

П.С. Митин

студент факультета государственного управления
Гжельского Государственного университета;
e-mail: pro.inter2017@yandex.ru

ТИПИЧНЫЕ ОШИБКИ МАШИННОГО ПЕРЕВОДА ОБЩЕПОЛИТИЧЕСКОЙ ЛЕКСИКИ (НА ПРИМЕРЕ ПЕРЕВОДА ПРЕССЫ АРАБСКОЙ РЕСПУБЛИКИ ЕГИПЕТ)

В предлагаемой статье автор предпринял попытку исследования наиболее типичных ошибок “Google Translate (GT)” — при машинном переводе (МП)¹ с арабского языка на русский информативных новостных текстов. В ходе исследования на основе анализа оценивалась целостность и семантическая адекватность результатов перевода при определении его качества с целью выявления степени необходимости профессиональной коррекции квалифицированным переводчиком готового машинного перевода.

В своём исследовании автор использовал метод контент-анализа, а также количественный и качественный описательный подходы для оценки качества ошибок перевода. В качестве исследуемого материала были отобраны примеры перевода из интернет-газет.

Ключевые слова: машинный перевод, программа, нейронная сеть, идентификация, коррекция, газета, типология, орфография, лексика, семантика, морфология, омонимия, упущение, редактирование.

За последние годы существенно вырос объём переводимых документов, при этом тенденция увеличения объёма машинного перевода продолжает увеличиваться.

Это обусловлено целым рядом факторов:

- экономическим развитием нашей страны;
- ростом делового сотрудничества российских компаний с зарубежными партнёрами;
- увеличением объём деловой корреспонденции, контрактов, бизнес-планов и другой коммерческой информации.

Технический прогресс требует не только увеличения объёмов технического перевода, но и сокращение сроков перевода информации.

¹ Далее — «машинный перевод» или МП.

Следовательно, без автоматизации части работ и использования искусственного интеллекта и внедрения систем машинного перевода практически невозможно выполнить требования, которые диктует рынок.

Термин «машинный перевод» рассматривается учёными в нескольких смыслах.

Так, например, специалисты компании «Промт» дают следующее определение: «Машинный перевод (Machine Translation, МТ) — это технология связного перевода текстов компьютерной программой с одного естественного языка на другой»².

При этом, машинный перевод рассматривается как в узком, так и в широком смысле. Машинный перевод в узком смысле — это процесс перевода некоторого текста с одного естественного языка на другой, реализуемый компьютером полностью или почти полностью. В ходе данного процесса на вход машины подаётся текст, словесная часть которого не сопровождается никакими дополнительными указаниями, а на выходе получается текст на другом языке, являющийся переводом входного. Преобразование входного текста в выходной происходит без вмешательства человека.

Машинный перевод в широком смысле — это область научных исследований, находящаяся на стыке лингвистики, математики, кибернетики, и имеющая целью построение систем, реализующих машинный перевод.

Своё развитие система перевода получила в конце 40-х — начале 50-х гг.

Характеризуя периоды развития теории перевода, Н.К. Гарбовский пишет: «Третий период, современный, начинается в сороковые годы появлением первых статей по теории машинного перевода. Начало этого периода Стейнер связывает с именами русских и чешских учёных, которые, унаследовав, по его мнению, идеи формализма, пытались применить лингвистическую теорию и статистические методы к исследованию перевода. В этот период предпринимаются попытки установить соответствие между формальной логикой и моделями языковых трансформаций. Период отмечен интенсивными научными разысканиями в области перевода» [Гарбовский, 2015: 6].

Датой рождения машинного перевода как научной проблемы учёные считают 1946 год, когда директор отделения есте-

² Статья: Нейросетевой машинный перевод: революция на наших глазах. <https://www.promt.ru/technology/neural-machine-translation/>

ственных наук Рокфеллеровского фонда Уоррен Уивер впервые сформулировал концепцию машинного перевода в меморандуме “Translation”.

В концепции отмечалось, что целью использования систем машинного перевода является стремление заменить человека в данной сфере деятельности. Целостная система машинного перевода “IBM Mark II”, созданная компанией IBM совместно с Джоржтаунским университетом была представлена в Нью-Йорке в 1954 году³.

С развитием вычислительной техники в конце 70-х годов (появление микрокомпьютеров, развитие сетей, увеличение ресурсов памяти) машинный перевод вошёл в эпоху «Ренессанса».

В течение последующих десятилетий бурное развитие персональных компьютеров и информационных технологий, появление сети Интернет (которая стала интернациональной и многоязыкой) сделали возможным использовать достижения не только точных наук, но и лингвистики, филологии и языкознания для развития систем машинного перевода.

