

Дина Викторовна Володина

Сибирский государственный университет путей сообщения, кандидат филологических наук, доцент,
доцент кафедры «Иностранные языки», Новосибирск, Россия
e-mail: dina.volodina@mail.ru

Юлия Сергеевна Юрьева

Сибирский государственный университет путей сообщения, кандидат педагогических наук, доцент,
доцент кафедры «Иностранные языки», Новосибирск, Россия
e-mail: jurjewaju@yandex.ru

**Возможности использования больших языковых моделей
при выполнении перевода студентами технического вуза**

Аннотация. Статья посвящена эволюции и возможностям применения искусственного интеллекта. Особое внимание уделено понятию больших языковых моделей и сферам их употребления. Описаны преимущества внедрения нейронных сетей в образование. Изучены возможности применения больших языковых моделей как вспомогательного инструмента обучения дисциплине «Иностранный язык» студентов технических вузов. В практической части исследования были проведены анкетирование и устный опрос обучающихся для выявления наиболее востребованного онлайн-сервиса машинного перевода. Приведены результаты опроса среди студентов технического вуза и интернет-пользователей. Проанализирован опыт применения машинного перевода при выполнении заданий учебной, научной и профессиональной (переводческой) деятельности.

Ключевые слова: искусственный интеллект, большие языковые модели, иностранный язык, машинный перевод, технический вуз, нейронные сети, функциональность.

Dina V. Volodina

Siberian Transport University, Candidate of Philological Sciences, Associate Professor,
Associate Professor of "Foreign Languages" Department, Novosibirsk, Russia
e-mail: dina.volodina@mail.ru

Yulia S. Yurieva

Siberian Transport University, Candidate of Pedagogical Sciences, Associate Professor,
Associate Professor of "Foreign Languages" Department, Novosibirsk, Russia
e-mail: jurjewaju@yandex.ru

**Possibilities of Using Large Language Models
in Translation by Students of a Technical University**

Abstract. The article is devoted to the evolution and possibilities of artificial intelligence application. Special attention is paid to the concept of large language models and spheres of their use. The advantages of implementing neural networks in education are described. The possibilities of application of large language models as an auxiliary tool for teaching the discipline 'Foreign Language' to students of technical universities are studied. For the practical part of the study, questionnaires and oral survey of students were conducted to identify the most demanded online service of machine translation. The results of the survey among technical university students and Internet users are presented. The experience of using machine translation in the performance of tasks of educational, scientific and professional (translation) activity is analysed.

Keywords: artificial intelligence, large language models, foreign language, machine translation, technical university, neural networks, functionality.

Введение (Introduction)

Современные цифровые технологии активно внедряются во все сферы деятельности человека, совершенствуя разнообразные процессы. Особого внимания заслуживает развитие искусственного интеллекта (ИИ) и его широкая сфера применения, начиная от использования в приборах бытовой техники до имитации работы человеческого интеллекта. Исследования по развитию ИИ направлены на изучение его способностей мыслить, рассуждать, принимать решения, т. е. тех способностей, которые присущи человеку [1].

Безусловно, с момента создания в 1950 г. до настоящего времени искусственный интеллект значительно эволюционировал: расширились как сферы его употребления, так и приросли его функции — от способности мыслить подобно человеку до потенциальной замены когнитивных и оперативных способностей человека [2]. Кроме того, изменения коснулись и самого понятия: на смену термину «машинный интеллект», предложенному А. Тьюрингом, ученый в сфере информационных технологий Джон Макартни ввел в употребление в 1961 г. название «искусственный интеллект».

Далее первый промышленный робот с ИИ заменил человека на конвейерной ленте. Позднее Джозефом Вайзенбаумом был изобретен чат-бот для общения с человеком: он представлял собой простую программу, которая использовала распознавание образов для имитации человеческого разговора, превращая вводимые пользователем данные в вопрос и генерируя ответ на основе набора заранее определенных правил. Создание такого робота положило начало исследованиям в области обработки естественного языка и разработке более сложных больших языковых моделей (БЯМ) [3].

Внедрение в 1997 г. сетей долгой и краткосрочной памяти (LSTM) позволило создать более глубокие и сложные нейронные сети, способные обрабатывать более значительные объемы данных. А в конце 1990-х гг. Синтия Бризил создала робота Кисмет, способного взаимодействовать с людьми и моделировать человеческие эмоции и мимику [4]. С 2002 г. появились виртуальные помощники, опции которых расширились.

