УДК 81’322.4

ОЦЕНКА КАЧЕСТВА МАШИННОГО ПЕРЕВОДА (на материале городского портала)

А. А. Новожилова, С. А. Королькова
Волгоградский государственный университет
Российская Федерация
E-mail: novozilova@volsu.ru, korolkova@volsu.ru

Настоящая статья посвящена анализу системы машинного перевода Яндекс. Переводчик и ее использованию в сети Интернет для перевода текстов городских порталов. Авторы оценивают современное состояние систем машинного перевода, анализируют ошибки, допускаемые программой при переводе урбанистических текстов с немецкого языка на русский.

Ключевые слова: информационные технологии; машинный перевод; качество перевода; семантические ошибки; урбанистический дискурс

В эпоху всеобщей глобализации и стремительного развития цифровых технологий и искусственного интеллекта все чаще можно слышать прогнозы о том, что скоро миру не нужны будут переводчики и люди со знанием иностранных языков. Такие утверждения можно слышать не только от разработчиков систем машинного перевода, но и от специалистов в области переводоведения, которые отмечают, что объем текстовых документов и материалов, предназначающихся для перевода, «растет экспоненциально и в этой ситуации классический (ручной) подход к переводу не всегда оправдывает ожидания, так как это требует значительных временных и финансовых вложений. В некоторых случаях целесообразнее использовать машинный (автоматический) перевод» [4, с. 585]. Другие специалисты менее категоричны в выводах, но подчеркивают, что «существующие системы машинного перевода пока не способны конкурировать с живыми переводчиками,

однако быстро их догоняют» [1]. Однако пока еще сложно предположить, когда все-таки искусственный интеллект сможет полноценно выполнять работу вместо профессиональных переводчиков, но объективную оценку эффективности машинных переводчиков на настоящем уровне их развития можно дать, проведя лингвистический анализ качества транслята, под которым понимается текст, полученный в результате письменного или устного перевода [2, с. 267].

Настоящая статья посвящена описанию популярной среди пользователей Рунета системы машинного перевода Яндекс.Переводчик, который 10 лет назад, на момент своего запуска 16 марта 2011 имел в своем ассортименте лишь 3 языка, а на сегодняшний день расширил свой языковой арсенал до 97. Данная система сочетает в себе статистическую модель (поиск двуязычных соответствий в параллельных текстах) и искусственную нейронную сеть, которая самообучается на множестве доступных ей примеров. В качестве анализируемого материала послужили тексты урбанистического дискурса, а именно контент официального городского портала г. Берлин. Такие сайты интересны самым разным целевым группам пользователей интернета, начиная от местных жителей и иностранных туристов и заканчивая крупными бизнесменами и инвесторами, собирающими информацию о той или иной территории.

Контент официального городского портала г. Берлина «*Berlin.de Das offizielle Hauptstadtportal*» [3] содержит разножанровые тексты, описывающие достопримечательности немецкой столицы, ее административную систему, транспортную структуру и т.п. Здесь размещаются также новостные сообщения, предложения о работе, рекламные объявления, афиши, различные рекомендации и советы туристам и жителям города и многое другое. Портал представлен в четырех языковых версиях – немецкий, английский, французский, итальянский языки. В связи с этим, использование машинного перевода может стать востребованным для русскоязычных посетителей сайта, не владеющих ни одним из перечисленных языков, но желающих получить официальную и самую актуальную информацию о столице Германии.

Проанализируем перевод, выполненный программой «Яндекс. Переводчик» и оценим эффективность ее работы. В работе предлагается анализ переводов, выполненных машиной с немецкого языка на русский, поскольку данный язык является официальным языком Германии в целом и Берлина в частности. Кроме того, интересно было проследить работу машинного переводчика в паре языков, не содержащей английского, поскольку с английским языком таких исследований

имеется уже достаточно много, а работ, выполненных на материале других языков, еще недостаточно.

Во-первых, стоит отметить, что программа выполняет перевод практически мгновенно, затрачивая на это не более 2-3 секунд. При первом быстром просмотре обнаруживается, что некоторые фрагменты остаются непереведенными, и это относится, прежде всего, к тексту который в немецкоязычной версии представлен на английском или других иностранных языках. Очевидно, что это связано с тем, что программа настроена только на одну пару языков, без учета возможности иноязычных вкраплений.