Как вид языковой деятельности перевод затрагивает все уровни языка — от распознавания программой графем (или фонем при переводе устной речи) до передачи смысла высказывания и текста. Кроме того, для перевода характерна обратная связь и возможность сразу проверить теоретическую гипотезу об устройстве тех или иных языковых уровней и эффективности предлагаемых алгоритмов.

Лингвистическая сторона машинного перевода всё больше привлекает внимание теоретиков, в результате чего возникают новые теории автоматизации перевода и формализации языковых процессов.

«Сегодня рождается новое понятие — «цифровой перевод», определяющее новый вид технологии перевода, систему сетевого взаимодействия переводчика и цифровых информационно-коммуникационных средств, искусственного интеллекта, призванный повысить эффективность переводческого искусства и качество переводческой продукции [Гарбовский, Костикова, 2019: 3].

Возрастающая роль практики машинного перевода в современном мире обусловлена также необходимостью условия межъязыковой коммуникации. С этой точки зрения можно утверждать, что альтернативы машинному переводу нет и разработка качественных и высокопроизводительных систем перевода будет способствовать

³ «Википедия» — интернет-ресурс.

разрешению важнейших социально-коммуникативных задач. Аргументами данного положения являются:

- высокая скорость перевода;
- низкая стоимость;
- конфиденциальность;
- универсальность.

Системы машинного перевода могут быть классифицированы по трём признакам:

- *нейронные системы машинного перевода;*
- *машинный перевод на основе правил;*
- *статические системы машинного перевода.*

Системы машинного перевода, построенные на базе нейронных сетей, продемонстрировали сразу несколько компаний, среди которых “Google”, “Microsoft” и “SYSTRAN”. Начиная с 2016 г. система “Translate” проделала путь от статических систем машинного перевода до нейронных. Для прогнозирования вероятности корректного употребления слов нейронная система использует искусственную нейронную сеть, обычно моделируя целые предложения. Эта система не требует большой объёма оперативной памяти. В основе работы нейронного переводчика заложен механизм двунаправленных рекуррентных нейронных сетей (Bidirectional, Recurrent, Neural, Networks), построенный на матричных вычислениях, который позволяет формировать существенно более сложные вероятностные модели, чем статистические машинные переводчики.

Нейронные системы в процессе перевода используют машинное обучение. В процессе обучения нейронные системы оперируют не отдельными фразами и словосочетаниями, а целыми предложениями. Основная проблема заключается в том, что для тренировки такой системы требуется существенно больше вычислительных мощностей. При этом тренировка одной нейронной модели требует от 1 до 3 недель, тогда как статистическая модель такого же размера настраивается за 1–3 дня.

Преимуществом использования “Google Translate” как системы машинного перевода является бесплатный доступ. Сильной стороной “Google Translate” является возможность взаимодействия с переводческими базами данных, т.е. с переводами, которые были добавлены пользователями. Исследователи считают, что GT является наиболее популярной системой машинного перевода. С момента создания данной системы ею воспользовались больше 200 млн человек⁴.

⁴ <https://blog.google/products/translate/ten-years-of-google-translate/>

Специалисты считают, что в большинстве случаев нейронные системы машинного перевода позволяют получать более качественный конечный перевод по сравнению со статическими системами машинного перевода. Преимущество нейронных систем машинного перевода заключается в том, что они используют искусственный интеллект для одновременного отображения всего предложения на необходимом языке.

В связи с этим К. Хандшух (2013) во время своего исследования машинных систем перевода использовал следующие онлайн-системы машинного перевода для сравнения конечного перевода с немецкого на английский язык, а именно: “Google Translate”, “SYSTRAN” и Вавилон. По результатам своего исследования он сделал вывод о том, что перевод, получаемый в результате использования систем машинного перевода, является недостаточно качественным [Handschuh, 2013]. М.Х. Кешаварц (1999) утверждал, что основными лексико-семантическими ошибками машинного перевода являются: неправильное использование времен, ошибки в использовании групп глаголов, некорректный порядок слов, некорректное использование предлогов [Keshavarz, 1999].

На сегодняшний день, ряд исследователей (например: De Camp; Vilar Xu, D’Haro & Ney) показали, что одной из наиболее часто встречающихся ошибок машинного перевода является пропуск слов, а другая группа учёных (например, Gaspari, Toral & Naskar; Valotkaite & Asadullah) установила, что наиболее частой ошибкой при работе с системами машинного перевода является искажение смысла перевода.

Лингвисты Лотц и Ренсбург (2016) провели исследования двух типов текстов, а именно: тексты из презентаций и тексты из онлайн-репортажей. В ходе исследования, проведённого в 2010–2013 гг., для определения качества перевода ГТ учёные использовали одинаковые тексты, и они сделали вывод о том, что основными ошибками являются упущения, некачественный перевод, пропуски и несоответствие грамматическому строю исходного предложения [Lotz, Rensburg, 2014].

