**УДК 81'25; 81'33** https://doi.org/10.22250/24107190-2024-10-3-131

**Парина Ирина Сергеевна[1]✉, Салаев Павел Владимирович[2]**
**[1]Национальный исследовательский Нижегородский государственный университет им. Н. И. Лобачевского, [2]Гарда Технологии, г. Нижний Новгород, Российская Федерация**
parina@fsn.unn.ru

# Чат-бот для создания двуязычных глоссариев в паре языков английский-русский для обучающихся переводу

**Аннотация**
В статье предлагается сервис для автоматизации составления двуязычных глоссариев студентами переводческих специальностей. Представлена тестовая версия разработанного нами сервиса в формате чат-бота, рассмотрены результаты его оценивания студентами и преподавателями переводческого факультета лингвистического вуза, произведены доработки по результатам тестирования. Анализ литературы по теме показал, что в настоящее время существуют программы для составления двуязычных глоссариев или для извлечения терминов из текста, однако они либо дают неудовлетворительный результат для русского языка, либо выполняют свои функции без ограничений только в платной версии. Целью проекта по созданию сервиса было облегчить работу с текстами и лексикографическими источниками, сохранив обучающую составляющую при подготовке глоссария так, чтобы студенты самостоятельно принимали решения при подборе двуязычных эквивалентов. В версии сервиса, предложенной для тестирования, пользователь направлял чат-боту файл с текстом и получал в ответном сообщении список ключевых слов (имён нарицательных с фрагментами контекста, а также прецизионной лексики) в виде гиперссылок на страницу двуязычного контекстного словаря или, для имён собственных, на результаты поиска по соответствующему запросу в Интернете. Таким образом, автоматизировался поиск лексических единиц в словаре, однако эквивалент в переводящем языке пользователь выбирал сам на основании текста оригинала и предложенных в контекстном словаре соответствий и контекстов с переводом. По результатам тестирования в формат запроса были внесены изменения: в обновлённой версии пользователи не отправляют файл, а копируют текст в онлайн-форму. Изменён и формат ответа чат-бота: в новой версии это файл с текстом, где ключевые слова преобразованы в гиперссылки. Опрос участников тестирования показал, что студенты рассматривают сервис преимущественно как инструмент для ускорения предварительного анализа лексического состава текста.

**Ключевые слова:** чат-бот, глоссарий, специальный перевод, извлечение терминов



**Для цитирования:** Парина И. С., Салаев П. В. Чат-бот для создания двуязычных глоссариев в паре языков английский-русский для обучающихся переводу // Теоретическая и прикладная лингвистика. 2024. Вып. 10, № 3. С. 131–143. https://doi.org/10.22250/24107190-2024-10-3-131

**Irina S. Parina[1]✉, Pavel V. Salaev[2]**
**[1]National Research Lobachevsky State University of Nizhny Novgorod, [2]Garda Technologies Nizhny Novgorod, Russian Federation**
parina@fsn.unn.ru

# Chatbot for creating bilingual glossaries in the English-Russian language pair for students of translation and interpreting

**Abstract**
In the article, we propose a chatbot for automating the process of compiling bilingual glossaries for students of translation and interpreting. A preliminary literature overview showed that the programs for compiling glossaries

available provide unsatisfactory results for the Russian language or function without restrictions only in the paid version. The project aims at making it easier for students to work with texts and lexicographic sources, and, at the same time, to let them make their own decisions when selecting bilingual equivalents. In the version of the service tested, the user sent a file with the text to the chatbot and received a message in response, containing a list of keywords (common nouns with context, as well as precision words) in the form of hyperlinks to the page of a bilingual contextual dictionary or, in case of proper names, to search engine results for a corresponding query. Thus, the search for lexical units in the dictionary was automated, but the users chose the equivalent in the target language by themselves. After the testing, the request format was updated, so that in the current version the users copy the source text into an online form. The format of the chatbot response was changed as well: in the new version it is a file with the text where the keywords are supplied with hyperlinks. The user survey showed that students view the service primarily as a tool for speeding up the preliminary analysis of the lexical composition of a text.

**Keywords:** chatbot, glossary, special translation, term extraction



**For citation:** Parina, I. S., & Salaev, P. V. (2024). Chat-bot dlya sozdaniya dvuyazychnykh glossariev v pare yazykov angliyskiy-russkiy dlya obuchayushchikhsya perevodu [Chatbot for Creating Bilingual Glossaries in the English-Russian Language Pair for Students of Translation and Interpreting]. *Teoreticheskaya i prikladnaya lingvistika [Theoretical and Applied Linguistics], 10* (3), 131–143. https://doi.org/10.22250/24107190_2024_10_3_131

## 1. Введение [Introduction]

Согласно «Словарю лингвистических терминов» Т. В. Жеребило, термином «глоссарий» обозначают либо составленное на основе конкретного текста собрание непонятных слов или выражений с толкованием или переводом на другой язык, либо словарь малоупотребительных слов с толкованием [Жеребило, 2010, с. 75]. В. В. Дубичинский определяет глоссарий как «словарь глосс (непонятных, устаревших, диалектных слов)», «относительно короткий перечень лексических единиц с минимальной информацией о них» или «краткий словарь специальной лексики» [Дубичинский, 2008, с. 383–384]. Отличительные особенности глоссария состоят, таким образом, в меньшем объёме и в отборе лексики по тем или иным принципам, которые определяются его предназначением. В настоящей статье будут рассматриваться глоссарии, которые составляются на основе конкретных текстов.

Глоссарии играют важную роль в работе как устного, так и письменного переводчика. Так, на необходимость разработки глоссариев при подготовке к синхронному переводу указывал ещё в 1978 году Г. В. Чернов [Чернов, 1978, с. 149]. Глоссарий упоминается в числе предметов, необходимых устному переводчику на мероприятии, в практических рекомендациях, подготовленных Союзом переводчиков России [Синхронный и последовательный перевод, 2015, с. 24]. О положительном влиянии подготовленного заранее глоссария на качество синхронного перевода пишут К. Фантинуоли [Fantinuoli, 2017] и К. Штолль [Stoll, 2009].

