www.volsu.ru

ПОТЕНЦИАЛ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В АВТОМАТИЧЕСКОЙ ОБРАБОТКЕ ЕСТЕСТВЕННОГО ЯЗЫКА И МАШИННОМ ПЕРЕВОДЕ

DOI: https://doi.org/10.15688/jvolsu2.2024.5.1

UDC 81’322.4
LBC 81.184

Submitted: 13.05.2024
Accepted: 20.08.2024

LEXICOGRAPHIC PROBLEMS OF MACHINE TRANSLATION SYSTEMS: ON THE WAY FROM LITERAL TO NEURAL

Larisa N. Beliaeva
Herzen State Pedagogical University of Russia, Saint Petersburg, Russia

Olga N. Kamshilova
Herzen State Pedagogical University of Russia, Saint Petersburg, Russia;
Saint Petersburg University of Management Technologies and Economics, Saint Petersburg, Russia

Abstract. The article discusses some current issues of interpreting out-of-vocabulary words by modern machine translation systems (MT systems) in the context of changing forms and ways of maintaining an automatic dictionary. It provides a critical outline of the typology of MT systems and strategies for their development. It describes the impact of fast developing software and technologies on these strategies and analyzes the changes they bring into the forms of dictionary support. The research shows that the linguistic support and the structure of automatic dictionaries, whatever the MT system is, are fundamentally important for ensuring the quality of translation. Despite all the success of neural MT (NMT) systems, their automatically updated vocabulary databases do not record words characterized by terminological specificity and low frequency in the special texts and text corpora on which the system is trained. Analysis of translations performed by two popular NMT systems – Google Translate and Yandex Translate – has proven that they fail to process and unify the translation of words that are not entered in the system dictionaries, a task used to be solved easily by users of all types of MT systems with the help of automatic dictionaries. With statistic-based automatic dictionaries it remains a pressing problem and requires a special approach when editing MP results.

Key words: machine translation, machine translation strategy, typology of machine translation systems, automatic dictionary, out-of-vocabulary words, linguistic support.

Citation. Beliaeva L.N., Kamshilova O.N. Lexicographic Problems of Machine Translation Systems: On the Way from Literal to Neural. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2024, vol. 23, no. 5, pp. 6-19. (in Russian). DOI: https://doi.org/10.15688/jvolsu2.2024.5.1

УДК 81’322.4
ББК 81.184

Дата поступления статьи: 13.05.2024
Дата принятия статьи: 20.08.2024

ЛЕКСИКОГРАФИЧЕСКИЕ ПРОБЛЕМЫ СИСТЕМ МАШИННОГО ПЕРЕВОДА: НА ПУТИ ОТ БУКВАЛЬНОГО ДО НЕЙРОННОГО

Лариса Николаевна Беляева

Российский государственный педагогический университет им. А.И. Герцена, г. Санкт-Петербург, Россия

Ольга Николаевна Камшилова

Российский государственный педагогический университет им. А.И. Герцена, г. Санкт-Петербург, Россия;
Санкт-Петербургский университет технологий управления и экономики, г. Санкт-Петербург, Россия

Аннотация. В статье рассматриваются актуальные вопросы интерпретации современными системами машинного перевода (МП) лексики, неизвестной этим системам (out-of-vocabulary words), в контексте изменений форм и ведения автоматического словаря. Дан критический очерк типологии систем МП и стратегий их развития. Описаны особенности этих стратегий и влияние на них развивающихся программных средств и технологий. Проанализированы формы ведения словарной поддержки, меняющиеся под воздействием технологических условий. Показано, что при любой системе МП ее лингвистическое обеспечение и структура автоматических словарей становятся принципиально важными для поддержания качества перевода. При всем успехе развития нейронных систем МП (НМП) их автоматически пополняемые словарные базы не фиксируют слова, характеризующиеся терминологической спецификой и низкой частотой в массивах и корпусах текстов, на которых обучается система. На примере анализа результатов двух востребованных НМП – Google Translate и Yandex Translate – доказано, что обработка и унификация перевода слов, не вошедших в словари системы, прежде легко решавшаяся пользователями всех типов систем МП на основе пополнения и ведения автоматического словаря, остается по-прежнему актуальной проблемой и требует особого подхода при редактировании результатов НМП.

Ключевые слова: машинный перевод, стратегия машинного перевода, типология систем машинного перевода, автоматический словарь, неизвестное слово, лингвистическая поддержка.

Цитирование. Беляева Л. Н., Камшилова О. Н. Лексикографические проблемы систем машинного перевода: на пути от буквального до нейронного // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2024. – Т. 23, № 5. – С. 6–19. – DOI: https://doi.org/10.15688/jvolsu2.2024.5.1

Введение

Длинный путь, пройденный прикладной лингвистикой от идеи машинного перевода (далее – МП) до ее современной реализации в виде сетевых систем нейронного перевода, требует осмысления тех трудностей, с которыми сталкивались разработчики лингвистического и программного обеспечения систем. Начавшись с чудовищной по затратам труда разработки автоматических словарей и процедур бинарного МП, пройдя через изменения самого парка компьютеров, современные сетевые системы предлагают высокий уровень перевода.

Тип конкретной системы МП определяется особенностями реализованных в нем грамматических (лексических, грамматических и собственно синтаксических анализаторов – парсеров) и программных средств. Однако часть лексических проблем, в частности обработка слов, не вошедших в словари системы, которые легко решались пользователями всех типов систем МП, осталась по-прежнему актуальной и требует особого подхода при редактировании результатов МП.

Развитие идеологии и теории машинного перевода и разработка систем, реализующих предлагаемые разработчиками подходы, всегда непосредственно зависели от уровня вычислительной техники: скорости обработки текста в оперативной памяти, скорости обмена с внешними (периферийными) устройствами, объемом этих устройств и памяти конкретного компьютера, универсальности программного обеспечения системы МП или его зависи-

мости от типа доступного разработчикам устройства [Almansoori et al., 2020].

Огромные усилия в ранние годы МП были потрачены на преодоление технических трудностей, особенностей представления текста в цифровой форме, его перфорацию, детальную проверку результатов перфорации и преодоление ограничений перфорационной техники с точки зрения полноты алфавитов, отсутствия надстрочных знаков и т. п. Сложность оцифровки печатных текстов была несколько снижена разработкой и усовершенствованием сканирующих устройств, первые из которых способны были распознавать только один шрифт и требовали специальной сложной перенастройки на другой. Появление современного типографского оборудования, мощных сканнеров и соответствующего программного обеспечения облегчили перевод текстов в цифровую форму и позволили оперативно создавать и корпуса текстов, и базы данных.