Для осуществления процесса перевода используют различные типы текстов. Согласно исследованию К. Райс (1977–1989 гг.) тексты можно разделить на три группы: экспрессивные, информативные и оперативные [Reiss, 1977/1989]. Информативными являются те тексты, которые несут в себе информацию, знания, имеют чёткую структуру. Основное внимание уделяется теме данного текста. Примерами аналогичных текстов могут служить научные и новостные

статьи. Экспрессивные тексты — те тексты, которые имеют эстетическую ценность. Во время перевода подобного типа текстов основное внимание должно уделяться форме, ритму и другим эстетическим составляющим. К эстетическим и оперативным типам текстов можно отнести рекламные материалы, т.к. они побуждают либо ограждают читателя от каких-либо действий (к покупке, продаже и т.д.). Основываясь на данной классификации, был сделан вывод о том, что информативные тексты в большей степени подходят для машинного перевода, т.к. основное внимание данного типа текстов направлено на их название. Таким образом, мы приходим к выводу о том, что для работы с машинным переводом наиболее подходящим типом текстов являются новостные статьи.

По мнению учёного — доктора филологических наук Н.К. Гарбовского, существуют две основные классификации видов перевода:

а) по характеру переводимых текстов (жанрово-стилистические особенности оригинала);

б) по характеру речевых действий переводчика в процессе перевода (формы перевода).

«Жанровая» — это специализация отчётливо проявляется в среде профессиональных переводчиков, отмечает Н.К. Гарбовский: «Весь рынок труда переводчиков разделён на следующие сферы общественно-экономической деятельности:

- 1) искусство и литература;
- 2) коммуникация;
- 3) экономика, политика, управление;
- 4) право;
- 5) техника и промышленность;
- 6) гуманитарные науки;
- 7) истинные и прикладные науки;
- 8) туризм и развлечения» [Гарбовский, 2007: 61–62].

Жанрово-стилистическая классификация переводов в зависимости от жанрово-стилистических особенностей оригинала выделяет три функциональных вида перевода: художественный; общественно-политический; специальный. Объектом художественного перевода являются художественные произведения. Основная задача любого художественного произведения заключается в достижении образно-эмоционального и эстетического воздействия на читателя. В целях достижения определённого эстетического воздействия на переводной язык используется огромное количество разнообразных языковых средств, от эпитета (красочное определение) до ритмико-синтаксического построения фразы. Такая эстетическая

направленность отличает художественный перевод от остальных видов перевода. Машинному переводу по указанным причинам художественные тексты поддаются труднее всего.

Общественно-политическим переводом называется перевод текстов общественно-политического и публицистического характера с пропагандистской или агитационной установкой. Общественно-политический перевод характеризуется яркой эмоциональной окраской с большой насыщенностью различной терминологией.

Специальный перевод обслуживает различные предметные отрасли знаний, имеющие специфическую терминологическую номенклатуру. Объектом специального перевода являются материалы, которые относятся к различным сферам человеческого знания и практики науки и техники. Эти материалы характеризуются предельно точным выражением мысли, следовательно, широким использованием терминологии. Такой перевод проще всего реализуем в машинном виде, поскольку специальные тексты характеризуются строгой логикой построения текста, жёсткой структурой предложений, отсутствием эмоциональной окраски и подтекста. Под формами перевода понимается способ, при котором осуществляется перевод: письменный (письменно-письменный, зрительно-письменный, письменный перевод на слух); устный (устный перевод на слух, зрительно-устный перевод или перевод с листа, т.е. устный перевод зрительно воспринятого исходного письменного текста). Машинной реализации проще поддаётся письменный перевод, поскольку устный требует решения дополнительной задачи — распознавания и синтеза устной речи.

Для сбора информации, поиска ошибок, их коррекции и идентификации был использован описательный метод анализа. Исследователь также оценил и дал количественную оценку полученных ошибок.

Материалом для анализа информационных текстов послужили электронные версии газет «Анадолу» и «Аль-Масрун». «Анадолу» была выбрана в качестве одного из источников информативных текстов из-за высокого профессионализма канала и высокого уровня арабского языка. Что касается «Аль-Масрун», то он был выбран из-за стандартизированного языка, используемого в большинстве его новостных статей. Для осуществления перевода исходных текстов был использован “Google Translate”, так как он является бесплатным и находится в широком доступе.

Для анализа автор использовал модель Дж. Хсу (2014), которая отображает более подробную типологию ошибок (см. табл. 1).