В 2011 г. был запущен Google Brain, предоставивший исследователям доступ к мощным вычислительным ресурсам и наборам данных, а также расширенные функции (например, встраивание слов), позволяющие системам нейролингвистического программирования (НЛП) лучше понимать контекст. Google Brain включает в себя машинное обучение, использующее вычислительные ресурсы Google. Кроме того, следует отметить значительные достижения данного проекта в сфере машинного перевода: был запущен нейронный машинный переводчик Google (GNMT), что значительно повысило качество перевода, позднее они разработали многоязыковую версию. Создание трансформера в 2017 г. помогло обрабатывать последовательности в тексте и реферировать его, что, в свою очередь, позволило создавать более крупные и сложные БЯМ, такие как GPT-3 от OpenAI (Generative Pre-Trained Transformer), которые послужили основой для ChatGPT и множества других невероятных приложений, управляемых искусственным интеллектом [5].

Методы (Methods)

Сегодня большие языковые модели (БЯМ) имеют хорошие показатели в обработке естественного языка. БЯМ определяется [6] как тип искусственного интеллекта, который имитирует работу интеллекта человека, используя статистические модели и методы обучения с целью обработки больших текстов.

Соответственно, БЯМ могут использоваться во время беседы, где они генерируют ответы, которые не только органично встраиваются в контекст, но и поддерживают согласованность на протяжении всего диалога. Широкое распространение чат-ботов и виртуальных помощников свидетельствует о растущем спросе на БЯМ в разных сферах деятельности.

В нашем исследовании будет предпринята попытка рассмотреть БЯМ как вспомогательный инструмент обучения студентов технических вузов. Образование как одна из наиболее востребованных на данный момент сфер применения современных технологий, включая разработки в сфере ИИ, предлагает достаточно широкий спектр опций или операций, придающих особую привлекательность использованию нейронных сетей в процессе обучения. БЯМ, как одно из достижений в сфере ИИ, позволяет анализировать и отслеживать прогресс обучающегося, корректировать его продвижение, предлагать контент, специально сгенерированный под потребности и способности каждого человека. Здесь важно отметить изменение отношений между участниками образовательного процесса: обучающиеся в гораздо меньшей степени зависимы от административных структур, устанавливающих стандарты обучения; от образовательных учреждений, реализующих на практике эти стандарты и предоставляющих возможности для обучения; от родителей, создающих условия для получения образования и осуществляющих контроль за успеваемостью детей. Что касается учителей/преподавателей, то они, помимо трансляции информации и организации учебного процесса, имеют возможность полноценного сотрудничества с обучающимися в освоении и использовании возможностей БЯМ.

Итак, согласно определению, использование БЯМ для обучающих целей позволяет моделировать и реализовывать ситуации реального общения (устного и письменного). Под моделированием устных форм общения понимается создание диалоговых систем. Письменные же охватывают генерацию текстового контента: от научной статьи и профессионального текста по определенной теме до написания эссе.

Наряду с вышеперечисленными сферами использования БЯМ, необходимо выделить и машинный перевод (МП). Благодаря их совокупности сложные наборы лингвистических правил, требуемых для машинного перевода, стали более эффективными и точными, а доля участия человека сократилась.

Согласно поставленной цели данного исследования, в фокусе находятся возможности БЯМ при осуществлении перевода текстов. Поэтому особый интерес представляет собой прикладной аспект БЯМ и один из способов его реализации — машинный перевод, а также определение роли последнего в образовательном процессе технического

вуза. Выбор МП обусловлен несколькими причинами. Во-первых, перевод, наряду с чтением, говорением, аудированием и письмом, является обязательным видом речевой деятельности при изучении иностранного языка на всех этапах обучения. Во-вторых, изучение иностранного языка сопряжено с поиском и переработкой большого количества аутентичной информации, в том числе профессионально ориентированной.

Результаты и обсуждение (Results and Discussion)

В современном информационном пространстве существует ряд многоязычных сервисов автоматического перевода, так называемые онлайн-переводчики, использование которых давно вошло в жизнь практически каждого человека настолько прочно, что не все подразумевают, что они созданы на основе БЯМ.

