



DOI: <http://dx.doi.org/10.15688/jvolsu2.2014.3.8>

УДК 81'322.4

ББК 81.184

МАШИННЫЕ СИСТЕМЫ ПЕРЕВОДА: КАЧЕСТВО И ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ

Новожилова Анна Алексеевна

Кандидат филологических наук,
доцент кафедры теории и практики перевода
Волгоградского государственного университета
novozilova@rambler.ru, tipp@volsu.ru
просп. Университетский, 100, 400062 г. Волгоград, Российская Федерация

Аннотация. В статье подробно рассматриваются две системы машинного перевода – «ПРОМТ» и «Google Translate», которые в настоящее время наиболее популярны среди пользователей интернета. Работа этих систем организована по разным принципам: «ПРОМТ» работает по так называемому принципу «перевода по правилам» и копирует переводческую деятельность человека; «Google Translate» осуществляет переводы по технологии, основанной на принципе статистики. На примере фрагментов текстов разных жанров предлагается лингвистический анализ качества переводов с английского языка на русский, выполненных данными системами. В ходе тестирования выявляются типичные ошибки, допускаемые в машинном переводе, и определяются возможности и сферы использования каждой из переводческих систем. Устанавливается, что данные системы непригодны для работы с текстами, содержащими большое количество сложносочиненных и сложноподчиненных предложений. Эти программы работают в основном на уровне словосочетания, и их можно успешно применять для перевода формализованных текстов, например технической документации, потребительских инструкций, формальных описаний и т. п., для которых характерно использование простых распространенных предложений и в которых не содержатся предложения со сложными синтаксическими конструкциями. Делается вывод о том, что профессиональные переводчики могут прибегать к помощи этих систем в своей практической деятельности при работе с определенными типами текстов, что помогает экономить время и оптимизировать процесс перевода. Однако обязательным условием достижения высокого качества переводов, выполненных с помощью компьютерных программ, является их постредактирование.

Ключевые слова: информационные технологии, электронные словари, автоматизация перевода, машинный перевод, качество перевода.

Научно-технический прогресс влечет за собой возрастание потоков информации, которой обмениваются носители разных языков, что способствует расширению межъязыковых границ и увеличению объема текстов, требующих перевода. «В настоящее время перевод выполняет главную функцию в процессе межкультур-

ной коммуникации, позволяя человечеству преодолевать трудности в аккумулировании полезной и необходимой информации» [9, с. 133]. Современные переводчики уже не могут использовать в своей работе лишь словари. В переводоведении все больше внимания уделяется анализу электронных средств, позволяющих

ускорить и оптимизировать процесс перевода. Отечественные и зарубежные ученые-лингвисты, практики и теоретики-переводоведы, в особенности специалисты в области переводческого терминоведения и машинного перевода, отмечая возрастающую значимость информационных технологий в лингвистике в целом и в переводе в частности, разрабатывают различные стратегии и методы их наиболее эффективного применения в профессиональной деятельности [2; 4; 6; 11; 13; 14].

Процессы глобализации и общедоступность интернета обусловили интенсивное развитие и совершенствование систем информационного обеспечения коммуникативной деятельности [5, с. 70]. По мнению В.Н. Шевчука, начало нового тысячелетия совпало с наступлением эры информационной революции и в практической деятельности профессиональных переводчиков и филологов за последние 20 лет произошли гигантские изменения [10, с. 8]. Многие ученые считают, что появление компьютеров и интернета стало серьезным шагом вперед в области практического переводоведения, позволило значительно повысить качество перевода, особенно при переводе текстов с родного языка на иностранный, а переводчикам работать стало проще и быстрее [3; 7; 10; 12; 14]. А.Н. Усачева утверждает, что благодаря интернету «переводчик приобрел уникальную возможность выхода в мировую информационную сеть, ему стали доступны данные отовсюду. Изменения, которые это внесло в профессию переводчика, настолько колоссальны, что все последствия этого сейчас вряд ли возможно оценить» [8, с. 82].