Классификация ошибок систем машинного перевода

Виды ошибок				
Морфологические ошибки	Орфографические	Лексические	Синтаксические	Сематические
<p>Неверное употребление рода существительных.</p> <p>Объединение простой и сложной форм сравнения.</p> <p>Использование полной формы прилагательного в функции сказуемого.</p> <p>Неверное склонение числительных.</p> <p>Ошибка в сочетании дробного числительного с существительным.</p> <p>Неверное употребление собирательных числительных.</p> <p>Неверное употребления личного местоимения, создающее двусмысленность.</p> <p>Неверное употребление форм глаголов.</p> <p>Несоблюдение единства видо-временных форм глаголов.</p>	<p>Исключения из правил.</p> <p>Написание большой буквы в составных собственных наименованиях.</p> <p>Случаи раздельного и слитного написания не с прилагательными и причастиями.</p> <p>В написании и и ы после приставок.</p> <p>Трудные случаи различения не и ни (Куда он только не обращался! Куда он ни обращался, никто не мог дать ему ответ).</p>	<p>Лексическая несочетаемость.</p> <p>Неоправданный пропуск слова.</p> <p>Многословие: - неоправданный повтор слова; - тавтология; - плеоназм.</p> <p>Неоправданное употребление взаимственных слов, устаревшей лексики неологизмов, жаргонизмов.</p> <p>Неверное использование многозначных слов омонимов, паронимов.</p> <p>Смещение понятий.</p> <p>Неверный выбор синонимов.</p>	<p>Ошибки с предложениями с причастным оборотом.</p> <p>Ошибки в предложениях с деепричастным оборотом.</p> <p>Ошибки в способе построения предложения.</p> <p>Смещение прямой и косвенной речи.</p> <p>Нарушение границ предложения.</p>	<p>Некорректное использование переменных (до инициализации, использование индексов, выходящих за границы массивов и т.п.).</p> <p>Ошибки вычисления (некорректное использование целочисленной арифметики, незнание приоритетов выполнения операции, деление на 0, извлечение корня из отрицательного числа и т.п.).</p> <p>Ошибки межмодульного интерфейса (игнорирование системных соглашений при передаче параметров, нарушение области действия локальных и глобальных переменных, и т.п.).</p>

Данная модель классифицирует ошибки на орфографические, морфологические, лексические, семантические и синтаксические.

В своём исследовании автор стремился определить типичные ошибки GT при переводе информативных текстов с арабского на русский.

Автором выявлены следующие типы ошибок.

Пункт 1. Типы ошибок.

Синтаксические ошибки. Как показано в табл. 1, синтаксические ошибки могут возникать из-за опечаток в статьях, неправильной расстановки союзов, предлогов, нарушенного синтаксического порядка слов и их категорий. Примеры таких ошибок приведены ниже.

Пример 1

Исходный текст:

بدأ، مساء الأحد، إغلاق صناديق الاقتراع في الانتخابات الرئاسية المصرية بالخارج، التي يتنافس فيها مرشحان أحدهما الرئيس الحالي عبد الفتاح السيسي⁵.

Полученный машинный перевод: В воскресенье вечером избирательные участки закрылись на египетских президентских выборах за границей, в которых участвуют два кандидата, в том числе нынешний президент Абдель Фаттах Эль-Сиси.

Из примера 1 видно, GT обеспечил точный перевод исходного текста. В целом смысл понятен, однако синтаксис в некоторой степени искажён, что приводит к двусмысленности. Смысл был бы яснее, если переформулировать предложение «В воскресенье вечером на египетских избирательных участках за рубежом завершилось голосование по случаю выборов главы государства», а не «В воскресенье вечером избирательные участки закрылись на египетских президентских выборах за границей». Общий смысл перевода также был бы более ясным, если его разделить на два предложения, которые говорят о двух различных событиях, происходящих в разных районах мира, хотя и явно взаимосвязанных. И возможно, указанные разные формы времени служат для того, чтобы запутать читателя — совместить различные события и места действия. Тема первой части — событие произошло в прошлом времени. Тема второй части — «два конкурирующих кандидата», и ход процесса происходит в настоящем времени. Читатель теряется: «Неужели выборы происходят в Египте? Выборы за границей или

⁵ <https://www.aa.com.tr/ar/إغلاق-مراكز-إغلاق-مصر/العربية-الدول/1100172>

опросы за границей? Неясно, закрылись ли заграничные избирательные участки, на которых египетские избиратели, живущие за границей, могут голосовать во время выборов на родине? Закрылись/прекратились ли одновременно и заграничные выборы, и выборы в Египте? Конкурируют ли кандидаты за границей?

Другими словами, из-за представленного перевода смысл может быть истолкован так, что кандидаты за рубежом конкурируют в египетских выборах или в опросах, когда на самом деле два кандидата конкурируют за должность в Египте и оба находятся в этой стране. Существуют египетские представительства, задачей которых является регистрация голосов египетских граждан, проживающих за рубежом. Если бы результат был более чётко выражен с точки зрения синтаксиса, не было бы никакой путаницы или двусмысленности.