Глоссарий рассматривается как важный инструмент обеспечения и повышения качества перевода и в статье В. Ю. Еолян и Э. Д. Муратовой, где речь идёт о письменном переводе научно-технической документации [Еолян, Муратова, 2017, с. 83]. Е. О. Лешканова и А. С. Бубнова, говоря о переводе специальных текстов, отмечают, что глоссарий позволяет переводчику сохранять единство терминологии и избегать ошибок и неточностей [Лешканова, Бубнова, 2018, с. 223].

В настоящее время работа с глоссариями может осуществляться с помощью технических средств. Так, некоторые программы позволяют хранить и систематизировать лексику, которую переводчик вносит через интерфейс самой программы (например, Flashterm [Flashterm]) или импортирует в виде текстового файла (например, Interplex

Glossary Software for Interpreters and Translators [Interplex]). Отбор единиц для глоссария в этом случае осуществляется самим переводчиком в ручном режиме. Как правило, переводческие глоссарии включают в себя лексические единицы, имеющие однозначные соответствия в рамках определённого проекта, текстов определённой тематики или области знаний в целом, – в первую очередь термины.

Существуют приложения для автоматического извлечения терминологии из текста – к бесплатным относятся, в частности, Word Tabulator 2.2.3 [Word Tabulator 2.2.3] и Concordancer for Windows 3.0 [Concordancer for Windows 3.0]. Как показал проведённый нами анализ, текст на русском языке указанные программы обрабатывают некорректно, так что списки извлечённых с их помощью терминов полностью или частично представляют собой нечитаемый набор символов (подробнее результаты описаны в [Парина, Салаев, 2022]). Ещё одна программа для автоматического составления глоссариев – InterpretBank [InterpretBank] – предназначена для самостоятельной подготовки синхронных переводчиков и обладает широким набором функций. Она позволяет извлекать термины из документа, с веб-страницы по заданному адресу, из результатов поиска по заданной теме в Интернете (для этого достаточно указать одно или несколько ключевых слов), а также автоматически подбирать для термина соответствие в переводящем языке. Описывая возможности системы InterpretBank, её создатель К. Фантинуоли [Fantinuoli, 2017] сообщает, что поиск соответствий производится по терминологической базе IATE [IATE] и двуязычным словарям – таким, как Beolingus [Beolingus], LEO [LEO], Dict.cc [Dict.cc]. Кроме того, в систему можно загружать пары текстов в оригинале и переводе и составлять двуязычный глоссарий на их основе. Помимо этого, система снабжена функцией добавления глосс в документ: импортировав в программу текст, можно автоматически «надписать» над всеми словами и словосочетаниями, которые уже зафиксированы в двуязычном глоссарии, их соответствия на переводящем языке. Программа работает с большим количеством языков, однако для русского языка функции автоматического извлечения терминологии из импортированного текста и подбора соответствий на основе загруженных в систему параллельных текстов недоступны. InterpretBank является платной программой, хотя у неё есть и пробная 14-дневная версия.

Обзор систем для автоматизации составления глоссариев для письменного перевода представлен на сайте переводческого агентства Primavista [Девятов, 2018]. Автор обзора положительно отзывается о сервисе Tilde Terminology, который поддерживает русский язык, достаточно точно извлекает из текста термины, в том числе двухсловные, и позволяет в автоматическом режиме подбирать соответствия.

В настоящее время Tilde [Tilde] представляет собой комплекс из систем машинного перевода, автоматического субтитрирования, управления терминологией и других. Доступ к ним предоставляется в платном и бесплатном режиме. В бесплатной версии допускается одновременная работа только с одним проектом, состоящим из файлов общим объемом не более 2 Мбайт.

Положительную оценку в обзоре М. Девятова [Девятов, 2018] получает и система SynchroTerm от канадской компании LogiTerm, которая также поддерживает русский язык [SynchroTerm]. Однако эта программа – самая дорогостоящая из всех рассмотренных. Наиболее эффективной автор обзора называет бесплатную программу Prospector [Prospector] от компании Logrus Global. Но и у неё есть недостаток: программа поддерживает только один язык – английский.

В обзоре [Девятов, 2018] также рассматриваются возможности по извлечению терминологии с помощью CAT-системы MemoQ [MemoQ] в версии MemoQ 8.2 и программы SDL MultiTerm Extract [SDL MultiTerm Extract] от компании-разработчика системы памяти переводов Trados в версии 2017 года. М. Девятов приходит к выводу, что использование этих систем для извлечения терминологии в автоматическом режиме нецелесообразно,

поскольку в сформированные ими списки попадает достаточно большое количество единиц, не являющихся терминами, и не попадают некоторые термины.

Функция извлечения терминов из текста есть и у облачной платформы для организации и выполнения переводов Memsource (в 2021 году объединилась с платформой Phrase) [Phrase]. Однако, как показал анализ, представленный в [Бухаров и др., 2021, с. 68–69], её использование едва ли сократит время работы переводчика над глоссарием: система, фактически, выстраивает в список все словоформы, из которых состоит текст.

Задачу составления глоссария на основе предложенного текста выполняют и основанные на генеративных предобученных моделях чат-боты, возможности которых в последнее время стремительно расширяются. Видеообзор официально недоступного в России, но наиболее известного из подобных чат-ботов – ChatGPT, – представленный профессиональным переводчиком Э. Гамшариком под ником Germling [Germling], позволяет убедиться, что бот быстро составляет двуязычный глоссарий на основе параллельных текстов (в частности, статей многоязычной электронной энциклопедии). Однако автор обзора отмечает, что для ряда терминов предпочтительно использовать не то соответствие, которое вошло в глоссарий. Иными словами, пользователь ChatGPT получает «готовое» решение, требующее проверки и постредактирования.

Проведённое нами тестирование российского аналога – чат-бота YandexGPT [YandexGPT API] – позволило убедиться, что к предъявляемым текстам и на русском, и английском языке сервис создаёт глоссарий, представляющий собой подробный русскоязычный список терминов с дефинициями. Использование только русского языка как языка выдачи YandexGPT подтверждают и представители компании «Яндекс» [Патрушева, Овчинникова, 2023]. Соответственно, чат-бот не выполняет задачу формирования двуязычного глоссария. Создаваемые им одноязычные глоссарии, бесспорно, могут использоваться переводчиком как полезное дополнение.