Сегодня в распоряжении переводчика находятся разнообразные электронные инструменты, ускоряющие и облегчающие переводческий процесс. Наиболее известными из них являются электронные переводные словари («Lingvo», «Мультитран», «Мультилекс», «Babylon», «Polyglossum» и др.), системы класса Translation Memory («TRADOS», «Déjà vu», «Wordfast» и др.), программы автоматического редактирования текстов, программы распознавания устной речи, электронные библиотеки, терминологические базы данных, а также сама глобальная сеть Интернет как хранилище информационных ресурсов. Все эти инструменты являются лишь вспомогательными в

деятельности профессионального переводчика, хотя ученые еще с середины прошлого столетия работают над созданием систем, которые автоматически выполняли бы безупречные переводы с одного языка на другой. Первая система машинного перевода – IBM Mark II, разработанная компанией IBM совместно с Джорджтаунским университетом, была представлена в 1954 г. в Нью-Йорке. С тех пор исследования и разработки в сфере машинного перевода не прекращаются, а специалисты в области перевода на протяжении многих лет ведут споры о целесообразности применения программ автоматического перевода. В среде профессиональных переводчиков высказывается много аргументов за их использование и против него, но как противники, так и сторонники машинного перевода остаются единодушны в том, что у этих программ есть одно неоспоримое преимущество – значительная экономия времени, затрачиваемого на перевод больших объемов текста. На сегодняшний день существуют различные системы машинного перевода, некоторые из них выполняют перевод на довольно высоком уровне. Но для того, чтобы выявить, какие именно системы подходят для работы с теми или иными языками или типами текстов, необходимо осуществлять тестирование этих систем и анализировать качество перевода на основе больших фрагментов текстов. Это входит в задачи лингвистов-переводоведов, а разработчики автоматизированных программ перевода стараются исправлять ошибки в системе, развивать и совершенствовать свои продукты, опираясь на получаемые результаты и теоретические обоснования специалистов в области переводоведения.

К основным факторам, затрудняющим машинный перевод, исследователи относят:

- языковую неоднозначность, которая может быть как лексического, так и грамматического характера;
- наличие сложных синтаксических структур, которые могут значительно различаться в языке оригинала и в языке перевода;
- различия в порядке слов в предложении (прямой / обратный, строгий / свободный);
- наличие анафорических связей в тексте;
- наличие идиом, смысл которых невозможно передать посредством пословного перевода;

– наличие неологизмов;
– существование культурных различий у языковых сообществ и т. д. [10, с. 222–223].

Среди наиболее известных автоматических переводчиков можно выделить системы машинного перевода «ПРОМТ» и «Google Translate». Автоматический переводчик «ПРОМТ» работает по принципу «перевода по правилам». Технология этого перевода состоит в применении алгоритмов, в соответствии с которыми программа анализирует текст и на основе проведенного анализа синтезирует вариант перевода. Считается, что работа такого машинного переводчика похожа на процесс мышления человека [1]. Работа системы машинного перевода «Google Translate» осуществляется по принципиально иной технологии, основанной на статистическом вычислении вероятности совпадений. Этой системой используется множество баз параллельных текстов, в которых попарно хранятся словосочетания и их переводы. В процессе перевода осуществляется статистический анализ: система подбирает эквивалент для перевода, основываясь на частоте употреблений, и в итоге подставляет вариант, имеющий наиболее высокий процент совпадений. Следует отметить, что «Google Translate» изначально разрабатывался для перевода с английского языка и на него, и до сих пор английский является языком-посредником при работе с другими парами языков. Иначе говоря, перевод осуществляется не напрямую: сначала происходит трансфер текста с языка-оригинала на английский, а уже потом – на необходимый язык перевода, что во многом влияет на качество перевода.

Сопоставим примеры переводов, выполненные системами «ПРОМТ» и «Google Translate» с английского языка на русский. При переводе фрагмента текста технического характера – инструкции по эксплуатации электронного устройства (игровой приставки) – были получены следующие результаты:

Оригинал:

Keep food and beverages away from product. Do not spill liquid of any kind on this product as it may render it inoperative. Clean with a slightly damp cloth (cold water). Do not use soap, detergent or other chemicals. Never submerge the unit in water.