Довольно долго накапливаемые массивы текстов использовались при создании автоматических словарей систем МП для извлечения актуальной лексики и ее дальнейшего преобразования в форму, требуемую конкретной системой МП, для исследования грамматических особенностей текстов и разработки соответствующих алгоритмов, основанных прежде всего на внесении в автоматический словарь специальных характеристик лексических единиц (слов и оборотов), используемых в дальнейшем при реализации алгоритмов синтаксического анализа (парсинга) [Dankers, Bruni, Hupkes, 2022].

Традиционно в любой системе МП принято выделять грамматические алгоритмы и лингвистическое обеспечение – автоматический словарь или словари, однако подход к структуре, составу и способам взаимодействия этих компонентов обусловлен функциями системы [Brottrager et al., 2022; Беляева, 2022]. Выбор конкретной стратегии машинного перевода является и теоретической, и практической проблемой МП. Этот выбор сводится к обсуждению возможных теоретических стратегий, в число которых входят традиционные подходы: прямой бинарный МП, разные версии перевода с трансфером (см., например: [Zhuang et al., 2021; Tars, Tättar, Fišel, 2022]) и расширенными сетями переходов, перевод с языком-посредником; а также новые стратегии: статистический МП и нейронный МП [Zhu et al., 2022].

При всем разнообразии существующих систем МП их типологические различия попрежнему определяются спецификой реализованных в них грамматических, лексических и программных средств.

Дискуссия

Опыт построения и эксплуатации давно применяемых систем *прямого бинарного МП* показывает, что эффективность этих систем обеспечивается в основном за счет мощных словарных средств, накопленных за десятилетия их функционирования, и специальных средств редактирования. Однако прямое бинарное сопоставление, осуществляемое на основе отдельных лексико-морфологических и семантико-синтаксических подструктур, жестко ориентирует лингвистическое и программное обеспечение системы не только на особенности конкретной пары языков, но и на специфику подъязыков. Замена одного из членов такой пары немедленно требует перестройки лингвистического, а иногда и программного обеспечения.

Вторая стратегия, предусматривающая *применение трансфера в системах МП*, основывается на теоретическом предположении о невозможности создания записи для прямой связи текстов разных языков. Эта связь может быть установлена только путем последовательных преобразований на различных уровнях репрезентации: лексическом, морфологическом, уровне синтаксических структур (в терминах конфигурационного и реляционного синтаксиса) и на уровне синтаксического описания, для чего активно используются варианты расширенных сетей перевода.

Использование трансфера предполагает последовательный переход от поддеревьев, описывающих структуру входного предложения на различных уровнях анализа, к глубинной ролевой структуре, используемой в качестве базового грамматического представления, и далее к входным поддеревьям. При этом использование процедуры преобразования именно для локальных поддеревьев позволяет получать МП, не моделируя пол-

ное понимание предложения. Решение этой задачи не обязательно при МП, так как глубинные семантико-синтаксические интерпретации могут быть оставлены человеку-пользователю, то есть специалисту, работающему с переводом, полученным с помощью компьютера. Такой подход к использованию трансфера характерен не только для большинства работающих зарубежных алгоритмов [Liu et al., 2022], но и для отечественных систем МП.

Использование трансфера в реально действующих системах МП подтверждает целесообразность выбора этой стратегии. В то же время уровень развития систем искусственного интеллекта и новые возможности вычислительной техники определили новое пробуждение интереса к третьей стратегии МП – моделированию понимания с помощью языков-посредников.

Увязка *систем МП, реализованных на основе языка-посредника*, и систем искусственного интеллекта определяется необходимостью использовать в них сложные базы знаний, с помощью которых должно быть распознано значение предложения или содержание целого текста и автоматически сгенерирован его пересказ на выходном языке.

Моделирование концептуального уровня понимания, являющегося промежуточным уровнем систем искусственного интеллекта, основывается на нескольких базовых составляющих [Беляева, 2016]:

– языках представления знаний, требующих организации словарей в виде статей, включающих узлы-понятия и типы связей (свойства);

– системных грамматиках, осуществляющих преобразование по принципу «признак – значение».

Язык представления знаний, используемый в подобных системах, и является языком-посредником, осуществляющим семантическую «разметку» входного предложения относительно семантической сети и генерацию выходной фразы из этого внутреннего представления. Такая генерация реализуется по образцам (шаблонам, фреймам и т. п.). При всей общетеоретической привлекательности этого подхода он останется экспериментальным до тех пор, пока не будут созданы лингвистические основы и программные средства для построения мощных семантических сетей и онтологий на основе автоматизированных лексиконов.

Практически до конца XX в. методы статистического (даже, точнее, количественного) анализа использовались при разработке систем МП на уровне пополнения словарей за счет актуальной терминологии. Расширение объемов оцифрованных текстов, доступных в сети Интернет, изменило подход к возможностям применения статистики. Особая роль в этом процессе принадлежит текстам параллельным, в которых одно и то же содержание представлено на нескольких языках благодаря переводу документов.

Наиболее часто используемыми параллельными текстами были протоколы заседаний или другие официальные документы стран с несколькими официальными языками (Канада, Швейцария, Гонконг и др.) и документы международных организаций. Одна из причин использования таких текстов при решении задач обработки данных на естественном языке заключается в том, что они открыты, их достаточно легко получить, у таких текстов, как правило, большой объем. Кроме того, предполагается, что требования точности заставляют переводчиков подобных материалов использовать корректные литературные переводы. Наличие такого рода материалов явилось стимулом для нового витка интереса к применению статистических методов и в решении задач МП, включая вероятностное моделирование, теорию информации и линейную алгебру.

Основой использования статистических методов является идея о том, что значение слова определяется распределением контекстов, в которых используются слова и более сложные выражения. *Системы статистического машинного перевода* (*Statistical Machine Translation – SMT*) доминировали на протяжении нескольких десятилетий там, где не было необходимости в жестко предметно-ориентированном переводе. Базой для их реализации стала разработка параллельных корпусов текстов, в которых данные выравнивались на уровне отдельных слов, словосочетаний и предложений. В основе процедуры перевода лежит количественная оценка выровненных сегментов в параллельном корпусе [Раренко, 2021].

