

О.А. Никишова, В.Ю. Потапова

ИСПОЛЬЗОВАНИЕ СИСТЕМ МАШИННОГО ПЕРЕВОДА ПРИ ПЕРЕВОДЕ ТЕКСТОВ НАУЧНО-ТЕХНИЧЕСКОГО СОДЕРЖАНИЯ

Аннотация. Почти семьдесят лет прошло со времён знаменитого Джорджтаунского эксперимента, однако вопрос создания качественных систем машинного перевода, позволяющих быстро и качественно выполнить даже не машинный, а автоматизированный перевод, до сих пор стоит очень остро и занимает умы лучших отечественных и зарубежных программистов и филологов. Несмотря на существенный прогресс в данной области, письменный перевод, выполненный с использованием систем машинного перевода, по-прежнему требует со стороны переводчика большой работы по пред- и постредактированию. В статье мы хотим представить наш взгляд на данную проблему с точки зрения специалистов, которым ежедневно приходится выполнять значительные по объёму письменные переводы научно-технической (прежде всего – медицинской) направленности, выделить моменты, которые наиболее часто требуют постредактирования.

Ключевые слова: машинный перевод, автоматизированный перевод, технология машинного перевода, система машинного перевода, предредактирование, постредактирование.

О.А. Nikishova, V.Yu. Potapova

USE OF MACHINE TRANSLATION SYSTEMS IN THE TRANSLATION OF TEXTS OF SCIENTIFIC AND TECHNICAL CONTENT

Abstract. Almost seventy years have passed since the famous Georgetown-IBM experiment, however, the issue of creating high-quality machine translation systems that allow you to perform quickly and efficiently not even machine, but automated translation is still very critical and occupies the minds of the best domestic and foreign programmers and philologists. Despite significant progress in this area, a written translation made using machine translation systems still requires a lot of pre- and post-editing work on the part of the translator. In our article, we want to present our view of this problem from the point of view of specialists who daily have to perform significant written translations of scientific and technical (primarily medical) orientation, to highlight the points that most often require post-editing.

Keywords: machine translation, automated translation, machine translation technology, machine translation system, pre-editing, post-editing.

После распада СССР, когда бывшие советские граждане получили возможность массово выезжать за рубеж по туристическим, студенческим и приглашительным визам, свободно принимать участие в международных симпозиумах, выставках, конференциях, работать на совместных предприятиях и т. д., на рынке труда резко

возрос спрос на профессиональных переводчиков. Требовались специалисты, владеющие навыками как устного, так и письменного перевода. Перевод научно-технических и деловых материалов и сейчас пользуется большим спросом. Если в первой половине XX века при подготовке переводчиков акцент прежде всего делался на переводе

Никишова Ольга Александровна

старший преподаватель кафедры иностранных языков и профессионального перевода Медицинского института, Тамбовский государственный университет имени Г.Р. Державина, город Тамбов. Сфера научных интересов: лингвистика, перевод в сфере профессиональной коммуникации, теория и практика перевода, когнитивная лингвистика. Автор девяти опубликованных научных работ. SPIN-код: 8421-0751, AuthorID: 1150681.

Электронный адрес: Nicolette1@mail.ru

Потапова Валентина Юрьевна

старший преподаватель кафедры иностранных языков и профессионального перевода Медицинского института, Тамбовский государственный университет имени Г.Р. Державина, город Тамбов. Сфера научных интересов: лингвистика, перевод в сфере профессиональной коммуникации, когнитивная лингвистика. Автор более 10 опубликованных научных работ. SPIN-код: 5909-7385, AuthorID: 1153084.

Электронный адрес: zotvalen@mail.ru

лучших литературных произведений зарубежных писателей и писателей советских республик, то после распада СССР было организовано множество фирм, специализировавшихся на переводе личных и нотариальных документов, научно-технической документации.

Сегодня спрос на перевод различных документов, учебных пособий, технических инструкций, программ мероприятий, рекламных проспектов, да и литературных произведений продолжает расти. Современному письменному переводчику приходится работать с огромным количеством материала. Как и в других сферах деятельности, здесь человеку пришли на помощь машинные технологии.

В конце прошлого века В.Н. Комиссаров писал: «На протяжении многовековой истории перевода неоднократно делались попытки теоретически осмыслить и объяснить деятельность переводчиков, сформулировать критерии оценки качества переводов, определить факторы, оказывающие влияние на ход и результат процесса перевода. Однако, как это часто бывает в других видах

человеческой деятельности, переводческая практика значительно опередила теорию перевода» [1, с. 253]. Е.В. Бреус [2] совсем недавно в своих лекциях вновь указал, что, несмотря на многочисленные попытки теоретического осмысления, по состоянию на настоящее время многие исследователи признают, что в машинном переводе каких-либо прорывов в реализации иных моделей, отличных от моделей, основанных на применении лингвистической теории структурализма и последующем разделении переводческой деятельности на два компонента: собственно перевод, осуществляемый по заданным правилам, и интерпретацию, включающую интерпретацию влилингвистических данных, не произошло. Для решения данной проблемы нужны глубокие теоретические и экспериментальные исследования. Конечно, теоретически осмыслить то, что каждый день развивается, дополняется, совершенствуется, очень сложно. В нашей статье мы предлагаем рассмотреть те приёмы, которыми сегодня пользуются современные письменные переводчики, работающие с узкоспециализированными

Использование систем машинного перевода при переводе текстов научно-технического содержания

текстами и использующие в своей работе онлайн-переводчики.