«ПРОМТ» (режим – общая тематика, без подключения специализированных словарей):

Держите отдельно еду и напитки от продукта. Не проливайте жидкость никакого вида на этом продукте, поскольку это может отдалить его недействующий. Чистый с немного влажной тканью (холодная вода). Не используйте мыло, моющее средство или другие химикаты. Никогда не погружайте единицу в воду.

«Google Translate» (основной предлагаемый вариант):

Держите еду и напитки от продукта. Не допускайте попадания какой-либо жидкости об этом продукте, как это может сделать его неработоспособным. Очистите со слегка влажной тканью (холодная вода). Не используйте мыло, моющие средства или другие химические вещества. Никогда не погружайте прибор в воду.

Следует отметить, что обе системы правильно распознали формы повелительного наклонения, используемые в тексте оригинала. Переводчик «Google Translate» верно определил частеречную принадлежность всех языковых единиц. Это составляет его преимущество по сравнению с «ПРОМТ», который выбрал для перевода лексемы *clean* прилагательное *чистый*, а не глагол *очистить*. Эта проблема связана с широко распространенным в английском языке явлением грамматической омонимии. «ПРОМТ», выполняя перевод по грамматическим правилам, неверно дешифровал исходную синтаксическую конструкцию и допустил ошибку, поскольку в программе не был учтен тот факт, что английский язык относится к аналитическому типу языков с достаточно строгим порядком слов и обязательным наличием сказуемого в предложении: именно оно при переводе не было распознано. «Google Translate» определил наличие глагола в этом предложении и его грамматическую форму (императив), но не распознал его видового значения, поскольку категория вида в английском языке не является обязательной для каждой формы глагола. В данном предложении профессиональный переводчик обязательно использовал бы глагол несовершенного вида, чтобы подчеркнуть неоднократность предписываемого действия.

Приведенные переводы показывают, что обе системы не всегда хорошо справляются с выбором управления глаголов и часто работа-

ют на уровне словосочетаний, стоящих в непосредственной близости друг от друга (*попадания какой-либо жидкости об этом продукте; не проливайте жидкость никакого вида на этом продукте* и др.). При этом осуществлен и корректный выбор управления, например: *Никогда не погружайте единицу в воду / Никогда не погружайте прибор в воду; Не используйте мыло, моющее средство или другие химикаты / Не используйте мыло, моющие средства или другие химические вещества.*

Различия обнаруживаются и при переводе наречий. Так, «ПРОМТ» перевел наречие *away* с помощью наречия *отдельно*, а «Google Translate» опустил это наречие, ограничившись переводом предлога *from*. Однако применение функции просмотра и выбора доступных вариантов перевода позволяет увидеть, что в «Google Translate» единицы *away* и *from* рассматриваются как связанная пара и предлагаются более точные варианты перевода – *вдали от, далеко от*.

В системе «ПРОМТ» есть функция выбора специальной терминологической области. При переводе данного отрывка с подключением тематической области «техника: гаджеты» был получен следующий текст:

*Держите отдельно еду и напитки от продукта. Не проливайте жидкость никакого вида на этом продукте, поскольку это может **представить** его недействующий. Чистый с немного влажной тканью (холодная вода). Не используйте мыло, моющее средство или другие химикаты. Никогда не погружайте **модуль** в воду.*

Подключение специализированной тематической области обусловило замену лексики *единица* лексемой *модуль*, что в данном контексте можно рассматривать как более удачный вариант. При переводе английского глагола *render* был использован глагол *представить*, а не глагол *отдать*. Однако оба варианта нельзя считать корректными, поскольку в данном случае глагол *render* следовало бы перевести посредством словосочетания *приводить в состояние* (*это может привести его в неисправное состояние*).

Если воспользоваться в системе «Google Translate» предлагаемыми для отдельных словосочетаний вариантами перевода, то можно «собрать» следующий текст:

Храните еду и напитки вдали от продукта. Избегайте попадания каких-либо жидкостей об этом продукте, как это может сделать его неработоспособным. Очистите со слегка влажной тканью (холодная вода). Не используйте мыло, моющие средства или другие химические вещества. Никогда не погружайте прибор в воду.