Под машинным переводом понимается выполняемое на компьютере с выходом в интернет действие по преобразованию текста на одном естественном языке в эквивалентный по содержанию текст на другом языке, а также результат такого действия [7]. В связи со стремительной технологизацией и цифровизацией современного общества авторам представляется возможным расширить определение и к компьютеру как средству, осуществляющему доступ к сервисам машинного перевода, добавить такие устройства, как смартфон или планшет, поддерживающие доступ в один из мобильных браузеров: Google Chrome, Safari, Opera, Mozilla Firefox, Microsoft Edge.

Первые упоминания о машинном переводе датируются 1930 г. и связаны с именами Жоржа Арцруни и Петра Троянского [8]. Следующий шаг предпринят в 1949 г. Уореном Уивером в эру компьютеров, где машинный перевод был основан на базе декодирования [9]. Первый компьютерный перевод сделан в 1954 г. в Нью-Йорке IBM-701, где были переведены 60 предложений с русского на английский язык [10]. В эти периоды МП характеризовался пословным переводом. Позднее прямой подход в 1980 г. был изменен на косвенный, т. е. машина анализирует синтаксис исходного языка, упрощает его и создает текст на целевом языке. Данный подход и получил свое развитие в будущем. В 2000 г. принцип работы машинного переводчика заключался в работе не со словом, а с фразой. Прорывом же в сфере машинного перевода стало появление нейронного машинного перевода в середине 2010 г. При данном подходе к переводу слова определяются/переводятся в соответствии с контекстом. Именно данный принцип и используется в современных переводчиках сегодня.

Не вызывает сомнения, что сервисами МП пользуется значительная часть студентов. Для выявления наиболее популярных были опрошены 100 студентов, обучающихся по направлению подготовки 09.03.02 «Информационные системы на транспорте». Выбор обучающихся данного направления обусловлен спецификой их будущей профессиональной деятельности. Кроме того, большая часть специальной литературы по программированию написана на английском языке, соответственно, необходим адекватный перевод материала.

Чаще всего студенты прибегают к помощи сервисов онлайн-переводчиков «Гугл Транслейт» (Google Translate)

(56 %), а 31 % обучающихся переводят текстовую информацию, используя «Яндекс Транслейт» (Yandex Translate). В последнее время для перевода профессионально ориентированной информации респонденты стали обращаться к сервису DeepL (9 %). Одинаковое процентное соотношение (2 %) обнаружили Prompt и Bing.

Далее, исходя из опыта пользования сервисами онлайн-перевода, студентам было предложено оценить по 5-балльной шкале функциональность каждого из них. Кроме того, было принято решение привлечь к опросу интернет-пользователей (97 человек).

Здесь стоит пояснить, что под функциональностью сервисов онлайн-перевода понимается соблюдение нескольких параметров, а именно: количество переводимых знаков; точность перевода; скорость перевода; удобство и простота обслуживания; количество поддерживаемых языков. Результаты представлены на рисунке.

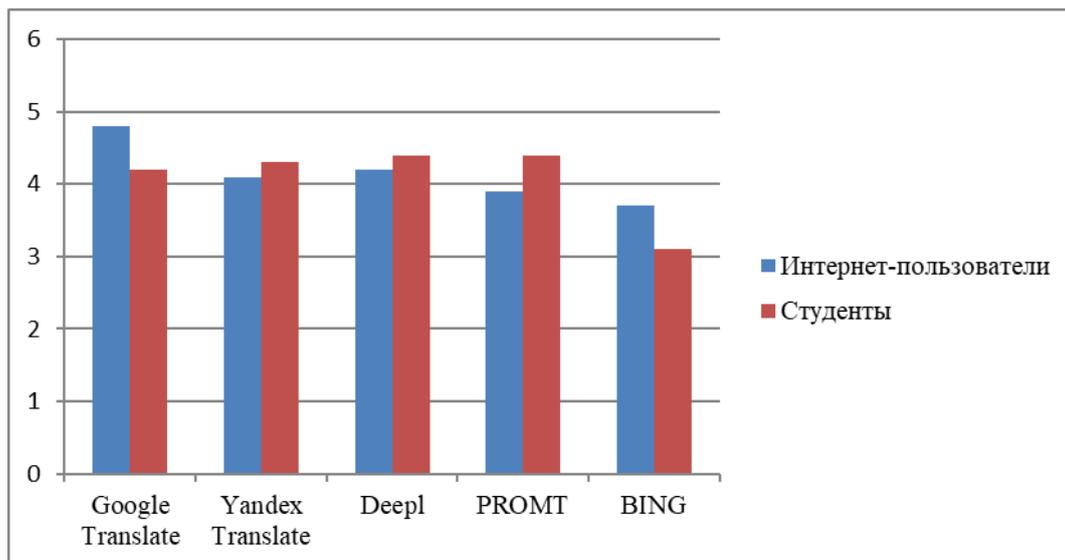
Проведенный опрос позволяет наблюдать расхождение мнений среди двух групп респондентов: абсолютным лидером по функциональности интернет-респонденты назвали «Гугл Транслейт» (Google Translate) (4,8 балла), в то время как у студентов данная программа машинного перевода набрала всего 4,2 балла. Студенты отдают предпочтение DeepL и Prompt (4,4 балла). Такое расхождение может быть связано со сферой применения сервисов для онлайн-переводов: студенты используют их для перевода академической и научной литературы, а интернет-пользователи — для деловой и профессиональной коммуникации.