Системы *нейронного машинного перевода* (далее – НМП) сегодня стали доминирующим инструментом в области МП, свидетельством чему является их включение в программные средства компаний Google, Microsoft, Яндекс, DeepL и др., которые ранее использовали статистический МП.

Нейронными сетями (далее – НС) называются структуры, которые способны имитировать процессы, моделирующие работу человеческого мозга. Важной особенностью нейронных сетей является способность к обучению, что обеспечивает их преимущество перед классическими алгоритмами, основанными на переборе вариантов.

У нейронных сетей есть несколько существенных преимуществ по сравнению с порождающими языковыми моделями на основе N-грамм. Архитектура таких сетей легко адаптируется для того, чтобы проецировать входные слова в представления меньшей размерности (например, измерения размером 500, где слова представлены в виде 500-позиционных векторов). Это позволяет автоматически группировать похожие слова во вложенном пространстве (*embedding space*). Представления слов в таком пространстве в настоящее время описываются как векторные представления слов (*word embeddings*). Более того, конкретные лексико-грамматические свойства слов кодируются в соответствии с размерами этих распределенных представлений [Popović, 2017].

Сегодня принято различать два основных вида нейронных сетей: рекуррентные (*recurrent neural networks* – RNN) и сверточные нейронные сети (*convolutional neural networks* – CNN). Для решения различных задач анализа текстов на естественном языке и распознавания звучащей речи используются рекуррентные нейронные сети. Сверточные нейронные сети хорошо справляются с такими задачами, как классификация изображений и предложений, а также сентимент-анализ. Особенностью рекуррентных сетей является возможность использовать предыдущие выходные данные в качестве входных, имея при этом скрытые состояния. Применение этих сетей обеспечивает поддержания модульной структуры системы, поскольку они поддерживают передачу параметров, что особенно важно для анализа текста при МП. При этом размеры модели необязательно увеличиваются с увеличением объема ее входных данных. Обычно используемые варианты RNN включают два варианта архитектуры: двунаправленную (*bidirectional* – BRNN) и глубинную (*deep* – DRNN) [Lankford, Afli, Way, 2021, p. 49–50].

Предполагается, что в отличие от систем «перевода по правилам» системы нейронного перевода работают с текстом в целом, а не с отдельными предложениями. Это утверждение, к сожалению, не соответствует действительности: анализ переводов слов, не зарегистрированных в системе, демонстрирует несовпадение их переводов. Архитектура системы нейронного машинного перевода включает две рекуррентные нейронные сети, одна из которых ответственна за обработку входной текстовой последовательности, а другая – за формирование выходного текста-перевода [Нуриев, 2019]. Соответственно, в системах НМП есть два важных компонента: кодер и декодер, которые работают и с исходным текстом, и с текстом, порождаемым системой. Системы имеют механизмы машинного обучения, позволяющие развивать лингвистическое обеспечение по мере использования системы, основу этого обеспечения составляют словари.

Во всех традиционных алгоритмах МП трудно переоценить роль автоматического словаря, который всегда был базой для лексико-морфологического анализа обрабатываемого текста и основой для реализации алгоритмов парсинга. Автоматический словарь (далее – АС) являлся и остается ядерной частью системы МП, так как именно на основе заключенной в нем информации реализовалось все программное обеспечение лингвистических алгоритмов [Беляева, 2022].

При отборе лексики в словари систем МП ранее учитывался не только терминологический статус лексической единицы – слова или словосочетания (машинного оборота), но и ее распространенность в конкретном языке для специальных целей. Автоматические словари не являлись словарями нормативными, поскольку в качестве заглавия словарной статьи используются все встречающиеся варианты номинации объектов, а перевод соответствует рекомендуемому для конкретного

языка перевода и предметной области [Беляева, 2016]. Объем АС зависит от типа (уровня аналитизма) языка и реализованных в системе алгоритмов морфологического анализа. Эти алгоритмы определяют и способы хранения информации в АС, и саму структуру АС (включение в нее машинных основ, словоформ, фрагментов слов или единиц, больших, чем слово), и сами алгоритмы морфологического анализа. Эти алгоритмы морфологического анализа, детально разработанные для «традиционных» систем МП, используются и при нейронном МП, предполагая, в частности, процедуры морфологического анализа слов, в системе не зафиксированных, но имеющих в своем составе латинские или греческие префиксы, частотные суффиксы, устойчивое значение которых позволяет вычислять значение такого неопознанного слова как сумму значений префикса, суффикса и слова, уже зарегистрированного в базе лексических данных [Sennrich, Haddow, Birch, 2015].

Однако сами словари систем и статистического, и нейронного МП организованы иначе. Они создаются как электронные таблицы (*spreadsheets*), в которых выделяются строки, столбцы и отдельные ячейки. Ячейками в электронной таблице называются пространства, содержащие единицы информации. Каждая ячейка маркируется своим местом в таблице (например, A1, A2, A3, ...) и может иметь абсолютную или относительную отсылку к ячейкам, ее окружающим. Они могут хранить различную информацию, которая обрабатывается (извлекается, суммируется и т. д.) с помощью прикладных программ. Наполнение таких таблиц осуществляется по мере развития (машинного обучения) системы на материалах получаемых переводов. В электронные таблицы вводятся пары лексических единиц, извлекаемые из предметно-ориентированных и параллельных корпусов текстов. Основу элементов, включаемых в таблицы, составляют многокомпонентные словосочетания, являющиеся основным способом номинации элементов предметных областей. Поэтому такие таблицы принято называть таблицами словосочетаний (*phrase-table*).

В общем случае электронные таблицы предназначены для хранения количественных данных и коротких текстовых строк. Кроме того, они обеспечивают графическое представление отношений между данными. Электронные таблицы могут структурировать и маркировать элементы данных так же полно, как базы данных, и обычно не предполагают возможности обращаться к базе данных с запросами. Функции для работы с электронными таблицами предоставляют Google, Microsoft Excel, Lotus 1-2-3, VisiCalc.