Приведенный перевод свидетельствует о том, что пользователю чаще всего предлагаются варианты, которые практически не различаются ни лексически, ни грамматически. Однако если сравнить оригинал и перевод второго предложения, то видно, что система произвела грамматическую и лексическую трансформации, заменив императив глагола *spill* в отрицательной конструкции (*Do not spill liquid of any kind* – дословно: *не проливайте жидкостей какого-либо рода*) на конструкцию «глагол + существительное» (*не допускайте / избегайте попадания*), но правильного управления для существительного (*попадание куда-л.*) распознать не смогла. При переводе второй части этого предложения не было дано корректного варианта и для многозначного союза *as* («так как, поскольку»), использующегося для выражения причинно-следственных отношений.

Результаты анализа представлены в приведенной ниже таблице.

Сравнение машинных переводчиков «ПРОМТ» и «Google Translate»

Языковой параметр сопоставления	Система машинного перевода	
	«ПРОМТ»	«Google Translate»
Наклонение	+	+
Часть речи	+	+/-
Вид	+/-	+/-
Управление	+/-	+/-
Лексика	+/-	+/-

Из таблицы видно, что обе системы хорошо распознают наклонение; «ПРОМТ» допускает меньше ошибок с определением членности принадлежности лексических единиц; обе системы не всегда справляются с выбором верного управления, подбором правильного лексического эквивалента и нормами употребления глагольного вида в русском языке. Однако, несмотря на наличие ошибок и неточностей, допущенных программами, можно сделать вывод о том, что общий смысл переведенного фрагмента понятен, и после

внесения корректуры в переводы, выполненные каждой из систем, можно получить текст, который будет вполне соответствовать нормам русского языка.

Приведем в качестве еще одного примера переводы текста другого жанра – энциклопедической статьи.

Оригинал (фрагмент):

Article 5 of the North Atlantic treaty, requiring member states to come to the aid of any member state subject to an armed attack, was invoked for the first and only time after the 11 September 2001 attacks, after which troops were deployed to Afghanistan under the NATO-led ISAF. The organization has operated a range of additional roles since then, including sending trainers to Iraq, assisting in counter-piracy operations and in 2011 enforcing a no-fly zone over Libya in accordance with U.N. Security Council Resolution 1973.

«ПРОМТ»:

Статья 5 Североатлантического соглашения, требуя, чтобы государства-члены пришли на помощь любому государству-члену, подвергающемуся вооруженному нападению, была призвана в течение первого и единственного времени, после 11 сентября 2001 нападает, после которого войска были развернуты в Афганистан под ВЕДОМЫМ НАТО ISAF. Организация управляла диапазоном дополнительных ролей с тех пор, включая отправку тренеров в Ирак, помощь в операциях противопиратства и в 2011 проводя в жизнь бесполетную зону по Ливии в соответствии с Резолюцией 1973 Совета безопасности ООН.

«Google Translate»:

Статья 5 Североатлантического договора, требующего государства-члены пришли на помощь любой стране-члене условию вооруженного нападения, был вызван в первый и единственный раз после 11 сентября 2001 года, после чего войска были развернуты в Афганистане под руководством НАТО ISAF. Организация осуществляет свою деятельность ряд дополнительных ролей с тех пор, включая отправку инструкторов в Ирак, помощь в борьбе с пиратством и в 2011 году приведения бесполетной зоны над Ливией в соответствии с Резолюцией 1973 Совета Безопасности ООН.

Исходный фрагмент текста состоит из двух повествовательных предложений, первое из которых является сложноподчиненным и содержит большое количество конструкций,

построенных на основе подчинительных связей между их компонентами. Второе является простым распространенным предложением с однородными членами, герундием и сложными синтаксическими конструкциями. Переводы позволяют увидеть, что обе системы не распознают сложных синтаксических связей, неверно определяют сказуемое, объектные, атрибутивные и другие отношения в предложении. Выполненные переводы представляют собой, скорее, произвольный набор слов и словосочетаний, не объединенных смысловыми связями. Их, в отличие от предыдущего примера, невозможно отредактировать, не обращаясь к оригиналу.