Таким образом, большинство рассмотренных средств автоматизированного формирования глоссариев не предоставляют удовлетворительного результата для русского языка, не формируют двуязычного глоссария или недоступны в бесплатной версии. Безусловно, настоящий обзор нельзя назвать исчерпывающим, и постоянно появляются новые приложения для работы с глоссариями. Однако тематические публикации свидетельствуют о том, что и устные, и письменные переводчики при составлении глоссариев часто ограничиваются использованием программ Word и Excel [Девятов, 2018 ; Fantinuoli, 2017 ; Techforword, 2023].

В настоящей статье нами предлагается средство автоматизации процесса составления двуязычных глоссариев для студентов переводческих специальностей. В отличие от профессионального переводчика, студенты в процессе обучения реже получают «рекомендации от заказчика», касающиеся перевода тех или иных терминов, и чаще ищут соответствия сами на основе словарей и параллельных текстов. Кроме того, они реже работают над масштабными проектами и чаще переводят тексты, небольшие по объему, но разнообразной тематики. В рамках занятий студенты получают доступ к некоторым системам помощи переводчику, но речь, как правило, идёт о системах переводческой памяти, в которых пользователь пополняет глоссарий в ручном режиме. Студенты могут не располагать средствами для самостоятельного приобретения рассмотренных выше специализированных программ. Однако они проявляют бесспорный интерес к развитию информационных технологий и ищут способы оптимизировать свою работу – что нередко приводит к использованию ими общеизвестных систем машинного перевода. Недостаток этих систем состоит в том, что они предлагают «готовый» текст на переводящем языке, не оставляя пользователю возможности принимать решения и учиться.

В связи с этим представляется актуальной задача разработки общедоступного и несложного в использовании сервиса для автоматического извлечения терминологии из

специального (нехудожественного) текста и составления глоссариев, который может применяться студентами в процессе обучения.

Ц е л ь настоящей работы состоит в том, чтобы представить тестовую версию разработанного нами сервиса для формирования двуязычных глоссариев, результаты его оценивания студентами и преподавателями переводческого факультета лингвистического вуза и доработки по результатам тестирования. З а д а ч и проекта – с одной стороны, ускорить и облегчить работу студентов с текстами и лексикографическими источниками. С другой стороны – предложить альтернативу системам машинного перевода, так, чтобы обучающая составляющая при подготовке глоссария сохранялась, студенты обращались к словарям и самостоятельно выбирали из нескольких возможных соответствий эквивалент лексической единицы в переводящем языке на основании контекста.

## 2. Описание сервиса [Description of the service]

Нами был разработан сервис для формирования глоссариев на основе текстов на русском и английском языке. Он дорабатывался с учётом пожеланий пользователей и сейчас доведён до состояния минимально жизнеспособного продукта (MVP).

Разработка представляет собой чат-бот в мессенджере Telegram [Telegram]. Такая форма была выбрана по нескольким причинам: во-первых, приложение Telegram позволяет принимать запросы от пользователя, анализировать их и отдавать результат напрямую пользователю. Во-вторых, в Telegram есть возможность подключаться к сторонним приложениям с помощью технологии API. Кроме того, при использовании приложения не требуется получать от пользователя согласие на сбор персональных данных: пользователь идентифицируется по номеру Telegram ID, представляющему собой набор цифр, по которому невозможно сделать выводы о других персональных данных. Приложение может использоваться и на смартфоне, и на компьютере, и получило широкое распространение, так что у многих потенциальных пользователей сервиса оно уже установлено. Формат чат-бота также уже известен многим пользователям. Чат-бот написан на языке программирования Python с использованием библиотек NLTK, SpaCy и PyMorphy2.

Основной принцип работы сервиса заключается в том, что он извлекает из полученного от пользователя текста ключевые слова по методике, предложенной А. Килгарриффом [Kilgarriff, 2009]. Ключевые слова в данном случае, – слова, частотность которых в рассматриваемом тексте выше, чем частотность слов в корпусе, с которым осуществляется сравнение (референсном корпусе). При выявлении ключевых слов в текстах на русском языке основанием для сравнения послужил Национальный корпус русского языка, а данные о частотности слов в нём были получены из Частотного словаря современного русского языка [Ляшевская, Шаров, 2009]. Для подсчёта частотности слов в предъявляемых текстах и для обращения к указанному словарю использовались возможности библиотеки NLTK. На материале английского языка основанием для сравнения послужил Брауновский корпус (Brown corpus), встроенный в пакет корпусов библиотеки NLTK [NLTK].

В первоначальной версии разработанного сервиса, предложенной пользователям для тестирования, для того, чтобы получить глоссарий к тексту, необходимо было отправить чат-боту сообщение, прикрепив к нему текст в виде файла в формате TXT. Впоследствии формат отправки пользователями запроса был изменён (внесённые изменения рассмотрены подробнее в разделе 4 настоящей статьи).

После отправки текста пользователь получал от чат-бота запрос, в котором предлагалось выбрать соотношение количества единиц глоссария и их точности по шкале от 1 до 5, где под точностью понимается сфокусированность результатов автоматического отбора лексем на единицах, являющихся ключевыми для анализируемого текста. Ины-

ми словами, выбор параметра «1» позволяет получить в результатах обработки большее количество лексических единиц, а выбор параметра «5» – меньшее количество единиц, но именно единицы, встречающиеся в рассматриваемом тексте существенно чаще, чем в корпусе, который служит основанием для сравнения.

Затем чат-бот высылал пользователю результат обработки текста (в протестированной пользователями версии – в виде текстового сообщения Telegram). Ответ от чат-бота представлял собой список лексических единиц, разделённых на две категории: «термины» (имена нарицательные) и «именованные сущности» (прецизионные слова).