Однако электронные таблицы не лишены недостатков. При обмене информацией между различными электронными таблицами они часто экспортируются как файлы, включающие значения, разделенные запятыми (*comma-separated value* – CSV). CSV является не единым форматом, а, скорее, дескриптором набора плохо определенных форматов, которые используют запятую, чтобы указать границы столбцов. Общая проблема обмена информацией между табличными представлениями состоит в том, что файлы в формате CSV могут быть в различных схемах кодирования и тип кодирования в файле не указывается, что делает их интерпретацию случайной. Microsoft Excel, возможно наиболее популярная программа работы с электронными таблицами, например, не может правильно загружать файлы в формате CSV, если они даны в коде UTF-8 (базовый формат для текстов, написанных на кириллице), и вместо этого требует кодировку ISO Latin-1, в результате чего для многих языков файлы CSV нельзя использовать в Excel.

Нейронные сети второго типа (CNN) работают с векторами слов и фильтрами, которые объединяют локальную информацию в рамках предложения. Чаще всего для получения наиболее эффективных результатов используется гибридный подход, предполагающий использование обоих видов сетей.

Для работы любой нейронной сети необходимо создание обучающих массивов данных. Процесс сбора и обработки данных делится на две различающихся процедуры: 1) сбор обучающих данных; 2) разработка и тестирование данных. Чаще всего для сбора данных, предназначенных для решения задач, связанных с обработкой текстов на естественных языках, используются уже существующие репрезентативные одноязычные или парал-

лельные корпусы текстов. Создание и предварительная разметка корпуса, ориентированного на новую с точки зрения предметной области или задач сферу, представляет собой чрезвычайно трудоемкую задачу. Поэтому при решении вопроса о возможности использовать нейронные сети в процессе обучения необходимо исследовать вопрос о наличии нейронных сетей, решающих требуемые задачи, и об их соответствии целям обучения [Peris, Casacuberta 2019].

Наибольшие улучшения были продемонстрированы, когда архитектура RNN или CNN была полностью отменена и заменена механизмом внимания, создающим гораздо более простую и быструю архитектуру, известную как Transformer [Devlin et al., 2019]. Модели трансформера фокусируются на ранее сгенерированных токенах. Такой подход позволяет моделям развивать длительную память, что особенно полезно в области перевода. Повышение производительности в CNN- и RNN-подходе может быть достигнуто за счет внедрения таких уровней в архитектуру перевода [Lankford, Afli, Way, 2021, p. 50]. В последние годы большие языковые модели на архитектуре трансформеров стали вершиной развития нейросетей в задачах обработки текстов на естественных языках.

Одной из мощнейших нейросетей, доступных в открытом доступе, является сеть ChatGPT, представляющая собой языковую модель, обученную OpenAI, которая использует глубинное обучение для генерации текста и ответов на вопросы. Эта модель была создана на основе технологии трансформеров, что позволяет обрабатывать большие объемы текста и понимать связи между словами и предложениями, для ее тренировки использовались методы обучения с учителем и с подкреплением.

На основе ChatGPT в 2021 г. была разработана модель YaLM. Нейросеть YaLM обучена на части индексируемых Яндексом страниц Рунета, включая не только Википедию, новостные статьи и книги, но и открытые записи пользователей социальных сетей и форумов. Чтобы не перегружать модель, из этого обучающего массива удаляются повторяющиеся, незаконченные и неестественные тексты, то есть данная сеть является предварительно настроенной (*pre-trained*). Эта сеть обучалась 65 дней на 1,7 ТБ текстов из Интернета, книг и множества других источников с помощью 800 видеокарт A100. Нейросеть YaLM содержит 100 млрд параметров и является самой большой из существующих моделей для русского и английского языков. Это реально позволяет использовать ее для решения большого круга задач, связанных с обработкой естественного языка. Языковые модели из семейства YaLM определяют принцип построения текста и генерируют новые.

Постановка проблемы

В отличие от автоматического словаря первых (бинарных) систем МП, состоящего из словарных статей, словари современных систем МП, реализованные как электронные таблицы, расширяются автоматически. Однако опыт использования систем НМП (например, Google Translate и Yandex Translate) демонстрирует провал работы этих систем при обработке лексических единиц, не зарегистрированных в электронных таблицах либо из-за их терминологической специфики, либо из-за низкой частоты использования в массивах и корпусах текстов. Обработка таких неопознанных слов является сегодня одной из самых острых проблем активно используемого нейронного МП, поскольку нарушает восприятие текста и затрудняет редактирование результатов МП [Araabi, Monz, Niculae, 2022].

В задачи настоящей статьи, наряду с определением типологии систем МП и описанием изменений в формах и способах ведения их словарных баз под влиянием технологических причин, входит выявление ошибок и анализ их причин в обработке слов, не опознанных системами нейронного МП.

Материалы и методы

Для оценки результатов распознавания неопознанных слов был проведен эксперимент с переводом 451 фрагмента текстов научных статей, посвященных современным проблемам перевода (см. табл. 1), включающих существительное *translationese* [Toral, 2019], ко-

торое отсутствует не только в словарном обеспечении двух наиболее часто используемых систем НМП Google Translate и Yandex Translate, но и в англо-русских переводных словарях. Хотя формально в этом слове можно выделить суффикс прилагательного *-ese* со значением '*обладающий данным качеством*', формально сформировать значение существительного *translationese* как сумму значений *translation* и *-ese* невозможно. Аналогичный вариант наблюдается со словом *interpretese*, где сочетаются глагол *interpret* и суффикс *-ese*.

Методы работы с экспериментальным материалом включали автоматическое извлечение контекстов с существительным *translationese,* их машинный перевод, а также количественный, морфологический, синтаксический и сопоставительный анализ предлагаемых вариантов перевода.