Предлагаемый отредактированный вариант перевода: **«В воскресенье вечером, на египетских избирательных участках за рубежом завершилось голосование по случаю выборов главы государства. В данных выборах участвовали два кандидата, в том числе и действующий президент Абдель Фатах Ас-Сиси».**

Семантические ошибки

Семантические ошибки возникают в результате выбора неэквивалентного варианта для перевода исходного слова. Это может произойти из-за лингвистической двусмысленности лексем исходного слова. Лингвистическая неоднозначность может возникнуть из-за гомофонии или синонимии исходного слова. Примеры семантических ошибок полученного текста после GT приведены ниже.

Пример 2

Исходный текст:

وتجرى الانتخابات الرئاسية دون وجود لمرشحين معارضين أو منافسين بارزين؛ جراء انسحابات سابقة من السباق الرئاسي متعلقة بالمشهد السياسي، لا سيما لكل من المرشحين المحتملين، خالد علي، ومحمد أنور السادات، والفريق المتقاعد، أحمد شفيق⁶.

Полученный машинный перевод: Президентские выборы проводятся без присутствия видной оппозиции или конкурентов; в результате предыдущих уходов из президентской гонки, связанных с политической сценой, особенно для каждого из потенциальных

⁶ <https://www.aa.com.tr/ar/العربية/الدول-الثاني/اليوم-الثاني-الرئاسيات-لليوم-الثاني/العربية/1091452>.
ج-يو اصلون-الإدلاء-بأصواتهم-في-الرئاسيات-لليوم-الثاني/العربية/1091452. مصريو-الخار

кандидатов, Халеда Али, Мухаммеда Анвара Садата и отставной команды Ахмеда Шафика.

Автором выявлена семантическая ошибка в переводе слова «الفريق», которое означает «генерал-майор»: в результате слово «генерал-майор» программа машинного перевода перевела как «команда». Эта ошибка произошла из-за того, что исходное слово «الفريق» является омонимом другого арабского слова, которое означает «команда». Омонимия часто вызывает языковую неоднозначность, которая приводит к семантическим ошибкам. Другая семантическая ошибка произошла в отбрасывании исходного слова в тексте полученного машинного перевода «جراة» (в результате), что повлияло на значение. Выражение исходного слова подразумевает, что отсутствие кандидатов и конкурентов на выборах было связано с уходом кандидата с выборов.

Лексические ошибки

Лексические ошибки, как показано в табл. 1, возникают из-за некорректного пропуска необходимых лексем или добавления ненужных / неправильных лексем или выражений. Примеры таких ошибок в выводах машинного перевода приведены ниже.

Пример 3

Исходный текст:

وفق قرار الهيئة الوطنية للانتخابات بمصر (رسمية)، تنتهي مساء الأحد، عملية الاقتراع في اليوم الثالث والأخير لتصويت المصريين بالخارج في التاسعة مساء بتوقيت كل دولة⁷.

Полученный машинный перевод: В соответствии с решением Национального избирательного органа в Египте (официально), в воскресенье вечером выборы в воскресенье закончатся на третий и последний день голосования египтян за границей в девять часов вечера по каждой стране.

Как видно из примера, полученный с помощью GT текст оставил некоторые части исходного текста непереведёнными. Ещё одна проблема, слово «воскресенье» является неопределённым. Неизвестно когда решение было принято. Они встретились в воскресенье и приняли решение в тот день? Или об этом только что сообщили в воскресенье? Кроме того, добавление пропущенного слова «жи-

⁷ https://msryon.com/Section_479-الاقتراع-في-رناسيات-مصر-بالخارج/الحياة-السياسية/
إغلاق-صناديق_1158122

вущий» для выражения «египтяне, проживающие за границей» звучит лучше.

Здесь следует указать, что в результате этого упущения значение смысла искажается. Похоже, что переведённая версия говорит о том, что процесс голосования для египтян, которые находятся за границей, закончится в 9 часов вечера. Но что делать в другое время, если голосование в Египте и за рубежом продлится ещё три дня? Отсутствие уточнения, кто, где находится и когда, что происходит, создаёт путаницу. Абсолютная конкретность необходима в каждой части предложения. Ссылки не могут быть неоднозначными, так как часовые пояса в разных странах могут быть различными. Именно это отсутствие конкретики, неправильное использование слов, путаница в ссылках и т.д. затрудняет читателям понимание истинного значения смысла.

Пример 4

Исходный текст:

قال الكاتب الصحفي مصطفى بكرى عضو مجلس النواب، إن الفريق أحمد شفيق تعجل في إعلان ترشحه للرئاسة، حيث كان مقرر إعلان ذلك من فرنسا يوم ديسمبر القادم⁸.

Полученный перевод: Журналист Мустафа Бакри, член Палаты представителей, заявил, что команда Ахмеда Шафика поспешила выдвинуть его кандидатуру на пост президента, поскольку было запланировано выдвинуть её от Франции 22 декабря.