Наименование «термины» в данном случае используется условно для обозначения имён нарицательных, частотность которых в рассматриваемом тексте выше, чем в референсном корпусе. С помощью этих слов в тексте передается основная когнитивная информация. Поскольку сервис предназначен для обработки нехудожественных текстов, предполагалось, что ключевыми в первую очередь будут слова для точного обозначения специальных предметов и понятий. Как показал анализ сформированных сервисом глоссариев, из текстов извлекаются не только собственно термины, но и слова, принадлежащие к общеупотребительной лексике – например, из статьи о велосипедах, не только *derailleur, wheelbase, cantilever*, но и *bicycle, bike*. Однако и извлечённые общеупотребительные слова в большинстве случаев являются ключевыми для понимания и передачи смысла текста. В переводе для них необходимо подобрать точное соответствие (в то время как использование приемов описательного перевода или опущения может привести к информационным потерям), и в этом плане они близки к терминам, даже если принадлежат к общеупотребительной лексике, изучаемой студентами на начальных этапах.

В сообщении от чат-бота, содержащем список лексики к тексту, лексемы приводились к исходной форме (лемме) и были представлены в виде гиперссылок. «Термины» сопровождались контекстом – фрагментом из текста оригинала.

Гиперссылки от «терминов» позволяли перейти к статьям, описывающим соответствующие лексемы, на сайте контекстного словаря Glosbe [Glosbe] для пары языков английский-русский. Контекстный словарь в качестве лексикографического источника был выбран потому, что в нём пользователю, помимо собственно эквивалентов, предлагается большое количество текстовых фрагментов, содержащих исходную единицу, с переводами. Благодаря этому, пользователь имеет возможность оценить решения, ранее принятые переводчиками при передаче искомой лексической единицы, и изучить контексты, в которых лексическая единица переводится тем или иным образом. Выбор в пользу словаря Glosbe [Glosbe] был сделан благодаря наличию в нём статей для пары языков английский-русский и технической возможности настроить его взаимодействие с чат-ботом. При переходе по ссылке от «именованной сущности» пользователь попадал на страницу с результатами поиска по соответствующему запросу в поисковой системе Google [Google], что позволяло ему при необходимости получить дополнительную информацию, связанную с прецизионными словами.

Таким образом, извлечение ключевых слов из текста и поиск словарных статей или справочной информации чат-ботом осуществлялись автоматически, а решение о выборе соответствия пользователь принимал сам.

## 3. Результаты тестирования сервиса пользователями [Results of user testing]

В тестировании чат-бота приняли участие 11 студентов бакалавриата профиля «Перевод и переводоведение», изучавших английский язык в течение пяти и более лет, и три преподавателя иностранных языков, устного и письменного перевода, изучавших английский язык более 10 лет. Пользователи проанализировали с помощью чат-бота 16

текстов на английском языке объёмом от 457 до 2259 слов (в среднем объём текстов составил примерно 890 слов) и два текста на русском языке объёмом 278 и 374 слова. По результатам тестирования студенты и преподаватели заполняли анкету: оценивали работу сервиса, высказывали замечания и предложения. Пользователи не были ограничены в выборе тематики текстов. Студенты отправляли чат-боту тексты на английском языке на такие темы, как спорт, кино, техника, политика, путешествия, законодательство, наука, финансы. Четверо студентов отметили в анкете, что тестировали глоссарий на текстах домашнего задания по иностранному языку, устному или письменному переводу, а семь респондентов подбирали текст специально для тестирования. Преподаватели загрузили тексты на английском языке научной и технической тематики, на русском языке – тексты, посвящённые политике и общественной жизни. В анкете они указали, что используют эти материалы на занятиях по иностранному языку и переводу.

В большинстве случаев респонденты не испытывали технических трудностей при использовании чат-бота. У одного обучающегося текст в после сохранения файла в формате TXT превратился в нечитаемый набор символов, но проблема была связана с кодировкой самого текста, а не с работой сервиса, и впоследствии была устранена. Также в комментариях один из пользователей высказал пожелание о добавлении формата DOC для текстов. В данном случае TXT как формат отправляемого чат-боту файла выбран по техническим причинам. Однако очевидно, что необходимость выполнять дополнительное действие по сохранению текста в файле формата TXT могла отпугнуть потенциальных пользователей.

Кроме того, в трёх случаях сервис выдал ошибку: «В тексте так много терминов, что они не убираются в сообщение Телеграма». Это объясняется тем, что в мессенджере Telegram существует ограничение, в соответствии с которым максимальная длина сообщения составляет не более 4096 символов. В указанных случаях список «терминов», фрагменты контекста к ним и список «именованных сущностей» на основе отправленного пользователем текста превышали по суммарному объёму максимально допустимое количество символов. В инструкции к сервису пользователям было рекомендовано загружать файлы объёмом от 500 до 3000 слов. Однако, как показало тестирование, проблема возникала с некоторыми текстами объёмом от 1200 слов. После того, как пользователи сократили эти тексты, сервис сформировал глоссарий к ним.

При заполнении анкеты пользователи также анализировали содержание глоссариев, сформированных сервисом. Так, необходимо было оценить, вошла ли в глоссарий вся лексика, необходимая пользователю, по шкале от одного до пяти, где 1 означает «скорее нет», а 5 – «скорее да». Среди студентов два респондента выставили по 3 балла, пятеро – по 4, четверо – по 5. Все преподаватели выставили по 3 балла.

В комментариях четыре пользователя из числа студентов и трое преподавателей указали лексические единицы, которые, на их взгляд, должны присутствовать в глоссарии, но не были выделены сервисом. Среди них преобладали составные термины, например: *climate change, animal fossil, polar desert, environmental DNA, bell curve, battery capacity, dental care; региональная площадка, субъект экономической деятельности*. Встречались также однословные лексемы: *awareness, literacy, конкурентоспособность, внешнеполитический, инновационный*.