Представленный анализ позволяет сделать вывод о том, что машинные переводчики «Google Translate» и «ПРОМТ» непригодны для переводов текстов, изобилующих сложносочиненными и сложноподчиненными предложениями, распространенными определениями, метафорическими сравнениями, сложными синтаксическими конструкциями и т. п. Перевод таких текстов может выполнить лишь человек, поскольку только он способен к осуществлению глубинных когнитивных речемыслительных процессов. Однако машинные системы можно использовать для перевода с английского на русский язык текстов, которые достаточно формализованы в речевом плане (техническая документация, потребительские инструкции, контракты и т. п.). Однако и в этом случае выполненные машиной переводы требуют обязательного редактирования и доработки «вручную». Профессиональные переводчики, от которых социум, как правило, ожидает переводы высокого качества, могут прибегать к помощи «ПРОМТ» и «Google Translate» при переводе текстов, ограниченных в сфере функционирования, поскольку названные компьютерные системы часто предлагают вполне адекватные, не требующие редактирования, варианты переводов для отдельных предложений или небольших фрагментов текста либо варианты, требующие минимальной корректуры, которая занимает у опытного переводчика незначительное время. Таким образом, системы «ПРОМТ» и «Google Translate» при правильном использовании способствуют ускорению процесса перевода и оптимизации труда практикующих переводчиков.

СПИСОК ЛИТЕРАТУРЫ

1. Андреев, А. Машинный перевод: правила против статистики / А. Андреев. – Электрон. текстовые дан. – Режим доступа: <http://www.computerra.ru/cio/old/offline/2007/63/329838/>. – Загл. с экрана.
2. Zubov, A. V. Информационные технологии в лингвистике / А. В. Zubov, И. И. Зубова. – М.: Академия, 2004. – 208 с.
3. Королев, Э. И. Промышленные системы машинного перевода / Э. И. Королев. – М.: Всесоюз. центр переводов, 1991. – 104 с.
4. Марчук, Ю. Н. Компьютерная лингвистика / Ю. Н. Марчук. – М.: АСТ: Восток-Запад, 2007. – 320 с.
5. Новожилова, А. А. Обучение студентов-переводчиков работе с электронными ресурсами как основа их будущей конкурентоспособности и успешности / А. А. Новожилова, Е. А. Шовгенина // Вестник Волгоградского государственного университета. Серия 6, Университетское образование. – 2013. – № 14. – С. 70–76.
6. Соловьева, А. В. Профессиональный перевод с помощью компьютера / А. В. Соловьева. – СПб.: Питер, 2008. – 160 с.
7. Тиссен, Ю. В. Интернет в работе переводчика / Ю. В. Тиссен // Мир перевода. – 2000. – № 2. – С. 45–62.
8. Усачева, А. Н. Инновационные технологии в профессиональном переводе / А. Н. Усачева // Коммуникативные аспекты современной лингвистики и лингводидактики: материалы Междунар. науч. конф., г. Волгоград, 29 янв. 2008 г. – Волгоград: Волгогр. науч. изд-во, 2008. – С. 81–87.
9. Усачева, А. Н. Перевод: от лингвистической теории к когнитивной модели / А. Н. Усачева // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2011. – № 1 (13). – С. 131–137.
10. Шевчук, В. Н. Информационные технологии в переводе. Электронные ресурсы переводчика – 2 / В. Н. Шевчук. – М.: Зебра Е, 2013. – 384 с.
11. Шевчук, В. Н. Электронные ресурсы переводчика / В. Н. Шевчук. – М.: Либрайт, 2010. – 136 с.
12. Austermühl, F. Electronic Tools for Translators / F. Austermühl. – Manchester: St. Jerome, 2001. – 192 p.
13. Bowker, L. Computer-Aided Translation Technology: A Practical Introduction / L. Bowker. – Ottawa: Univ. of Ottawa Press, 2002. – 185 p.
14. Computers and Translation: A Translator's Guide / H. Somers (ed.). – Amsterdam; Philadelphia: John Benjamins Publ. Company, 2003. – 349 p.