Рассматривая интерпретацию результатов представленного выше опроса, стоит подчеркнуть, что анкетированы были в основном респонденты, изучающие и использующие в своей профессиональной деятельности английский язык.

Опыт работы с переводами с немецкого языка на русский, осуществленными с помощью онлайн-переводчиков, позволяет сделать следующее обобщение: сложные грамматические конструкции, специфические исключительно для немецкого языка, не всегда точно передаются при переводе; сложносоставные слова, состоящие из трех и более основ, также характерные для немецкого языка, при машинном переводе могут потерять отдельные семантические нюансы. Это может также объясняться тем фактом, что онлайн-переводчики, хотя и имеют в своих настройках несколько переводимых языков, изначально были ориентированы именно на английский язык.

Результаты опроса и индивидуальных бесед со студентами позволили прийти к следующим умозаключениям: выбираются те онлайн-переводчики, которые решают проблему передачи информации с одного языка (переводимого) на другой (переводящий) с учетом именно смысловой нагрузки. То есть для этой группы пользователей не важен чисто лингвистический (языковой) аспект перевода, включая грамматические, лексические и стилистические особенности переводимого текста, для них любой язык (родной или иностранный) является средством передачи сути повествования. Это имеет место, когда основной целью является сам перевод и, соответственно, передача смысла.

Если же приоритетными являются способы передачи того или иного грамматического явления либо семантики



Функциональность различных сервисов машинного перевода

слова, иными словами, лингвистические аспекты перевода, то существует ряд нюансов, которые необходимо учитывать при использовании сервисов для онлайн-перевода.

Авторами предлагается рассмотреть некоторые положения касательно сервисов онлайн-перевода в процессе обучения дисциплине «Иностранный язык» в техническом вузе: собственно, аудиторные занятия, научно-исследовательская деятельность студентов (НИРС) и программа переподготовки «Переводчик в сфере профессиональной коммуникации».

Актуальное состояние использования машинного перевода было описано выше. Здесь необходимо выделить следующее: если перевод текстовой информации не является предметом контроля в обучении, а лишь промежуточной операцией (например, перевод текста с последующим использованием информации из него в разработке проектных заданий), то применение машинного перевода как результата деятельности БЯМ является допустимым, но все же подлежит процедуре редактирования.

Спорным же является целесообразность использования онлайн-переводчиков в ситуациях учебного процесса, когда перевод — как вид учебной деятельности — является предметом оценивания и контроля.

Кафедра «Иностранные языки» Сибирского государственного университета путей сообщения (СГУПС) уделяет большое внимание НИРС: более 40 лет проводит олимпиаду «Письменный перевод», в которой участникам предлагается перевести текст объемом 1400 знаков в течение 60 мин. с иностранного (английского, немецкого, французского) на русский язык, а также небольшую рекламу по теме, где студенты могут проявить свои творческие способности в данном стиле. Текст относится к публицистическому стилю и отражает основные тенденции в развитии современного общества. Изначально перевод осуществлялся при помощи бумажного словаря. С развитием информационных технологий допустимым стало использование электронных словарей, например Abby Lingvo. В период пандемии было принято решение пере-

вести олимпиаду в дистанционный (онлайн-) формат, что привело к определенным последствиям: во-первых, значительно расширилась география, так как из городского мероприятия стало всероссийским (с международным участием); во-вторых, дистанционная форма проведения олимпиады позволяет участникам использовать сервисы машинного перевода, причем ни контроль за объемом их использования, ни полный запрет на них не представляется возможным. Абсолютно все участники олимпиады в разном объеме прибегают к помощи онлайн-переводчиков.