Следующий вопрос анкеты касался того, содержал ли сформированный системой глоссарий не интересовавшие пользователя лексические единицы. По шкале от одного до пяти, где 1 означает «их было много», а 5 – «их совсем не было», двое пользователей-студентов выставили по 3 балла, шестеро – по 4 балла, трое – по 5 баллов. Среди преподавателей двое выставили по 2 балла, один – 4 балла. В комментарии пользователи приводили пример «лишних», на их взгляд, лексических единиц. Анализ комментариев показал, что причиной их попадания в список было несколько факторов:

– ошибки сервиса, возникшие по неясным пока причинам – например, попадание в список «терминов» глагольных форм *got, wouldn't*;

– влияние капитализации, орфографических ошибок или опечаток на результаты отбора лексики. Например, в глоссарий к тексту о выращивании огурцов попало слово *Tasty*, которое упоминалось в названиях сортов и дважды было написано в тексте с заглавной буквы. Из текста на русском языке в глоссарий попала единица *28-ми* (цитата: … *форума, организованного, к тому же, на 28-ми региональных площадках*), написание которой не является корректным;

– неправильная обработка сервисом некоторых спецсимволов – в частности, дефиса, который классифицировался сервисом как разделитель слов. В результате такие термины, как *infra-red, Li-ion*, анализировались по частям, и эти части попадали в глоссарий в качестве «ключевых» слов к тексту;

– влияние тематики текста на результат. Так, один из пользователей указал на появление в списке «терминов» таких принадлежащих к базовой лексике слов, как *imaginable, iconic, boyfriend, lonely, romantic*. Исходный текст был сравнительно небольшим – объёмом 522 слова, был посвящён актёрской игре Николаса Кейджа в фильме «Ренфилд» и точных обозначений специальных предметов или понятий, то есть собственно терминов, не содержал.

На наш взгляд, наличие в глоссарии слов, уже известных пользователю (в отличие от слов с опечатками и ошибками) не является значительной проблемой, поскольку пользователь может игнорировать их, не переходя по гиперссылкам в двуязычный словарь.

Этап перехода по гиперссылкам не вызвал у студентов и преподавателей технических трудностей, но в одном из отзывов была указана проблема отсутствия в словаре Glosbe [Glosbe] статьи к спортивному термину *buildup play*. Один из преподавателей загрузил в сервис текст, содержащий сокращения (*product dimensions: L 123.4 mm, W 123.4 mm, H 132 mm*), и остался недоволен результатом: сокращение *H* в список извлечённой лексики не попало, *W* было распознано как «именованная сущность», а *L* – как «термин». Однако соответствующая статья в словаре Glosbe [Glosbe] не содержит эквивалента «длина», который следует использовать в переводе рассматриваемого контекста. На наш взгляд, перевод сокращений, ввиду их многозначности, требует особого внимания, и полностью полагаться на технические средства в этом вопросе не следует в любом случае.

Общая оценка применимости чат-бота пользователями была удовлетворительной. На вопрос, насколько часто они бы стали использовать чат-бот вместо двуязычного словаря по шкале от одного до пяти, где 1 означает «никогда», а 5 – «постоянно», двое пользователей-студентов выставили по 2 балла, пятеро – по 3, четверо – по 4. Преподаватели выставили 1, 2 и 3 балла. На вопрос, насколько часто чат-бот мог бы использоваться вместо систем машинного перевода, по шкале от одного до пяти, где 1 означает «никогда», а 5 – «постоянно», один студент выставил 1 балл, двое – по 2, шестеро – по 3, двое – по 4. Среди преподавателей двое выставили 1 балл, один – 3 балла.

Ещё один вопрос анкеты касался возможных целей применения сервиса. От студентов были получены следующие ответы (авторская орфография сохранена): «для предпереводческого анализа текста»; «в учебных целях, если возникнет необходимость быстро обработать/проанализировать текст»; «составление глоссария для коротких текстов, например, текстов на домашнее чтение»; «для будущей работы», «чтобы проверить, стоит ли читать текст большого объёма ради лексики. Если лексика знакомая, смысла в этом может не быть, т. е. бот сэкономит время»; «для перевода сложных текстов; для составления глоссария к дипломной или исследовательской работе»; «для препереводческого анализа текста и работы над самим переводом»; «для составления глоссария по рассказу (книги и т. п.), который надо было прочитать, для последующего отсеивания терминов, которые мне уже известны»; «в процессе устного перевода, когда

нужно быстро найти информацию по определённому термину/имени или названию, а также в процессе письменного перевода текстов»; «для быстрого составления глоссариев на английском языке». Кроме того, один студент указал, что сервис позволяет значительно сократить время на поиск неизвестных слов. Таким образом, студенты рассматривают сервис в первую очередь как инструмент для ускорения предварительного анализа лексического состава текста.

По мнению преподавателей, тестировавших сервис, он может использоваться «для составления глоссариев для текстов различного применения, в том числе для перевода». Однако в комментариях преподаватели отметили, что студенты должны учиться самостоятельно находить и составлять глоссарии к текстам, а чат-бот скорее пригодился бы уже в практической деятельности для экономии времени переводчика. Постановщик задачи по разработке чат-бота и соавтор настоящей статьи, будучи также преподавателем иностранного языка и перевода, не соглашается с этой точкой зрения, поскольку считает, что опыт работы с различными техническими средствами даёт начинающим переводчикам преимущества на рынке труда, и осваивать их и в то же время узнавать границы их возможностей, необходимо уже в процессе обучения.

## 4. Доработка сервиса по результатам тестирования и перспективы [Improvement of the service based on testing results and work prospects]

По результатам тестирования глоссария пользователями в принципы работы сервиса были внесены некоторые изменения, направленные на то, чтобы сделать его более простым в использовании. Во-первых, был изменён формат получения запроса от пользователей. Загружая файлы в формате TXT, пользователи не замечали в них опечатки, однако их наличие приводило к искажениям в работе глоссария. Кроме того, как показал анализ загруженных материалов, если текст был предварительно скопирован из файла в формате PDF, в нём появлялись символы конца абзаца в тех местах, где в исходном тексте заканчивались строки, в том числе в середине предложений, а знак переноса мог преобразоваться в дефис в середине слова. Необходимость переводить файл в формат TXT и в целом создавала неудобства для пользователей. Потому в более позднюю версию сервиса была интегрирована функция отправки сообщения чат-боту через сервис для проведения опросов Yandex Forms [Yandex Forms]. В форму Yandex Forms копируется собственно текст, так что нет необходимости сохранять его в виде отдельного файла. Для русского языка в конструкторе форм Yandex Forms есть функция проверки орфографии.

