E. A. (1986). *From One Language to Another (Functional Equivalence in Bible Translating)*. Nashville: Thomas Nelson Publishers.
- Welnitzova, K., Jakubickova, B., & Králik, R. (2021). Human-Computer Interaction in Translation Activity: Fluency of Machine Translation. *RUDN Journal of Psychology and Pedagogics*, 18(1), 217-234. DOI 10.22363/2313-1683-2021-18-1-217-234.

УДК 32.019.52, 323.234

<https://doi.org/10.25076/vpl.45.04>

Н. К. Радина

Национальный исследовательский Нижегородский
государственный университет им. Н. И. Лобачевского

А. В. Козлова

Харман

ТОНАЛЬНОСТЬ КАК ФАКТОР УСПЕШНОСТИ ЭЛЕКТРОННОЙ ПЕТИЦИИ (НА МАТЕРИАЛАХ CHANGE.ORG)

В статье представлены результаты анализа тональности русскоязычных электронных петиций с международной онлайн-платформы Change.org. Эмпирическую базу исследования составили 22450 петиций на русском языке, размещенных на интернет-ресурсе Change.org с 2012 по 2016 г. В статье сравниваются тональности популярных и победивших петиций. Исследователи ранее использовали анализ тональности электронных петиций на английском и французском языках, который показал, что успех петиции связан с лексикой, имеющей положительный оттенок. Данное исследование доказывает, что современный контекст российских политических отношений требует более сложного подхода при выборе эмоциональной окраски для того или иного текста. Было высказано предположение, что сложная лингвистическая стратегия для выявления тональности текста электронной петиции обусловлена экстралингвистическим контекстом, а также социальными и