Таблица 1. **Список использованных материалов с количественными характеристиками фрагментов со словом *translationese***

Table 1. **List of materials used with quantitative characteristics of fragments with the word *translationese***

Название статьи	Число фрагментов	Частота *translationese*
Baroni M., Bernardini S. A New Approach to the Study of Translationese: Machine-Learning the Difference Between Original and Translated Text // Literary and Linguistic Computing. 2005.	16	17
Toral A. Reassessing Claims of Human Parity and Super-Human Performance in Machine Translation at WMT 2019 // Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, 3–5 November 2020. Online Conference. URL: https://aclanthology.org/2020.eamt-1.pdf	2	2
Toral A. Post-editese: an Exacerbated Translationese // Proceedings of Machine Translation Summit XVII, 2019. Vol. 1. Research Track. URL: https://www.mtsummit2019.com/workshops	8	9
Riley P., Caswell I., Freitag M., Grangier D. Translationese as a Language in "Multilingual" NMT. arXiv:1911.03823v1	52	58
Graham Y., Haddow B., Koehn P. Translationese in Machine Translation Evaluation. arXiv:1906.09833v1	14	15
Koppel M., Ordan N. Translationese and Its Dialects // Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oregon, June 19–24, 2011, pp. 1318–1326. URL: https://aclanthology.org/P11-1000	34	40
Towards the Classification of the Finnish Internet Parsebank: Detecting Translations and Informality // Nodalida 2015. Proceedings of the 20th Nordic Conference of Computational Linguistics. URL: https://aclanthology.org/W15-1800.pdf	1	1
Larkin S., Simard M., Knowles R. Like Chalk and Cheese? On the Effects of Translationese in MT Training // MT summit 2021. URL: https://cmt3.research.microsoft.com/MTSUMMIT2021	35	46
Wang J., Meng F., Zhang T., Liang Y., Xu J., Li Z., Zhou J. Understanding Translationese in Cross-Lingual Summarization. arXiv:2212.07220v1	82	92
Nikolaev D., Karidi T., Kenneth N., Mitnik V., Saeboe L., Abend O. Morphosyntactic Predictability of Translationese // Linguistics Vanguard. 2020. Vol. 6, № 1. P. 20190077. DOI: https://doi.org/10.1515/lingvan-2019-0077	9	11
Wein S., Schneider N. Translationese Reductionusing Abstract Meaning Representation. arXiv:2304.11501v1	75	94
Chen S. Effect of Translationese on Machine Translation Quality. URL: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/custom/15721959.pdf	15	16
Edunov S., Ott M., Ranzato M.'A., Auli M. On the Evaluation of Machine Translation Systems Trained with Back-Translation // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, July 5–10, 2020. URL: https://aclanthology.org/2020.acl-main	38	40
He H., Daume H. III, Boyd-Graber J. Interpretese vs. Translationese: The Uniqueness of Human Strategies in Simultaneous Interpretation. URL: https://aclanthology.org/N16-1111	20	20
Итого	451	512

Результаты исследования

Проведенный анализ показал отсутствие предсказуемости в выборе перевода слова, которое не входит в лингвистическое обеспечение этих систем. Именно эта непредсказуемость затрудняет постредактирование переводов текстов, принадлежащих конкретной предметной области [Беляева, Камшилова, Шубина, 2023]. Всего в исследуемом массиве выделено 48 разных переводов слова *translationese*. Важно отметить, что методы транслитерации или транскрипции при передаче этого слова практически не используются. Единственным исключением является формирование гибридной формы *переводез*, зафиксированной в переводах системы Google Translate с частотой 2:

Translationese has originally been described... as the set of "fingerprints" that one language leaves on another when a text is translated between the two. → Первоначально **переводез** был описан... как набор «отпечатков пальцев», которые один язык оставляет на другом, когда текст переводится с одного на другой.

Interference Effects in **Translationese** → Интерференционные эффекты в **переводезе**.

К особым переводам, формируемым этой же системой, относится форма *переводский (язык)*. Такого словотворчества в переводах, выполненных системой Yandex Translate, не зафиксировано.

Рассмотрим особенности перевода слова *translationese* в позициях, встречающихся чаще других (табл. 2).

В наибольшем числе выбранных системами эквивалентов выделяются различные словоформы лексемы *перевод* в характерных для нее словосочетаниях, фиксируемых табличными словарями системы автоматически, поскольку, как отмечалось выше, частотные словосочетания являются основой элементов,

Таблица 2. **Самые частотные переводы слова *translationese* в позиции ядра однокомпонентной именной группы**

Table 2. **The most frequent translations of the word *translationese* in the position of the core of a single-component nominal group**

№ п/п	Переводы	Частота в Yandex Translate	Частота в Google Translate	Суммарная частота перевода
1	Перевод	71	63	134
2	Переводческий язык	24	94	118
3	Translationese	34	7	41
4	Язык перевода	25	1	26
5	Переводы	12	4	16
6	Переводческий	7	8	15
7	Трансляционный язык	–	13	13
8	Переводческий (в конце)	7	4	11
9	Трансляция	7	4	11
10	Переводческие ошибки	8	–	8
11	Переводной	–	7	7
12	Трансляционный	1	6	7
13	Переводной язык	–	6	6
14	Переводческий диалект	–	4	4
15	Переводчик	–	4	4
16	Количество переводов	2	1	3
17	Переведенный текст	1	2	3
18	Перевод с английского	3	–	3
19	Переводной текст	–	3	3
20	Переводные тексты	3	–	3
21	Трансляции	–	3	3
22	Переводез	–	2	2
23	Переводные слова	2	–	2
24	Уровень перевода	1	1	2
25	Языки перевода	–	2	2

включаемых в словари такого типа (*переводческий язык*, *язык перевода*, *переводной язык*, *переводной текст* и т. д.). Кроме того, варьируются формы множественного и единственного числа, например: *перевод* / *переводы*, *переводной текст* / *переводные тексты*, *язык перевода* / *языки перевода*.

В традиционных системах МП было принято правило, при котором неизвестное системе слово подвергалось стандартному морфологическому анализу и при отсутствии частотных аффиксов оставалось непереведенным. В общем массиве примеров всего 57 демонстрируют отсутствие переводов, причем логика выбора именно этого варианта частично прослеживается в результатах. Так, в системе Google Translate устойчиво не переводится слово *translationese* в форме притяжательного падежа (частота 4) и в угловых скобках (частота 2).

Второй случай также характерен для системы Yandex Translate. Во всех остальных случаях установить причины, по которым слово осталось без перевода, не представляется возможным. Поэтому системы переводят это слово исходя из заложенных в них результатов формирования нейросетей. Соответственно, эти результаты требуют дополнительного и сложного редактирования.