Подробный анализа транслята показывает, что лишь небольшой процент контента переведен на достаточно высоком языковом уровне, приблизительно 10-15% от общего объема. В основном машина неплохо справляется с простыми предложениями, короткими новостными заголовками, неперегруженными синтаксическими конструкциями. Большинство односложных названий рубрик переведены правильно:

Wirtschaft	*Экономика*
Tourismus	*Туризм*
Notdienste	*Экстренные службы*

Однако, следует отметить, что иногда именно короткие названия некоторых рубрик переведены некорректно, что связано, прежде всего, с многозначностью языковых единиц или с несовпадением частей речи в русском и немецком языках (например, *Kultur & Ausgehen – Культура и Выход,* где *выход* следовало бы перевести как *досуг*).

К адекватно выполненному переводу, понимание которого не вызывает затруднений, можно отнести также названия следующих рубрик и новостных заголовков:

Stellenmarkt für Berlin *Der Stellenmarkt von berlin.de - Finden Sie Ihren nächsten Job mit unserem Stellenmarkt.*	*Рынок вакансий для Берлина* *Вакансии от рынка berlin.de - найдите Вашу следующую работу с помощью нашего рынка труда.*
DDR Museum - Berlins interaktives Museum	*Музей ГДР - интерактивный музей Берлина*
Berlin Aktuell	*Берлин В Настоящее Время*
Berlin bietet kostenlosen digitalen Bibliotheksausweis	*Берлин предлагает бесплатный цифровой библиотечный билет*
Karneval der Kulturen in Berlin abgesagt	*Карнавал культур в Берлине отменен*
Wohnungssuche *Wohnungen, Häuser,*	*Поиск недвижимости* *Ищите и находите квартиры,*

Gewerbeimmobilien und mehr in Berlin suchen und finden!	*дома, коммерческую недвижимость и многое другое в Берлине!*
Berlin entdecken	*Откройте для себя Берлин*

Из приведенных примеров видно, что машинный переводчик справился с переводом данных фрагментов и в основном осуществил пословный перевод, сохранив синтаксис оригинала, что оказалось также приемлемым для русского языка. При этом был выбран наиболее контекстуально подходящий вариант перевода для многозначных слов. Кроме того, программа распознала сложные слова, которыми изобилует немецкий и перевела их словосочетаниями: *Stellenmarkt* – рынок вакансий, *Bibliotheksausweis* – библиотечный билет. Очевидно, что при переводе названия рубрики *Berlin Aktuell* программа выполняла пословный перевод и не связала его в единое целое, написав все слова с заглавной буквы. Это вызвано, скорее всего, тем, что прилагательное *aktuell* написано в оригинале с большой буквы. Такое употребление этого наречия является весьма частотным в заглавиях и названиях различного рода (например, *Deutsch Aktuell* - название рубрики на сайте Deutsche Welle). В последнем из приведенных примеров машинный переводчик подобрал удачный перевод для наречия в сравнительной степени – *mehr* (основное значение *более*) и перевел его словосочетанием *многое другое*, что является наиболее корректным вариантом для данного контекста, который был выбран по статистическому принципу из параллельных текстов, имеющихся в распоряжении у Яндекса в обширных многоязычных базах. Представляется, что перевод последнего примера, который весьма адекватно и стилистически корректно звучит на языке перевода, также был осуществлен благодаря поиску эквивалентов в корпусах текстов на русском и немецком языках.