Во-вторых, была изменена форма «ответа» от чат-бота пользователю. В новой версии чат-бот отправляет в качестве ответа пользователю файл с исходным текстом, в котором «термины» и «именованные сущности» выделены графически и представляют собой гиперссылки на статьи в двуязычном словаре Glosbe [Glosbe] и на результаты поиска в Google [Google], соответственно. Формат ответа в виде текста с «глоссами», на наш взгляд, более наглядный (так, в первоначальной версии контексты к «терминам» в ответном сообщении чат-бота представляли собой фрагменты предложений, и один из пользователей выразил пожелание, чтобы в качестве контекстов использовались только целые предложения). Переход к этому формату также позволил отказаться от использования в выдаваемом пользователю результате работы сервиса условных обозначений «термины» и «именованные сущности». Кроме того, отправка в качестве ответа пользователю текстового файла позволяет снять ограничения, касающиеся объёма анализируемых текстов. Объём файла, пересылаемого в Telegram в приложении к сообщению, может составлять до 2 Гб, что для текстового файла означает практически полное отсутствие ограничений по объёму.

## 5. Заключение [Conclusion]

Итак, в статье представлен разработанный нами сервис для формирования двуязычных глоссариев на основе специальных текстов на русском и английском языке, предназначенный для студентов переводческих специальностей. Потребность в этом сервисе связана с тем, что на рынке появляется всё больше программ для автоматизации тех или иных аспектов переводческой работы, однако рассмотренные программы для извлечения терминов из текста, доступные для использования без дорогостоящей подписки, предоставляют для русского языка неудовлетворительный результат.

Предлагаемый нами сервис представляет собой чат-бот в популярном мессенджере, который автоматизирует выявление в тексте лексических единиц, требующих обращения к словарю, и поиск их в двуязычном контекстном словаре. Тестирование показало, что сервис функционирует удовлетворительно. Выявленные ошибки в его работе преимущественно связаны с тем, что он чувствителен к нестандартной орфографии в загружаемых в него текстах.

В отличие от сервисов, рассмотренных в обзоре литературы, созданный нами чат-бот не предлагает пользователю однозначное соответствие, а позволяет выбрать из нескольких, указанных в двуязычном контекстном словаре. Такой формат не ускоряет работу переводчика, так что чат-бот может, на первый взгляд, представляться менее эффективным средством автоматизации составления глоссариев, чем сервисы из числа рассмотренных выше, корректно работающие с русским языком. Однако в ситуации обучения мы считаем этот формат более полезным, чем списки соответствий от рассмотренных сервисов, которые начинающему переводчику могут показаться готовыми к использованию, но зачастую требуют постредактирования.

Кроме того, предлагаемая в актуальной версии сервиса выдача результата в виде текста оригинала с глоссами-гиперссылками позволит студентам сконцентрироваться на контексте в большей степени, чем отдельные таблицы или карточки, содержащие только термин на исходном языке и соответствие в переводящем языке, которые создаются предназначенными для профессиональных переводчиков приложениями и удобны, в первую очередь, когда речь идёт о работе с большими объёмами текста одной тематики. В учебной ситуации тематики меняются чаще, чем в реальной практической деятельности переводчика. Как показал опрос пользователей, студенты рассматривают сервис в первую очередь как инструмент для быстрого предварительного анализа лексического состава текста, за которым последует более углубленная самостоятельная работа с текстом.

Хотя сервис, бесспорно, нуждается в дальнейших доработках, мы считаем выполненными основные задачи, поставленные перед ним: ускорить работу студентов с текстами и лексикографическими источниками и предложить альтернативу системам машинного перевода так, чтобы студенты выбирали переводческие соответствия самостоятельно.

В ходе дальнейшей работы планируется настроить функцию извлечения из текста неоднословных терминов, продолжить тестирование и расширить круг пользователей. Пользователями было высказано пожелание о том, чтобы сервис мог поддерживать другие иностранные языки, помимо английского, и их добавление также входит в перспективы работы. Кроме того, рассматривается вопрос о добавлении функции извлечения двуязычных соответствий из параллельных текстов. С одной стороны, она позволит составлять глоссарий с учётом «требований заказчика» и использовать эквиваленты, рекомендованные в рамках определённого проекта, как это принято в переводческой практике. С другой стороны, появление этой функции приведет к ещё большей автоматизации процесса составления глоссариев, которой, как показал опрос, опасаются преподаватели.

**Библиографический список**

Бухаров и др., 2021 – Компьютерная лингвистика. Начальные сведения. Учебно-методические материалы / В. М. Бухаров, Ю. В. Балакина, И. С. Парина, М. Б. Чиков. Н. Новгород : Нижегородский государственный лингвистический университет им. Н. А. Добролюбова, 2021. 117 с.

Девятов, 2018 – Девятов М. Средства для извлечения терминологии. Забава или экономия времени? // Prima Vista. 2018. URL : https://www.primavista.ru/blog/2018/08/17/sredstva-dlya-izvlecheniya-terminologiii?ysclid=lnxm77ady1483455647 (дата обращения : 21.10.2023).

Дубичинский, 2008 – Дубичинский В. В. Лексикография русского языка: учеб. пособие. М. : Наука: Флинта, 2008. 432 с.

Еолян, Муратова, 2017 – Еолян В. Ю., Муратова Э. Д. Глоссарий как инструмент повышения качества перевода // Международный научный журнал «Молодой ученый». 2017. № 31 (165). С. 83–85.

Жеребило, 2010 – Жеребило Т. В. Словарь лингвистических терминов. Изд. 5–е, испр. и доп. Назрань : ООО «Пилигрим», 2010. 486 с.

Лешканова, Бубнова, 2018 – Лешканова Е. О., Бубнова А. С. К вопросу о составлении переводческого глоссария // Филологический аспект. 2018. № 4 (36). С. 223–229.

Ляшевская, Шаров, 2009 – Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М. : Азбуковник, 2009. 1087 с.

Парина, Салаев, 2022 – Парина И. С., Салаев П. В. Инструменты создания автоматизированных глоссариев для подготовки устных переводчиков в специальных областях // Профессионально ориентированный перевод: реальность и перспективы : Сборник научных трудов / под редакцией Н. Н. Гавриленко. Выпуск 17. М. : Российский университет дружбы народов (РУДН), 2022. С. 290–301.