Важно отметить, что часто разные системы используют переводы, которые не используются другой системой или используются при переводе других предложений. Так, например, только система Google Translate использует в качестве переводов следующие слова и словосочетания: *переводез*, *переводной*, *переводной диалект*, *переводной текст*, *переводной язык*, *переводоведение*, *переводский язык*, *переводческий диалект*, *переводческий стиль*, *переводческий характер*, *переводчик*, *переводчики*, *трансляции*, *трансляционизм*, *трансляционность*, *трансляционный сигнал*, *трансляционный стиль*, *трансляционный язык*, *языковой перевод*, *языки перевода*. В противоположность этому только система Yandex Translate использует в качестве перевода следующие единицы: *объем перевода*, *ошибки перевода*, *перевод на русский язык*, *переводные слова*, *переводные тексты*, *переводческие ошибки*, *переводческое мастерство*, *письменный перевод*, *сигналы трансляции*, *сложность перевода*, *точность перевода*, *переводческие тексты*.

Сопоставительный анализ результатов работы систем показывает, что в системе Yandex Translate предусмотрено дополнение перевода за счет вставки слов, расширяющих толкование. Так, в конструкциях с переходным глаголом со значением изменения при переводе системой Yandex Translate добавляется существительное со значением объема, однако выбор этого существительного оказывается произвольным. Например, появление в предложении глагола *reduce* с прямым объектом *translationese* приводит к вставке в результат МП существительных с*ложность*, *объем*, *количество*, а перевод самого глагола *reduce* также варьируется в зависимости от выбранной модели преобразования синтаксической конструкции как *сокращение*, *уменьшить* (табл. 3).

Таблица 3. **Примеры расширенного перевода *translationese* в позиции прямого дополнения к *reduce***

Table 3. **Examples of extended *translationese* translation in the position of direct object to *reduce***

Текст на входе	Текст на выходе
As suggested by Freitag et al. (2020), diversified paraphrasing strategy might be a good solution to reduce the **translationese**	Как было предложено Фрейтагом и др. (2020), диверсифицированная стратегия перефразирования может быть хорошим решением для сокращения **сложности перевода**
Therefore, the post-editing strategy cannot reduce the **translationese**	Следовательно, стратегия постредактирования не может уменьшить **объем перевода**
In order to make the machine-translated documents or summaries suitable for evaluating model performance, some postprocessing strategies should be conducted to reduce the **translationese** in them	Для того чтобы документы или резюме, переведенные на машинный перевод, были пригодны для оценки производительности модели, следует использовать некоторые стратегии постобработки, чтобы уменьшить **количество переводов** в них

Иногда выбор такого дополнения определяется ядром словосочетания, в которое входит существительное *translationese*. Например, существительные *influence of* и *effect of* несмотря на один и тот же перевод на русский язык как *влияние*, явно имеют в системе МП разное описание типа воздействия (табл. 4).

Анализ результатов перевода слова *translationese* подтверждает, что при использовании систем НМП перевод слов, отсутствующих в словаре системы, произволен и должен особым образом проверяться и устанавливаться [Беляева, Камшилова, Шубина, 2023].

Заключение

Использование систем МП давно стало элементом профессиональной работы как специалистов в конкретных областях знаний, так и профессиональных переводчиков (см., например: [Khoong, Rodriguez, 2022]), при этом ни одна из предлагаемых сегодня систем не гарантирует профессионального качества продукта перевода без привлечения работы переводчика в качестве эксперта и редактора. Прикладная лингвистика прошла длинный путь от самой идеи машинного перевода к ее реализации в виде современных нейронных систем МП. Стратегии МП менялись под воздействием новых технологий разработки программного и лингвистического обеспечения, изменения самого парка компьютеров, характера доступности, способов обработки и хранения текстовых баз. При этом выбор конкретной стратегии МП – это всегда теоретическая и практическая задача, которая решается создателями конкретной системы МП. Однако при всех изменениях ядерной частью любой системы МП был и остается автоматический словарь.

Изменение форм ведения и способов пополнения словарной базы системы МП привело к тому, что словари наиболее востребованных и производительных нейронных систем перевода реализуются в виде автоматически пополняемых (и, соответственно, исключающих вмешательство переводчика-эксперта) таблиц и настроены на фиксацию частотных лексических единиц (слов и словосочетаний) в обрабатываемом массиве. При этом в словарную базу не попадают лексические единицы, претендующие на терминологический статус или характеризующиеся низкой частотой в обрабатываемых массивах или корпусах текстов.

Поскольку, вопреки декларируемым свойствам, нейронные системы МП работают не на уровне текста, а на уровне предложения (и, таким образом, не наследуют опыт перевода не опознанной словарем лексемы в пределах того же текста), интерпретация ими неизвестной лексики приводит к применению нескольких способов, обусловленных технологическими возможностями системы, ни один из которых не может считаться собственно переводом. Анализ работ систем Google Translate и Yandex Translate показал, что неопознанная лексема «переводится» следующими способами:

Таблица 4. **Примеры результатов перевода *translationese* в словосочетаниях *influence of* и *effect of***

Table 4. **Examples of *translationese* translation results in the phrases *influence of* and *effect of***

Текст на входе	Текст на выходе
Some researchers explore the influence of **translationese** on MT evaluation...	Некоторые исследователи исследуют влияние **языка перевода** на оценку МТ...
To control the effect of **translationese** on MT models, tagged training... is proposed to explicitly tell MT models if the given data is translated texts	Чтобы контролировать влияние **перевода** на модели МТ, предлагается обучение с тегами... чтобы явно указывать моделям МТ, являются ли данные переведенными текстами
...Mitigate the effect of **translationese** in cross-lingual transfer learning	...Смягчают влияние **перевода** при обучении межъязыковому переводу
In this paper, we investigate the influence of **translationese** on CLS	В этой статье мы исследуем влияние **языка перевода** на CLS
Subsequently, we can use the summaries in HT text and MT text as references, respectively, to train CLS models and analyze the influence of **translationese** on the model performance	Впоследствии мы можем использовать резюме в тексте НТ и тексте МТ в качестве ссылок, соответственно, для обучения моделей CLS и анализа влияния **перевода** на производительность модели

– формирование гибридной формы на основе комбинации опознанных частотных морфем – *переводез*, *переводский* (Google Translate);

– перевод близкой и более частотной лексемой или частотным словосочетанием с этой лексемой – *переводческий язык*, *язык перевода*, *переводной язык*, *переводной текст* (Google Translate, Yandex Translate);

– перевод словосочетаний с неопознанной лексемой с неоправданным расширяющим толкованием, обусловленным контекстом;

– без перевода.