В данном примере ошибка заключается в переводе слова «الفريق» как «команда». Как было отмечено автором выше, это является некорректным переводом исходного слова. Существует также семантическая двусмысленность анафорической ссылки, где «его» может относиться, как к М. Бакри, так и к А. Шафику. Эта ошибка могла быть исправлена переводом «генерал-майор А. Шафик поспешно выдвинул свою кандидатуру», а не «его». Опять же, отсутствие конкретики порождает отсутствие ясности и возникновение путаницы.

Морфологические ошибки

Морфологические ошибки встречаются на уровне глаголов, существительных и других аспектов морфологии. Пример такой отмечен автором в примере 4. Так, глагол *تعجل* ST, который является глаголом прошедшего времени, был переведён как «спешит»,

⁸ <https://www.almasryalyoum.com/news/details/1226294>

который является глаголом настоящего продолженного времени. В примере 1 лексические элементы ST إغلاق بدأ были представлены как «закрытые», что является морфологической ошибкой, поскольку существительное исходного текста إغلاق было представлено как глагол, а глагол исходного текста بدأ был опущен. Пропуск исходного глагола بدأ в полученном переводе отражает лексическую ошибку.

Основываясь на представленном анализе машинного перевода, проведённое автором исследование показывает, что встречаются такие типичные ошибки, как упущение и неправильный лексический выбор, искажение морфологии. Таким образом, автор приходит к выводу о необходимости редактирования машинного перевода профессиональным переводчиком. Вместе с тем, данное исследование также показывает, что GT является полезной системой, которая помогает ускорить процесс перевода.

Одним из путей совершенствования машинного перевода и избежания типических ошибок, по мнению автора, является использование корпусной лингвистики. «Уникальность корпусной лингвистики заключается в использовании современных компьютерных технологий, в сборе и структурировании большого объёма языковых данных, в методах, применяемых для их обработки, а также предполагаемых стратегий их применения во всех видах деятельности, связанных языком, в частности, в процессе перевода, и исследованиях результатов перевода и их оценке» [Цзинь Ифан, 2019: 4].

Так, например, национальный корпус русского языка включает целый ряд подкорпусов (основной, синтаксический, корпус современных СМИ, а также параллельные корпуса, «в которых можно найти все переводы для определённого слова или словосочетания на русский язык или с русского языка» [Там же: 6].

Хотя методика формирования корпусов текстов на материале индоевропейских языков, в том числе и русского, в значительной мере разработана в отношении других языков, в частности арабского, она далека от окончательного решения. Несмотря на формирование в последние годы корпусов текстов на различном материале, в частности на базе текста Корана, что имеет большое научное значение, арабский мир «не располагает арабским лингвистическим корпусом, сопоставимым с английским».

В значительной мере всё это объясняется тем, что многие решения, эффективные при обработке материала европейских языков с алфавитным письмом, богатой парадигмой словообразования и чётко выраженными на графике границами слов, оказываются

малопригодными для восточных языков с иероглифической письменностью, а также языков, письменность которых хотя и построена на алфавитном принципе, однако не всегда позволяет четко маркировать границы между словами, где критерии выделения слов могут трактоваться по-разному [Редькин, 2014].

Заключение

Подводя итоги, можно сделать вывод о том, что современные программы снабжены ёмким словарём и чёткими алгоритмами перевода, но восприятие общего смысла текста всё же часто затрудняется, требуя корректировки грамматических и лексических форм. Причиной тому является отсутствие у программы-переводчика так называемого «чутья» и осознания излагаемого смысла. Поэтому необходима работа профессионального переводчика и постредактирование. Автор считает, что наиболее эффективных результатов машинного перевода можно достичь для текстов, написанных в техническом и официально-деловом стиле. Машинный перевод художественных текстов изобилуют большими погрешностями, так как программы перевода не владеют тонкими языковыми нюансами, не понимают иносказаний и намёков, не воспринимают игру слов. Перевод художественных текстов — это всецело творческая работа, которая не под силу и программе машинного перевода.

Список литературы

Аль Кадими Махмуд Гази Чаллюб. Лингвистические и межкультурная коммуникация // Семантические и структурные признаки арабских заимствований и пути их освоения. Выпуск 4. Воронеж, 2008. С. 21–27.

«Википедия» — интернет-ресурс: URL: <https://blog.google/products/translate/ten-years-of-google-translate/>.

Гарбовский Н.К. Теория перевода: Учебник. 2-е изд. Вестник Московского университета. Изд-во МГУ, 2007.

Гарбовский Н.К. Системологическая модель о переводе. Трансдисциплинарность и система научных знаний. Вестник Московского университета. Серия 22. Теория перевода, 2015. № 1. С. 3–21.