Патрушева, Овчинникова, 2023 – Патрушева А., Овчинникова П. Chat GPT по-русски: на что способна нейронка от Яндекса // Блог Яндекс Практикума. 2023. URL : https://practicum.yandex.ru/blog/neyroset-yandexgpt-kak-polzovatsya/ (дата обращения : 08.06.2024).

Синхронный и последовательный перевод, 2015 – Синхронный и последовательный перевод. Рекомендации практикующим переводчикам. Вторая редакция / сост. Н. К. Дупленский. М. : Р. Валент, 2015. 64 с.

Чернов, 1978 – Чернов Г. В. Теория и практика синхронного перевода. М. : Междунар. отношения, 1978. 208 с.

Beolingus – Beolingus. URL : https://dict.tu-chemnitz.de/ (дата обращения : 20.10.2023).

Concordancer for Windows 3.0 – Concordancer for Windows 3.0. URL : https://www.englishelp.ru/soft/soft-for-translator/95-concordancer-for-windows.html (дата обращения : 21.10.2023).

Dict.cc – Dict.cc. URL : https://www.dict.cc/ (дата обращения : 20.10.2023).

Fantinuoli, 2017 – Fantinuoli C. Computerlinguistik in der Dolmetschpraxis unter besonderer Berücksichtigung der Korpusanalyse // S. Hansen-Schirra, S. Neumann, O. Čulo (Hrsg.): Annotation, exploitation and evaluation of parallel corpora. Berlin : Language Science Press, 2017. S. 111–146.

Flashterm – Flashterm. URL : https://www.flashterm.eu/ (дата обращения : 21.10.2023).

Germling. – Germling. Create a translation glossary using ChatGPT. URL : https://www.youtube.com/watch?v=rIwX-lhPRgY(дата обращения : 08.06.2024).

Glosbe – Glosbe. URL : https://glosbe.com/en/ru (дата обращения : 25.10.2023).

Google – Google. URL : https://www.google.ru/ (дата обращения : 05.11.2023).

IATE – IATE. URL : https://iate.europa.eu/ (дата обращения : 20.10.2023).

Interplex – Interplex Glossary Software for Interpreters and Translators URL : http://www.fourwillows.com/interplex.html (дата обращения : 21.10.2023).

InterpretBank – InterpretBank. URL : https://www.interpretbank.com/site/ (дата обращения : 20.10.2023).

Kilgarriff, 2009 – Kilgarriff A. Simple maths for keywords // Proc. of the Corpus Linguistics Conference (CL2009) / M. Mahlberg, V. González-Díaz, C. Smith (eds.). Liverpool. 2009. URL : https://www.sketchengine.eu/wp-content/uploads/2015/04/2009-Simple-maths-for-keywords.pdf (дата обращения : 27.10.2023).

LEO – LEO.org – Ihr Sprachangebot im Web. URL : https://www.leo.org/englisch-deutsch/ (дата обращения : 20.10.2023).

MemoQ – MemoQ. URL : https://www.memoq.com/memoq-versions/memoq-8-2 (дата обращения : 21.10.2023).

NLTK – NLTK. Documentation. Sample Usage for Corpus. URL : https://www.nltk.org/howto/corpus.html (дата обращения : 07.06.2024).

Phrase – Phrase (Frm. Memsource): Localization and Translation Software. URL : https://phrase.com/ (дата обращения : 21.10.2023).

Prospector – Prospector. URL : https://prospector.logrusglobal.com/ (дата обращения : 21.10.2023).

SDL MultiTerm Extract – SDL MultiTerm Extract. URL : https://tra-service.ru/multitermextract?ysclid=lo0j8whfa8578780139 (дата обращения : 21.10.2023).

Stoll, 2009 – Stoll Chr. Jenseits simultanfähiger Terminologiesysteme. Trier : Wvt Wissenschaftlicher Verlag, 2009. 342 S.

SynchroTerm – SynchroTerm. URL : https://terminotix.com/index.asp?content=category&cat=6&lang=en (дата обращения : 21.10.2023).

Techforword, 2023 – Stop using Word and Excel for interpreting glossaries // Techforword. 2023. URL : https://www.techforword.com/blog/stop-using-word-and-excel-for-interpreting-glossaries/(дата обращения : 25.10.2023).

Telegram – Telegram. URL : https://web.telegram.org/a/ (дата обращения : 25.10.2023).

Tilde – Tilde Terminology Services. URL : https://tilde.com/ (дата обращения : 21.10.2023).

Word Tabulator 2.2.3 – Word Tabulator 2.2.3. URL : https://www.englishelp.ru/topics/96.html (дата обращения : 21.10.2023).

Yandex Forms – Yandex Forms URL : https://cloud.yandex.ru/docs/forms/ (дата обращения : 25.10.2023).

YandexGPT API – YandexGPT API. URL : https://yandex.cloud/ru/services/yandexgpt (дата обращения : 07.06.2024).

**References**

Bukharov, V. M., Balakina, Yu. V., Parina, I. S., & Chikov, M. B. (2021). *Komp'yuternaya lingvistika. Nachal'nye svedeniya. Uchebno-metodicheskie materialy [Computational linguistics. Initial information. Educational materials]*. N. Novgorod : Linguistics University of Nizhny Novgorod Press. (In Russ.).

Devyatov, M. (2018). Sredstva dlya izvlecheniya terminologii. Zabava ili ekonomiya vremeni? [Terminology extraction tools. Fun or time saving?]. *Prima Vista*. Retrieved October 21, 2023 from <https://www.primavista.ru/blog/2018/08/17/sredstva-dlya-izvlecheniya-terminologiii?ysclid=lnxm77ady1483455647>. (In Russ.).

Dubichinskiy, V. V. (2008). *Leksikografiya russkogo yazyka: ucheb. posobie [Lexicography of the Russian language: a textbook]*. Moscow : Nauka Press : Flinta Press (In Russ.).

Eolyan, V. Yu., & Muratova, E. D. (2017). Glossariy kak instrument povysheniya kachestva perevoda [Glossary as a tool for improving translation quality]. *Molodoy uchenyy [Young Scientist], 31* (165), 83–85. (In Russ.).

Zherebilo, T. V. (2010). *Slovar' lingvisticheskikh terminov [Dictionary of linguistic terms]*. 5th edn, revised and enlarged. Nazran : Piligrim OOO Press. (In Russ.).