При этом сопоставление результатов перевода двух систем позволило определить различные подходы к интерпретации лексем, отсутствующих в словаре системы, но не выявило существенных преимуществ: обе системы оставляют решение этой проблемы эксперту-переводчику. При этом главная проблема переводчика – распознавание ошибок в результатах системы МП, которые (кроме случая с гибридными формами) не демонстрируют внешних признаков некорректности.

Оптимальным решением этой проблемы представляется ведение пользовательского словаря, формируемого в виде традиционного набора словарных статей, включающих узлы-понятия и типы связей, а также фиксирующих переводы новой лексемы и ее контексты для обеспечения единства перевода в рамках текста или группы текстов.

СПИСОК ЛИТЕРАТУРЫ

Беляева Л. Н., 2016. Лингвистические технологии в современном сетевом пространстве: language worker в индустрии локализации. СПб. : Кн. дом. 134 с.

Беляева Л. Н., 2022. Машинный перевод в современной технологии процесса перевода // Известия РГПУ им. А.И. Герцена. № 203. С. 22–30.

Беляева Л. Н., Камшилова О. Н., Шубина Н. Л., 2023. Научная статья в технологическом пространстве машинного перевода: правила и процедуры редактирования : учеб. пособие. СПб. : Кн. дом. 90 с.

Нуриев В. А., 2019. Архитектура системы нейронного машинного перевода // Информатика и ее применения. Т. 13, № 3. С. 90–96. DOI: https://doi.org/10.14357/19922264190313

Раренко М. Б., 2021. Машинный перевод: от перевода «по правилам» к нейронному переводу (Обзор) // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6, Языкознание : РЖ. № 3. С. 70–79. DOI: https://doi.org/10.31249/ling/2021.03.05

Almansoori A., Al Mansoori S., Alshamsi M., Salloum S. A., Shaalan K., 2020. Development of Machine Translation Models: A Systematic Review // International Journal of Control and Automation. Vol. 13, № 2. P. 1462–1483.

Araabi A., Monz C., Niculae V., 2022. How Effective is Byte Pair Encoding for Out-Of-Vocabulary Words in Neural Machine Translation? URL: https://arxiv.org/abs/2208.05225v1

Brottrager J., Stahl A., Arslan A., Brandes U., Weitin T., 2022. Modeling and Predicting Literary Reception // Journal of Computational Literary Studies. Vol. 1, iss. 1. P. 1–27. DOI: 10.26083/tuprints-00023250

Dankers V., Bruni E., Hupkes D., 2022. The Paradox of the Compositionality of Natural Language: A Neural Machine Translation Case Study // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Vol. 1. Long Papers. P. 4154–4175. DOI: https://doi.org/10.48550/arXiv.2108.05885

Devlin J., Chang M.-W., Lee K., Toutanova K., 2019. Pre-Training of Deep Bidirectional Transformers for Language Understanding // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Vol. 1. Long and Short Papers. P. 4171–4186. DOI: https://doi.org/10.18653/v1/N19-1423

Khoong E. C., Rodriguez J. A., 2022. A Research Agenda for Using Machine Translation in Clinical Medicine // Journal of General Internal Medicine. Vol. 37, iss. 5. P. 1275–1277. DOI: 10.1007/ s11606-021-07164- y

Lankford S., Afli H., Way A., 2021. Transformers for Low-Resource Languages: Is Feridir Linn! // Proceedings of the 18th Biennial Machine Translation Summit Virtual USA, August 16–20. Vol. 1. MT Research Track. P. 48–61. DOI: https://doi.org/10.48550/arXiv.2403.01985

Liu X., Sun T., He J., Wu J., Wu L., Zhang X., Jiang H., Cao Z., Huang X., Qiu X., 2022. Towards Efficient NLP: A Standard Evaluation and a Strong Baseline // Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle : Association for Computational Linguistics. P. 3288–3303.

Peris Á., Casacuberta F., 2019. Online Learning for Effort Reduction in Interactive Neural Machine Translation // Computer Speech & Language.

Vol. 58. P. 98–126. DOI: https://doi.org/10.48550/arXiv.1802.03594

Popović M., 2017. chrF++: Words Helping Character n-Grams // Proceedings of the Second Conference on Machine Translation. Copenhagen : [s. n.]. P. 612–618.

Sennrich R., Haddow B., Birch A., 2015. Neural Machine Translation of Rare Words with Subword Units. arXiv:1508.07909v5 [cs.CL]. DOI: https://doi.org/10.48550/arXiv.1508.07909

Tars M., Tättar A., Fišel M., 2022. Cross-Lingual Transfer From Large Multilingual Translation Models to Unseen Under-Resourced Languages // Baltic Journal of Modern Computing. Vol. 10, iss. 3. P. 435–446. DOI: https://doi.org/10.22364/bjmc.2022.10.3.16

Toral A., 2019. Post-Editese: An Exacerbated Translationese // Proceedings of Machine Translation Summit XVII. Vol. 1. Research Track. Dublin : European Association for Machine Translation. P. 273–281.

Zhu C., Yu H., Cheng Sh., Luo W., 2020. Language-Aware Interlingua for Multi-Lingual Neural Machine Translation // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroutsburg : Association for Computational Linguistics. P. 1650–1655.

Zhuang F., Qi Z, Duan K., Xi D., Zhu Y., Zhu H., Xiong H., He Q., 2021. A Comprehensive Survey on Transfer Learning // Proceedings of the IEEE. Vol. 109, iss. 1. P. 43–76. DOI: 10.1109/JPROC.2020.3004555

REFERENCES

Belyaeva L.N., 2016. *Lingvisticheskiye tekhnologii v sovremennom setevom prostranstve: language worker v industrii lokalizatsii* [Linguistic Technologies in the Modern Network Space: Language Worker in the Localization Industry]. Saint Petersburg, Kn. dom Publ. 134 p.

Belyaeva L.N., 2022. Mashinnyy perevod v sovremennoy tekhnologii protsessa perevoda [Machine Translation in Modern Translation Technology]. *Izvestiya RGPU im. A.I. Gercena* [Izvestia: Herzen University Journal of Humanities & Sciences)], no. 203, pp. 22-30.

Belyaeva L.N., Kamshilova O.N., Shubina N.L., 2023. *Nauchnaya statya v tekhnologicheskom prostranstve mashinnogo perevoda: pravila i procedury redaktirovaniya: ucheb. posobie* [Scientific Article in the Technological Space of Machine Translation: Editing Rules and Procedures. Textbook]. Saint Petersburg, Kn. dom Publ. 90 p.