Оценивание конкурсных работ осуществляется на основании «эталонного перевода», выполненного экспертной комиссией, состоящей из профессорско-преподавательского состава кафедры «Иностранные языки» и приглашенных преподавателей из вузов-партнеров олимпиады, и при соблюдении следующих критериев: адекватная передача лексического значения слова; выявление грамматических явлений и передача их на переводимый язык; соблюдение стилистического единства текста; соответствие структур предложения нормам русского языка. Несомненно, онлайн-переводчики нового поколения не только различают грамматические структуры, термины, но и контекст. Тем не менее присутствие машинного перевода в олимпиадных заданиях всегда различимо. Опытным путем профессорско-преподавательскому составу кафедры представилось возможным выявить некоторые особенности использования МП при выполнении заданий олимпиады, а именно:

1. Несмотря на то что переводимый текст не является научным или профессионально ориентированным, а относится к публицистическому стилю, предполагающему некоторую свободу изложения, перевод не всегда представляет собой стилистическое единство и зачастую превышает все границы «свободы» даже в рамках публицистики.

2. Основной сложностью при переводе олимпиадных заданий является перевод лексических единиц и словосочетаний, не имеющих эквивалентов в русском языке и/или не

зафиксированных в существующих электронных словарях. Онлайн-сервисы, скорее всего, предложат наиболее частотный вариант перевода по причине того, что сам переводчик основан на статистическом подходе.

3. Олимпиадные задания составляются таким образом, чтобы в одном тексте были по возможности представлены грамматические явления в различных своих функциях, соответственно, оценивается способность адекватно передать это явление с учетом особенностей переводящего языка.

4. Отдельного внимания заслуживает задание перевода рекламного текста. Несмотря на небольшой объем (до 90 слов), именно оно представляет собой наибольшую трудность при переводе: здесь необходимо соблюдение стилистики переводимой информации, наличие фоновых знаний, владение контекстом и достаточно высокий уровень русского языка. На данный момент выполнение всех этих требований одновременно практически неосуществимо программами машинного перевода.

При переводе профессионально ориентированной литературы современные онлайн-переводчики различают термины и контекст, однако и в данном виде деятельности существует ряд замечаний, которые подлежат вниманию и должны быть приняты в рассмотрение.

С 2009 г. кафедра «Иностранные языки» осуществляет подготовку слушателей по программе переподготовки «Переводчик в сфере профессиональной коммуникации» [11], позволяющую им вести переводческую деятельность по специальности/направлению подготовки основного образования. Во время обучения слушатели изучают основы стилистики, лексикологии, теоретической грамматики, которые составляют базу для владения грамматическими и лексическими трансформациями.

Составляющей итоговой аттестации слушателей является перевод текста по специальности/направлению подготовки основного образования объемом до 25 тыс. знаков. Слушателями осуществляются предпереводческий и переводческий анализ текста, а также его редактирование. Ключевым моментом при переводе является предпереводческий анализ текста с определением реципиента, стиля и типа текста, а также коммуникативного задания. Именно выполнение предпереводческого анализа определяет стратегию перевода. Переводческий анализ предполагает выявление лексических и грамматических трансформаций и трудностей на лексическом, грамматическом и прагматическом уровнях.

Так как переводимый слушателями текст имеет достаточный большой объем, не все сервисы онлайн-перевода могут обработать его как целый текст, а только фрагментами. В данном случае не соблюдается ни стилистическое единство, ни унификация терминологического аппарата. Таким образом затрудняется процесс выявления трансформаций.

Кроме того, текст, переведенный с помощью сервисов машинного перевода, не всегда учитывает его коммуникативное задание и реципиентов. Соответственно, на этапе редактирования текста искусственным путем создаются избыточные трудности.