Гарбовский Н.К., Костинова О.И. Интеллект для перевода: искусный или искусственный? Вестник Московского университета. Серия 22. Теория перевода, 2019. № 1. С. 3–23.

Гарбовский Н.К., Мишуков Э.Н. Военный перевод в современном мире (теоретико-методологические, лингвистические, военно-исторические и социально-политические аспекты). Вестник Московского университета. Серия 22. Теория перевода. № 2. М.: Изд-во МГУ, 2010. С. 16–41.

Кузнецов С.А. Большой толковый словарь русского языка. СПб.: Норинт, 2000.

Лехина И.В., Петрова Ф.Н. Словарь иностранных слов. М.: Издательство «Советская энциклопедия», 1964.

«Нейросетевой машинный перевод: революция на наших глазах». URL: <https://www.promt.ru/technology/neural-machine-translation/>.

Смирнов Л.Н. Словарь иностранных слов и выражений. М.: Издательство ЭКСМО-Пресс, 2000.

Редькин О.И. Формирование корпуса текстов и определений частотности слов в арабском языке: проблемы и решения. Вестник СПбГУ. Сер. 13, 2014. Вып. 1.

Фасмер М. Этимологический словарь русского языка в 4-х т. М., 1964.

Цзинь Ифан. Корпуса межъязыковых данных и перевод. Вестник Московского университета. Серия 22. Теория перевода. 2019. № 1. С. 4.

Castillo Gaspari, Moorkens and Way. Integrating machine translation into moocs. Proceedings of EDULEARN17 Conference, 3rd–5th July 2017, Barcelona, Spain. URL: https://www.researchgate.net/publication/318706512_INTEGRATING_MACHINE_TRANSLATION_INTO_MOOCS.

De Camp J. (2009). What is Missing User-Centric MT?. Proceeding soft the MT Summit. Ottawa, Canada.

Handschuh C. (2013). Human Translation vs. Machine Translation: A Triumph of Nuance over Nonsense? URL: <http://dcults.dcu.ie/dcu-languageservices-news/human-translation-vs-machine-translation>.

Keshavarz M.H. (1999). Contrastive Analysis and Error Analysis. URL: <http://bookyar.net/?paperno=142803>.

Lotz S., and Rensburg A. Van. (2014). Translation technology explored : Has a three-year maturation period done Google Translate any good ? Stellen bosch Papers in Linguistics Plus, 43(1), 235–259. URL: <http://doi.org/10.5842/43-0-205>.

Reiss Katharina. (1977/1989). “Text Types, Translation Types and Translation Assessment.” In Readings in Translation Theory, Translated and edited by A. Chesterman, 105–115. Helsinki: Oy Finn Lectura Ab.

Valotkaite J. and M. Asadullah. (2012). Error detection for post-editing rule-based machine translation. Proceedings of the Association for Machine Translation in the Americas, Workshop on Post-editing Technology and Practice. URL: <http://www.mt-archive.info/10/AMTA-2012-Valotkaite.pdf>.

Vilar D., Xu, J. D’Haro L., and Ney H. (2006). Error Analysis of Statistical Machine Translation Output. Proceedings of the LREC. 697–702.

URL: https://msryon.com/Section_479/ناسيات-مصر-بالخارج/الحياة-السياسية/
إغلاق-صناديق-الاقتراع-في-ر-1158122

URL: <https://www.aa.com.tr/ar/العربية-الدول/إغلاق-مصر-العربية-الدول/الرائسة-بانتخابات-الأول/1100172>

URL: <https://www.aa.com.tr/ar/العربية-الدول/الثاني-اليوم-الناسيات-لليوم-الثاني/العربية-الدول/مصر-يو-الخارج-يو-اصولون-الإدلاء/1091452>

URL: <https://www.almasryalyoum.com/news/details/1226294>

Pavel S. Mitin,

Student at the Department of Public Administration, Gzhel State University, Russia; e-mail: pro.inter2017@yandex.ru

TYPICAL ERRORS IN MACHINE TRANSLATION OF SOCIOPOLITICAL VOCABULARY: A CASE STUDY OF TRANSLATIONS OF EGYPT'S MASS MEDIA

In this article, the author made an attempt to investigate the most typical errors of “Google Translate (GT)” in the machine translation of news texts from Arabic into Russian. The integrity and semantic adequacy of the translation results were assessed when determining their quality in order to identify the degree of the need for professional correction of the completed machine translation by a qualified interpreter.

The author uses the method of content analysis, as well as quantitative and qualitative descriptive approaches to assess the quality of typical translation errors. Several excerpts from translations of e-newspapers were selected as the material for the study.

Key words: machine translation, program, neural network, identification, correction, newspaper, typology, spelling, vocabulary, semantics, morphology, homonymy, omission, editing.