Дальнейший анализ показал, что приблизительно 50% транслята можно отнести к продукту «приемлемого» качества. Восприятие такого текста не затруднено в смысловом отношении, но перевод содержит лексические и грамматические ошибки и неточности и стилистические нарушения. Чаще всего, такие нарушения языковой нормы связаны с многозначностью слов, с дословным переводом или непониманием машиной контекстуальных связей, лексической сочетаемости или скрытых подтекстов. Это наглядно демонстрируют следующие примеры:

Neue Primark-Filiale in den Gropius Passagen eröffnet	*Открылся новый филиал Primark в проходах Gropius*
Aussichtspunkte	*Наблюдательные пункты*
Mehrheit der Hotels und Gaststätten	*Большинство отелей и тракти-*

bangen um Existenz	*ров жаждут существования*

В первом примере налицо неправильный подбор значения для многозначной единицы *Passagen*, но получатель текста, даже не очень хорошо знакомый со столичным шоппингом, легко может предположить, что речь идет не о каких-то проходах, а о торговой галереи под названием Gropius, где размещаются различные магазины. А для пользователей, которым известна сеть магазинов ирландского ритейлера Primark, имеющая отделы по всей Европе, понимание данного заголовка не вызывает затруднений. Более того, при дальнейшем просмотре данного новостного сообщения, это становится очевидным. Второй пример представляет собой название рубрики в разделе *Достопримечательности*, из чего становится ясным что имеются ввиду *смотровые площадки* или другие места, откуда открывается красивый вид. Примечательно, что при переходе по ссылке для получения более подробной информации в данной рубрике, мы находим уже вариант *Смотровые площадки в Берлине*. В последнем из приведенных примеров получатель понимает, что программа использовала неподходящий эквивалент перевода для глагола *bangen um – опасаться за*. Здесь, скорее всего, был осуществлен поиск по статистическому принципу, и он сработал в отрицательную сторону, поскольку основное значение глагола *bangen*, которое зафиксировано в большинстве словарей, именно *опасаться, бояться чего-л*, а машина, видимо, нашла в немецко-русском корпусе параллельных текстов некий контекст, в котором данный глагол был переведен как *жаждать*, и осуществила механическую подстановку. Но тем не менее, для получателя понятно, что данные заведения хотят продолжить свое существование, а не быть закрытыми.

Переводы следующих фрагментов также не вызывают затруднений при их восприятии:

Ob Firmenevent, Produktpräsentation oder Hochzeitsfest: Berlin bietet Veranstaltungsorte in jeglicher Größe und für jeden Bedarf.	*Будь то корпоративное мероприятие, презентация продукта или свадебный пир: Берлин предлагает места любого размера и для любых потребностей.*
Im Januar holt die Berliner Stadtreinigung die ausgedienten Weihnachtsbäume ab.	*В январе Берлинская городская уборка забирает отработанные елки.*

В первом примере Яндекс.Переводчик выбрал контекстуально и стилистически некорректный эквивалент *свадебный пир* для лексической единицы *Hochzeitsfest*, но получатель понимает, что речь идет о помещениях для проведения свадебных торжеств. Вторая часть дан-

ного фрагмента переведена на русский язык также с нарушением языковых норм, но и здесь ясно, что описывается возможность подбора помещения для проведения любых мероприятий. Во втором примере адресат легко восстанавливает пропущенное в переводе слово *служба* и без труда понимает, что имеется в виду муниципальная служба, осуществляющая уборку улиц и других общественных мест в городе.

Одной из наиболее частотных ошибок является дословный перевод с нарушением лексической сочетаемости. Что мы наблюдаем, например, при переводе прогноза погоды:

Bedeckt mit leichtem Schneefall *Покрыто легким снегопадом*

Машинный переводчик выполнил дословный перевод и соединил прилагательное *легкий* и существительное *снегопад*, которые в русском языке не сочетаются лексически, в одно словосочетание. Семантика слова *снегопад* подразумевает выпадение обильных осадков в виде снега, и снегопад априори не может быть легким. Переводчик также не распознал синтаксические связи и значение предлога в этом фрагменте и *bedeckt mit* перевел как *покрыто чем-то,* передавая значение инструмента или средства, при помощи которых совершается действие, а не значение сопровождения, совместного действия. Кроме того, для причастия *bedeckt* также выбрано неверное контекстуальное значение для описания погодных явлений (*покрыто* вместо *облачно*).