Заключение (Conclusion)

Очевидно, что использование МП и процесс его внедрения как одного из способов реализации БЯМ в образовательный процесс невозможно остановить. Поэтому нам представляется возможным обратить внимание на некоторые значимые моменты:

1. Цель перевода определяет инструменты его реализации (электронный словарь или онлайн-сервисы автоматического перевода). Если нужно быстро передать общий смысл для последующего его использования в других видах деятельности, то машинный перевод вполне оправдан. В случае определения и выявления языковых явлений и особенностей переводимого текста машинный перевод скорее вызовет больше затруднений, чем облегчит работу.

2. При условии, что машинного перевода не избежать, существует потребность разработать задания/тренажеры для практики фрагментарного перевода.

Резюмируя положения о возможностях применения машинного перевода как одного из способов реализации БЯМ в образовательном процессе, рассмотренные в данной работе, авторы пришли к следующим выводам: онлайн-сервисы, преобразующие текстовую информацию на основе нейросетей с одного языка на другой, не могут уже восприниматься пользователями в гипотетической плоскости: «Что будет, если машинный перевод заменит традиционный?» Онлайн-переводчики сейчас представляют собой удобный инструмент для преобразования информации, и их дальнейшее применение, в том числе и в образовательном процессе, является аксиомой. Соответственно, существует необходимость разработки методического инструментария, позволяющего наиболее эффективно применять онлайн-переводчики в качестве вспомогательного средства осуществления перевода.

Библиографический список

1. Искусственный интеллект как ключевой фактор цифровизации глобальной экономики : исследование, проведенное консультантами J'son & Partners. 2017. URL: <https://www.crn.ru/news/detail.php?ID=117544> (дата обращения: 30.05.2024).
2. Голенков С. И. Сознание как испытание бытия/бытием, или Почему невозможно создать искусственный интеллект // Философские проблемы информационных технологий и киберпространства. 2011. № 2. С. 75–83.
3. Basset C. The Computational Therapeutic: Exploring Weizenbaum's Eliza as a History of the Present // AI & Society. 2019. Vol. 34. P. 803–812. DOI: 10.1007/s00146-018-0825-9
4. Cheshire T. How Cynthia Breazeal is Teaching Robots How to Be Human // Science. 2011. URL: <https://www.wired.com/story/best-friendbots/> (дата обращения: 30.05.2024).
5. Large Language Models 101: History, Evolution and Future // ScribbleData : [site]. URL: <https://www.scribbledata.io/blog/large-language-models-history-evolutions-and-future/> (дата обращения: 23.05.2024).

6. From Human Writing to Artificial Intelligence Generated Text: Examining the Prospects and Potential Threats of ChatGpt in Academic Writing / I. Dergaa, K. Chamari, P. Zmijewski, H. B. Saad // *Biology of Sport*. 2023. Vol. 40, no. 2. P. 615–622. DOI: 10.5114/biolSport.2023.125623

7. Скворцова О. В., Тихонова Е. В. Проблемы и преимущества автоматизированного и машинного переводов // *Молодой ученый*. 2016. № 9. С. 1287–1289.

8. Hutchins J. Two Precursors of Machine Translation: Artsrouni and Trojanskij // *International Journal of Translation*. 2004. Vol. 16 (1). P. 11–31. URL: <https://mt-archive.net/00/IJT-2004-Hutchins.pdf> (дата обращения: 01.05.2024).

9. Weaver W. Memorandum on Translation. 1949. URL: <https://mt-archive.net/Weaver-1949.pdf> (дата обращения: 13.04.2024).

10. История машинного перевода: от гипотез Лейбница и Декарта — до мобильных приложений и облачных сервисов // *Prompt* : [сайт]. 2019. 21 марта. URL: <https://www.prompt.ru/press/blog/istoriya-mashinnogo-perevoda-ot-gipotez-leybnitsa-i-dekarta-do-mobilnykh-prilozheniy-i-oblachnykh-se/> (дата обращения: 01.05.2024).

11. Володина Д. В., Юрьева Ю. С. Особенности подготовки инженеров для осуществления переводческой деятельности (на примере программы переподготовки «Переводчик в сфере профессиональной коммуникации») // *Иноязычное образование в современных реалиях: проблемы и векторы развития : материалы Междунар. науч.-практ. конф., посвящ. 55-летию Казан. гос. энергет. ун-та. Казань : [б. и.], 2023. С. 445–449.*