References

Al Qadimi Mahmoud Gazi Chalub. Linguistic i meschkulturnaja communication. Semantic i struktura priznaki arabskich zaimstvovanija i puti ich osvvaenija, [Linguistic and intercultural communication. Semantic and structural features of Arabic borrowings and ways of their development]. Issue 4. Voronezh, 2008, pp. 23–27 (In Russian).

Wikipedia is an Internet resource: URL: <https://blog.google/products/translate/ten-years-of-google-translate/>.

Garbovskij N.K. Teorija perevoda [The Theory of Translation]. Moscow: Izd-vo Moskovskogo universiteta, 2004. 544 p. (In Russian).

Garbovskij N.K. Sistemologicheskaya model' o perevode. Transdisciplinarnost' i sistema nauchnyh znanij [Systemological model about translation. Transdisciplinarity and the system of scientific knowledge]. *Vestnik Moskovskogo universiteta. Seriya 22. Teoriya perevoda*, 2015. No. 1, pp. 3–21 (In Russian).

Garbovskij N.K., Kostikova O.I. Intellekt dlya perevoda: iskusnyj ili iskusstvennyj? [Intelligence for Translation: Skillful or Artificial?]. *Vestnik Moskovskogo universiteta. Seriya 22. Teoriya perevoda*, 2019. No. 1, pp. 3–23 (In Russian).

Garbovskij N.K., Mishkurov E.N. Voennyj perevod v sovremennom mire (teoretiko-metodologicheskie, lingvisticheskie, voenno-istoricheskie i social'no-politicheskie aspekty) [Military translation in the modern world (theoretical and methodological, linguistic, military-historical and socio-political aspects)]. *Vestnik Moskovskogo universiteta. Seriya 22. Teoriya perevoda. No. 2, 2010*, pp. 16–41 (In Russian).

Kuznetsov S.A. Bol'shoy tolkovyj slovar' russkogo yazyka [Great Dictionary of Russian language.] St. Petersburg: Norint, 2000. (In Russian).

Lekhina I.V., Petrova F.N. Slovar' inostrannyh slov [Dictionary of foreign words]. Moscow: Izdatel'stvo "Sovetskaya enciklopediya", 1964 (In Russian).

Neurosetevoi machine perevod: revolischen na naschich glasach [Neural network machine translation: a revolution before our eyes]. URL: <https://www.promt.ru/technology/neural-machine-translation>. (In Russian).

Smirnov L.N. Slovar inostrannich slov i viraschenij, [Dictionary of foreign words and expressions]. Moscow: EKSMO-Press, 2000 (In Russian).

Red'kin O.I. Formirovanie korpusa tekstov i opredelenij chastotnosti slov v arabskom yazyke: problemy i resheniya [Formation of the text corpus and definitions of the frequency of words in the Arabic language: problems and solutions]. *Vestnik SPbGU. Ser. 13, 2014. Ussue. 1* (In Russian).

Fasmer M. Etimologicheskij slovar' russkogo yazyka v 4-h t. [Etymological dictionary of the Russian language in 4 volumes]. Moscow, 1964 (In Russian).

Jin Yifang. Korpusa mesjasikovich dannih I perevod, [Interlanguage Data Corps and Translation]. *Vestnik Moskovskogo universiteta. Seriya 22. Teoriya perevoda, 2019. No. 1. P. 4* (In Russian).

Castilho Gaspari, Moorkens and Way. Integrating machine translation into moocs. Proceedings of EDULEARN17 Conference, 3rd–5th July 2017, Barcelona, Spain. URL: https://www.researchgate.net/publication/318706512_INTEGRATING_MACHINE_TRANSLATION_INTO_MOOCS.

De Camp J. (2009). What is Missing User-Centric MT? Proceeding soft the MT Summit. Ottawa, Canada.

Handschuh C. (2013). Human Translation vs. Machine Translation: A Triumph of Nuance over Nonsense? URL: <http://dcuils.dcu.ie/dcu-language-services-news/human-translation-vs-machine-translation>.

Keshavarz M.H. (1999). Contrastive Analysis and Error Analysis. URL: <http://bookyar.net/?paperno=142803>.

Lotz S., and Rensburg A. Van. (2014). Translation technology explored: Has a three-year maturation period done Google Translate any good? Stellenbosch Papers in Linguistics Plus, 43 (1), 235–259. URL: <http://doi.org/10.5842/43-0-205>.

Reiss Katharina. (1977/1989). "Text Types, Translation Types and Translation Assessment." In Readings in Translation Theory, Translated and edited by A. Chesterman, 105–115. Helsinki: Oy Finn Lectura Ab.

Valotkaite J. and M. Asadullah. (2012). Error detection for post-editing rule-based machine translation. Proceedings of the Association for Machine Translation in the Americas, Workshop on Post-editing Technology and Practice. URL: <http://www.mt-archive.info/10/AMTA-2012-Valotkaite.pdf>.