Nuriev V.A., 2019. Arkhitektura sistemy neyronnogo mashinnogo perevoda [Architecture of a Machine Translation System]. *Informatika i ee primeneniya* [Informatics and Applications], vol. 13, no. 3, pp. 90-96. DOI: https://doi.org/10.14357/19922264190313

Rarenko M.B., 2021. Mashinnyy perevod: ot perevoda «po pravilam» k neyronnomu perevodu (Obzor) [Machine Translation: From Translation "According to the Rules" to Neural Translation (Review)]. *Sotsialnye i gumanitarnye nauki. Otechestvennaya i zarubezhnaya literatura. Seriya 6. Yazykoznanie: RZh* [Social Sciences and Humanities. Domestic and Foreign Literature. Series 6. Linguistics. Abstract Journal. INION RAN], no. 3, pp. 70-79. DOI: https://doi.org/10.31249/ling/2021.03.05

Almansoori A., Al Mansoori S., Alshamsi M., Salloum S.A., Shaalan K., 2020. Development of Machine Translation Models: A Systematic Review. *International Journal of Control and Automation*, vol. 13, no. 2, pp. 1462-1483.

Araabi A., Monz C., Niculae V., 2022. *How Effective Is Byte Pair Encoding for Out-Of-Vocabulary Words in Neural Machine Translation?* URL: https://arxiv.org/abs/2208.05225v1

Brottrager J., Stahl A., Arslan A., Brandes U., Weitin T., 2022. Modeling and Predicting Literary Reception. *Journal of Computational Literary Studies*, vol. 1, iss. 1, pp. 1-27. DOI: 10.26083/tuprints-00023250

Dankers V., Bruni E., Hupkes D., 2022. The Paradox of the Compositionality of Natural Language: A Neural Machine Translation Case Study. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Vol. 1: Long Papers*, pp. 4154-4175. DOI: https://doi.org/10.48550/arXiv.2108.05885

Devlin J., Chang M.-W., Lee K., Toutanova K., 2019. Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Vol. 1. Long and Short Papers*, pp. 4171-4186. DOI: https://doi.org/10.18653/v1/N19-1423

Khoong E.C., Rodriguez J.A., 2022. A Research Agenda for Using Machine Translation in Clinical Medicine. *Journal of General Internal Medicine*, vol. 37, iss. 5, pp. 1275-1277. DOI: 10.1007/s11606-021-07164-y

Lankford S., Afli H., Way A., 2021. Transformers for Low-Resource Languages: Is Feridir Linn! *Proceedings of the 18th Biennial Machine Translation Summit Virtual USA, August 16–20. Vol. 1*. MT Research Track, pp. 48-61. DOI: https://doi.org/10.48550/arXiv.2403.01985

Liu X., Sun T., He J., Wu J., Wu L., Zhang X., Jiang H., Cao Z., Huang X., Qiu X., 2022. Towards Efficient NLP: A Standard Evaluation and a Strong Baseline. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle, Association for Computational Linguistics, pp. 3288-3303.

Peris Á., Casacuberta F., 2019. Online Learning for Effort Reduction in Interactive Neural Machine Translation. *Computer Speech & Language*, vol. 58, pp. 98-126. DOI: https://doi.org/10.48550/arXiv.1802.03594

Popović M., 2017. chrF++: Words Helping Character n-Grams. *Proceedings of the Second Conference on Machine Translation*. Copenhagen, s.n., pp. 612-618.

Sennrich R., Haddow B., Birch A., 2015. *Neural Machine Translation of Rare Words with Subword Units*. arXiv:1508.07909v5 [cs.CL]. DOI: https://doi.org/10.48550/arXiv.1508.07909

Tars M., Tättar A., Fišel M., 2022. Cross-Lingual Transfer from Large Multilingual Translation Models to Unseen Under-Resourced Languages. *Baltic Journal of Modern Computing*, vol. 10, iss. 3, pp. 435-446. DOI: https://doi.org/10.22364/bjmc.2022.10.3.16

Toral A., 2019. Post-Editese: An Exacerbated Translationese. *Proceedings of Machine Translation Summit XVII. Vol. 1. Research Track*. Dublin, European Association for Machine Translation, pp. 273-281.

Zhu C., Yu H., Cheng Sh., Luo W., 2020. Language-Aware Interlingua for Multi-Lingual Neural Machine Translation. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Stroutsburg, Association for Computational Linguistics, pp. 1650-1655.

Zhuang F., Qi Z, Duan K., Xi D., Zhu Y., Zhu H., Xiong H., He Q., 2021. A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, vol. 109, iss. 1, pp. 43-76. DOI: 10.1109/JPROC.2020.3004555

Information About the Authors

Larisa N. Beliaeva, Doctor of Sciences (Philology), Professor, Department of Educational Technologies in Philology, Herzen State Pedagogical University of Russia, Reki Moiki Emb., 48, 191186 Saint Petersburg, Russia, lauranbel@gmail.com, https://orcid.org/0000-0002-8622-4595

Olga N. Kamshilova, Candidate of Sciences (Philology), Associate Professor, Department of Educational Technologies in Philology, Herzen State Pedagogical University of Russia, Reki Moiki Emb., 48, 191186 Saint Petersburg, Russia; Associate Professor, Department of Linguistics and Translation Studies, Saint Petersburg University of Management Technologies and Economics, Prosp. Lermontovsky, 44, 190020 Saint Petersburg, Russia, onkamshilova@gmail.com, https://orcid.org/0000-0002-1488-2206

Информация об авторах

Лариса Николаевна Беляева, доктор филологических наук, профессор кафедры образовательных технологий в филологии, Российский государственный педагогический университет им. А.И. Герцена, наб. реки Мойки, 48, 191186 г. Санкт-Петербург, Россия, lauranbel@gmail.com, https://orcid.org/0000-0002-8622-4595

Ольга Николаевна Камшилова, кандидат филологических наук, доцент кафедры образовательных технологий в филологии, Российский государственный педагогический университет им. А.И. Герцена, наб. реки Мойки, 48, 191186 г. Санкт-Петербург, Россия; доцент кафедры лингвистики и переводоведения, Санкт-Петербургский университет технологий управления и экономики, просп. Лермонтовский, 44, 190020 г. Санкт-Петербург, Россия, onkamshilova@gmail.com, https://orcid.org/0000-0002-1488-2206