Leshkanova, E. O., & Bubnova, A. S. (2018). K voprosu o sostavlenii perevodcheskogo glossariya [To the issue of creating a translation glossary]. *Filologicheskiy aspekt [Philological Aspect], 4* (36), 223–229. (In Russ.).

Lyashevskaya, O. N., & Sharov, S. A. (2009). *Chastotnyy slovar' sovremennogo russkogo yazyka (na materialakh Natsional'nogo korpusa russkogo yazyka) [Frequency dictionary of contemporary Russian (Based on the Russian National Corpus]*. Moscow : Azbukovnik Press. (In Russ.).

Parina, I. S., & Salaev, P. V. (2022). Instrumenty sozdaniya avtomatizirovannykh glossariev dlya podgotovki ustnykh perevodchikov v spetsial'nykh oblastyakh [Automated glossary management tools for interpreters in special fields]. In N. N. Gavrilenko (Ed.), *Professional'no orientirovannyy perevod: real'nost' i perspektivy [Professionally oriented translation: Reality and prospects]: Collection of scientific papers* (Vol. 17, pp. 290–301); Moscow : RUDN University Press. (In Russ.).

Patrusheva, A., & Ovchinnikova, P. (2023). Chat GPT po-russki: na chto sposobna neyronka ot Yandeksa [Chat GPT in Russian: what the neural network by Yandex is capable of]. *Yandex Practicum Blog*. Retrieved June 8, 2024 from <https://practicum.yandex.ru/blog/neyroset-yandexgpt-kak-polzovatsya/>. (In Russ.).

Duplenskiy, N. K. (Ed.). (2015). *Sinkhronnyy i posledovatel'nyy perevod. Rekomendatsii praktikuyushchim perevodchikam [Simultaneous and consecutive interpreting. Recommendations for practicing interpreters]*. 2nd edn. Moscow : R. Valent Press. (In Russ.).

Chernov, G. V. (1978). *Teoriya i praktika sinkhronnogo perevoda [Theory and practice of simultaneous interpreting]*. Moscow : Mezhdunarodnye Otnosheniya Press. (In Russ.).

Beolingus (n. d.). Retrieved October 20, 2023 from <https://dict.tu-chemnitz.de/>.

Concordancer for Windows 3.0 (n. d.). Retrieved October 21, 2023 from <https://www.englishelp.ru/soft/soft-for-translator/95-concordancer-for-windows.html>.

Dict.cc (n. d.). Retrieved October 20, 2023 from <https://www.dict.cc/>.

Fantinuoli, C. (2017). Computerlinguistik in der Dolmetschpraxis unter besonderer Berücksichtigung der Korpusanalyse. In S. Hansen-Schirra, S. Neumann, O. Čulo (Hrsg.), *Annotation, exploitation and evaluation of parallel corpora* (S. 111–146). Berlin : Language Science Press.

Flashterm (n. d.). Retrieved October 21, 2023 from <https://www.flashterm.eu/>.

Germling (n. d.). Retrieved June 7, 2024 from <https://www.youtube.com/watch?v=rIwX-lhPRgY>.

Glosbe (n. d.). Retrieved October 25, 2023 from <https://glosbe.com/en/ru>.

Google (n. d.). Retrieved November 5, 2023 from <https://www.google.ru/>.

IATE (n. d.). Retrieved October 20, 2023 from: <https://iate.europa.eu/>.

Interplex Glossary Software for Interpreters and Translators (n. d.). Retrieved October 21, 2023 from <http://www.fourwillows.com/interplex.html>.

InterpretBank (n. d.) Retrieved October 21, 2023 from <https://www.interpretbank.com/site/>.

Kilgarriff, A. (2009). Simple maths for keywords. In M. Mahlberg, V. González-Díaz, C. Smith (Eds), *Proc. of the Corpus Linguistics Conference (CL2009)*. Liverpool. Retrieved October 27, 2023 from <https://www.sketchengine.eu/wp-content/uploads/2015/04/2009-Simple-maths-for-keywords.pdf>.

LEO (n. d.). LEO.org – Ihr Sprachangebot im Web. Retrieved October 20, 2023 from <https://www.leo.org/englisch-deutsch/>.

MemoQ (n. d.). Retrieved October 21, 2023 from <https://www.memoq.com/memoq-versions/memoq-8-2>.

NLTK (n. d.). Retrieved June 7, 2024 from <https://www.nltk.org/howto/corpus.html>.

Phrase (n. d.). (Frm. Memsource): Localization and Translation Software. Retrieved October 21, 2023 from <https://phrase.com/>.

Prospector (n. d.) Retrieved October 21, 2023 from <https://prospector.logrusglobal.com/>.

SDL MultiTerm Extract (n. d.). Retrieved October 21, 2023 from <https://tra-service.ru/multitermextract?ysclid=lo0j8whfa8578780139>.

Stoll, Chr. (2009). *Jenseits simultanfähiger Terminologiesysteme.* Trier : Wvt Wissenschaftlicher Verlag, 2009.

SynchroTerm (n. d.). Retrieved October 21, 2023 from <https://terminotix.com/index.asp?content=category&cat=6&lang=en>.

Techforword. Stop using Word and Excel for interpreting glossaries. (2023). Retrieved October 25, 2023 from <https://www.techforword.com/blog/stop-using-word-and-excel-for-interpreting-glossaries/>.

Telegram (n. d.). Retrieved October 25, 2023 from <https://web.telegram.org/a/>.

Tilde Terminology Services (n. d.). Retrieved October 21, 2023 from <https://tilde.com/>.

Word Tabulator 2.2.3 (n. d.). Retrieved October 21, 2023 from <https://www.englishelp.ru/topics/96.html>.

Yandex Forms (n. d.). Retrieved October 25, 2023 from <https://cloud.yandex.ru/docs/forms/>.

NLTK (n. d.). Retrieved June 7, 2024 from URL : <https://www.nltk.org/howto/corpus.html>.

Статья поступила в редакцию 08.11.2023; одобрена после рецензирования 03.06.2024; принята к публикации 26.06.2024.
The article was submitted 08.11.2023; approved after reviewing 03.06.2024; accepted for publication 26.06.2024.