Вот еще два примера дословного перевода и демонстрации неумения программы выполнять адекватные замены:

Bis Ostern macht keine Bühne auf	*До Пасхи не будет сцены*
Krankenhäuser fordern freie Hand bei Corona-Mutations-Tests	*Больницы требуют свободной руки в тестах на мутацию короны*

В первом примере переводчик дословно перевел существительное *Bühne*, не распознав, что здесь имеются в виду *представления*, которые до Пасхи отменены из-за пандемии коронавируса. Но адресат, знакомый с ситуацией, понимает смысл переведенного новостного заголовка. Во втором примере мы видим кальку *свободная рука*, хотя речь идет о *свободе действий* при проведении тестов на коронавирус, что также становится ясным при прочтении дальнейшего текста.

Приблизительно 35% переведенного контента содержат грубые ошибки, полностью или частично искажающие смысл текста оригинала. Они вызваны тем, что современный уровень развития машинного перевода еще не позволяет выявлять скрытый подтекст, не всегда отслеживает анафорические и логические связи в предложении, зачастую не распознает метафору, эллипсис и также, как и в переводе

приемлемого качества выбирает абсолютно неподходящие значения многозначных слов. Рассмотрим несколько примеров:

Termine des Regierenden Bürgermeisters	*Даты правящего мэра*
Gegen 16.30 Uhr bemerkte eine Zivilstreife auf der Siriusstraße in einem Personenkraftwagen vier Personen, die alle keinen Mund-Nasen-Schutz trugen.	*Против 16.30 гражданский патруль заметил на Sirius улица легковых автомобилей четыре человека, которые несли все нет рта-носа-защитная.*
Mann funkt Flugzeugpiloten an: Gibt gefährliche Anweisungen	*Человек искрится пилотам самолета: дает опасные инструкции*

В первом фрагменте машина выбрала абсолютно некорректный вариант перевода из имеющихся значений для слова *Termine*. Из названия данной рубрики пользователю незнающему немецкий язык может быть абсолютно непонятно, о каких датах идет речь и, не раскрыв и не просмотрев всю рубрику, вряд ли он сможет догадаться, что здесь имеется в виду график рабочих встреч бургомистра. В следующем примере программа также выбрала неподходящий для данного контекста эквивалент перевода для предлога *gegen*, который в сочетании с указанием на время, имеет значение *около*, *приблизительно*. Кроме того, в этом предложении нарушены всякие синтаксические и логические связи. Единственное число существительного *Personenkraftwagen* программа перевела множественным числом несмотря на то, что перед ним стоит неопределенный артикль, который употребляется только с существительными в единственном числе. Здесь также опущен перевод предлога *in*, что нарушило дальнейшие синтаксические корреляции. О смысле придаточного предложения остается только гадать, поскольку оно представляет собой набор непонятных и несвязанных слов. Хотя в настоящих реалиях можно предположить об отсутствии масок, защищающих органы дыхания в эпоху пандемии коронавируса, но тем не менее относить это предложение к переводу приемлемого качества нельзя. Также название улицы *Siriusstraße* разделено на две части, первая из которых оставлена без перевода на языке оригинала, а вторая переведена, но это бы не вызвало бы недоумения, если бы все остальные элементы в предложения были переданы более или менее корректно. Перевод последнего примера, выполненный Яндекс.Переводчиком также не способствует пониманию смысла, поскольку программа не распознала верное значение глагола *anfunken – связаться по рации,* потеряв, по всей видимости, отделяемую приставку *an*. Отделяемые приставки в главных предложениях стоят в конце предложения, а здесь возможно сбой

произошел из-за использования знака препинания, не сигнализирующего о конце предложения, а именно, двоеточия.

Еще одной достаточно частотной ошибкой, которая обнаружилась в ходе анализа, является путаница с переводом имен собственных. Так с одной стороны, машина может переводить имена собственные как имена нарицательные, что является весьма ожидаемым в отношении немецких фамилий, поскольку многие из них зачастую имеют форму общеупотребительных лексем:

3D Rundgang: Eintauchen in die Installation «I hope» der japanischen Künstlerin Chiharu Shiota in der Berliner König Galerie.	*3D экскурсия: погружение в инсталляцию «Я надеюсь» японского художника Чихару Шиоты в берлинской галерее короля.*

В данном примере фамилия *König* переведена дословно: *король*, хотя здесь имеется в виду Йоханн Кениг, основавший данную галерею. Стоит отметить, что иностранные имена и фамилии программа по большей части транслитерирует, как она это сделала в этом же примере с именем Чихару Шиоты.