REFERENCES

1. Andreev A. *Mashinnyy perevod: pravila protiv statistiki* [Machine Translation: Rules against Statistics]. Available at: <http://www.computerra.ru/cio/old/offline/2007/63/329838/>.
2. Zubov A.V., Zubova I.I. *Informatsionnye tekhnologii v lingvistike* [Information Technologies in Linguistics]. Moscow, Akademiya Publ., 2004. 208 p.
3. Korolev E.I. *Promyshlennyye sistemy mashinnogo perevoda* [Industrial Machine Translation Systems]. Moscow, Vsesoyuznyy tsentr perevodov Publ., 1991. 104 p.
4. Marchuk Yu.N. *Kompyuternaya lingvistika* [Computer Linguistics]. Moscow, AST, Vostok-Zapad Publ., 2007. 320 p.
5. Novozhilova A.A., Shovgenina E.A. *Obuchenie studentov-perevodchikov rabote s elektronnyimi resursami kak osnova ikh budushchey konkurentosposobnosti i uspeshnosti* [Teaching Prospective Translators to Use Electronic Resources as Base of Their Future Competitiveness and Success]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 6, Universitetskoe obrazovanie* [Science Journal of Volgograd State University. University Education], 2013, no. 14, pp. 70-76.
6. Solovyeva A.V. *Professionalnyy perevod s pomoshchyu kompyutera* [Professional Translation by Means of Computer]. Saint Petersburg, Piter Publ., 2008. 160 p.
7. Tissen Yu.V. *Internet v rabote perevodchika* [Internet for Translators and Interpreters]. *Mir perevoda*, 2000, no. 2, pp. 45-62.
8. Usacheva A.N. *Innovatsionnye tekhnologii v professionalnom perevode* [Innovative Technologies in Professional Translation]. *Kommunikativnye aspekty sovremennoy lingvistiki i lingvodidaktiki. Materialy Mezhdunarodnoy nauchnoy konferentsii. Volgograd, 29 yanvarya 2008* [Communicative Aspects of Modern Linguistics and Linguodidactics: Proceedings of International Scientific Conference, Volgograd, January 29, 2008]. Volgograd, Volgogradskoe nauchnoe izd-vo, 2008, pp. 81-87.
9. Usacheva A.N. *Perevod: ot lingvisticheskoy teorii k kognitivnoy modeli* [Translation: From Linguistic Theory Towards a Cognitive Model]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2, Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2011, no. 1 (13), pp. 131-137.
10. Shevchuk V.N. *Informatsionnye tekhnologii v perevode. Elektronnyye resursy perevodchika – 2* [Information Technology in Translation. Electronic

Translator's Resources – 2]. Moscow, Zebra E Publ., 2013. 384 p.

11. Shevchuk V.N. *Elektronnye resursy perevodchika* [Electronic Translator's Resources]. Moscow, Librayt Publ., 2010. 136 p.

12. Austermühl F. *Electronic Tools for Translators*. Manchester, St. Jerome, 2001. 192 p.

13. Bowker L. *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa, University of Ottawa Press, 2002. 185 p.

14. Somers H., ed. *Computers and Translation: A Translator's Guide*. Amsterdam, Philadelphia, John Benjamins Publishing Company, 2003. 349 p.

MACHINE TRANSLATION SYSTEMS: QUALITY AND POSSIBLE WAYS OF USE

Novozhilova Anna Alekseevna

Candidate of Philological Sciences,
Associate Professor, Department of Translation Theory and Practice,
Volgograd State University
novozhilova@rambler.ru, tipp@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Abstract. The article considers two machine translation systems, PROMT and Google Translate, which work on different principles. PROMT works on the so-called “translation by the rules” principle. Google Translate uses the technology which is based on the statistics principle. The article presents the linguistic analysis of the quality of the translations of text fragments of various genres made from English into Russian by these systems. The most typical mistakes made by the systems are revealed, the possibilities and spheres of use are determined. The author reveals that the systems are not useful for translating texts which are full of compound and complex sentences. The programs work mainly on the phrase level and can be fruitfully used for translating texts of formal style, for example, technical documents, manuals, etc., which contain simple extended sentences and do not contain sentences with complex syntactic constructions. The author considers that professional translators can use these systems while working with certain text types; however, the translations made by the systems must be checked thoroughly and verified.

Key words: information technologies, electronic dictionaries, computerized translation, machine translation, quality of translation.