С другой стороны, машинный переводчик иногда переводит имена нарицательные как имена собственные:

Karneval der Kulturen will klein feiern	*Карнавал культур хочет отпраздновать Клейн*

В этом примере прилагательное *klein* (*маленький*) транслитерировано, написано с заглавной буквы и выглядит в тексте транслята как имя собственное. Возможно, здесь сыграл роль статистический поиск, в ходе которого в корпусах текстов было обнаружено такое имя собственное, или же программа не смогла выполнить переводческую трансформацию и передать смысл того, что карнавал культур не хотят праздновать широкомасштабно.

Помимо проанализированных встречаются также и прочие ошибки, которые либо частично нарушают смысл, заключенный в оригинале, либо полностью искажают его и практически не поддаются какой-либо классификации и описанию. Приведем для наглядности один пример такого искажения:

Lärmschutz: Erneut liegen Planungs-Unterlagen aus	*Снова взвыл он*

Действия машины при переводе данного фрагмента невозможно интерпретировать, поскольку из элементов, имеющихся в оригинале, Яндекс.Переводчик перевел только слово *erneut* - *снова*, остальное же абсолютно не имеет ничего общего с тем, о чем говорится в оригина-

ле. Стоит отметить, что процент таких переводов невелик, но они все же присутствуют и вызывают полное непонимание у получателей.

Проведенный анализ показывает, что, несмотря на имеющиеся недостатки, развитие машинного перевода находится уже на достаточно высоком уровне. Его применение допустимо для перевода текстов определенного жанра и типа, например, для рассмотренных нами текстов урбанистического дискурса, которые не перегружены специализированной терминологией, как, например, научные или технические публикации или образностью, как произведения художественной литературы. Полученный транслят может быть использован заинтересованными пользователями в целях быстрого ознакомления и получения общего представления о контенте, размещаемом на сайте. Однако, следует отметить, что перевод такого качества ни в коей мере не может претендовать на безупречность и абсолютную корректность. Из этого следует вывод, что машина и искусственный интеллект пока еще не могут составить конкуренцию профессиональным переводчикам, но при грамотном использовании и наличии навыков постредактирования могут быть для них полезными инструментами, ускоряющими и оптимизирующими их работу. Кроме того, системы машинного перевода становятся всё более популярными среди большого количества пользователей интернета, предоставляя им доступ к информации на самых разных языках мира, тем самым открывая им новые горизонты и разрушая языковые барьеры.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Ершов, А. Google сделал рекордно точную систему машинного перевода на основе нейросетей / А. Ершов // N+1 Интернет-издание. [Электронный ресурс]. – Режим доступа: https://nplus1.ru/news/2016/09/28/neurogoogletrans - Дата доступа: 11.02.2021.

2. Нелюбин, Л.Л. Толковый переводоведческий словарь / Л.Л. Нелюбин - 3-е издание, переработанное. — М.: Флинта: Наука, 2003. – 318с.

3. Berlin.de Das offizielle Hauptstadtportal. [Электронный ресурс]. – Режим доступа: http://www.berlin.de/ - Дата доступа: 12.02.2021.

4. Panasenkov, N.A., Korneeva L.I. Groundbreaking Technologies in Machine Translation / N.A. Panasenkov, L.I. Korneeva // Язык в сфере профессиональной коммуникации: сборник материалов международной научно-практической конференции преподавателей, аспирантов и студентов (Екатеринбург, 28 мая 2020 г.). – Екатеринбург: ООО «Издательский Дом «Ажур», 2020. – С. 584-589.

EVALUATION OF THE QUALITY OF MACHINE TRANSLATION
(based on the material of the city portal)

A. A. Novozhilova, S. A. Korolkova
Volgograd State University, Russia
E-mail: novozilova@volsu.ru, korolkova@volsu.ru

The article is devoted to the analysis of the Yandex.Translate machine translation system and its use on the Internet for translating texts of city portals. The authors evaluate the current state of machine translation systems, analyze the errors made by the program while translating urban texts from German into Russian.

Keywords: information technologies, machine translation, translation quality, semantic errors, urban discourse