На сегодняшний момент в развитии машинных технологий перевода можно выделить четыре разных направления:

1) статистическое направление – Statistical Machine Translation (SMT);

2) машинный перевод на основе правил – Rule-based Machine Translation (RBMT);

3) гибридное направление – Hybrid Machine Translation (HMT);

4) нейронный машинный перевод – Neural Machine Translation (NMT).

SMT-технология при переводе с одного языка на другой использует статистические модели, которые являются частью естественного языкового процесса. Для построения моделей проводится анализ большого количества двуязычного материала (параллельные тексты, словарные соответствия). Целью данного анализа является выявление закономерностей и связей между словами и фразами на разных языках. Именно эти закономерности и связи используются для построения вероятностных моделей, которые в дальнейшем будут использоваться при переводе новых предложений или документов. Одним из ключевых преимуществ *SMT-технологии* специалисты считают возможность обрабатывать широкий спектр языковых пар и решать практические задачи перевода: от простых фразовых переводов до более сложных нейронных машинных переводов. Данная технология используется в широком спектре приложений, включая лингвистические аспекты локализации программного обеспечения и веб-сайтов, перевод контента для многоязычных веб-сайтов и трансграничную коммуникацию для предприятий и организаций.

RBMT-технология основана на лингвистических правилах, которые позволяют использовать слова в различных значениях в зависимости от контекста. Технология

оперирует большим количеством лингвистических правил на трех разных этапах: анализ, передача и генерация. Эти правила разрабатываются экспертами по языкам и программистами, в задачу которых входит сопоставление правил между двумя языками. Также *RBMT* опирается на словари, созданные вручную, некоторые из которых пользователи могут редактировать и уточнять для улучшения перевода. Однако в некоторых случаях внедрение и поддержка этих усовершенствований могут занять большое количество времени, а иногда и привести к снижению качества перевода из-за двусмысленности терминов. Переводы, выполненные с помощью данной технологии, вполне понятны для понимания, но очень часто выглядят более «машинными». То есть на практике данная технология подходит для передачи смысла сообщения, но требует значительной последующей доработки, чтобы точно передать цель и характер письменного высказывания. Учитывая недостатки технологии, разработчики *RBMT* попытались исправить их путём применения некоторых методов *SMT*. Эти новые продукты были представлены как *Hybrid MT*. На сегодняшний день существует несколько моделей гибридного машинного перевода, каждая из которых имеет свои преимущества и недостатки.

Технологии Hybrid Machine Translation основаны на многочисленных подходах к машинному переводу в рамках одной системы. Именно невозможность достичь удовлетворительного качества перевода при использовании какого-либо отдельного метода послужила толчком для разработки *HMT*. Например, Р.Г. Мифтахова и Е.А. Морозкина [3] относят к гибридным системам перевода системы, где используется сочетание «статического метода с нейронными сетями», где две технологии «взаимно компенсируют существующие в них недо-

статки». Таким образом, есть несколько форм данной технологии, наиболее известными из которых являются Rules Post-Processed by Statistics и Statistics Guided by Rules.

Технологии *Neural Machine Translation* начали разрабатываться в 1990-е гг. и сегодня являются основным видом машинного перевода. Данная технология разрабатывалась с целью улучшить качество машинного перевода и основана на использовании механизмов сложной работы нейронных сетей головного мозга. Однако, по мнению некоторых исследователей, например, Р.Г. Мифтаховой и Е.А. Морозкиной, «несмотря на все преимущества новой технологии перевода, на сегодняшний день она не может полностью заменить уже существующие системы» [3, с. 497].

Говоря о машинном переводе, мы используем данный термин в узком смысле, то есть как процесс перевода текста с одного естественного языка на другой с использованием технологий и систем машинного перевода, реализуемый компьютером с возможным участием человека – переводчика, владеющего иностранным языком.

Специалистам, которые работают в области письменного перевода, конечно, в первую очередь важна не сама технология, а результат её использования. Чем качественнее конечный продукт, тем меньше времени переводчик тратит на постредактирование, тем самым увеличивая производительность своего труда.

Сегодня в сети Интернет можно найти много сайтов, где проводится детальный анализ процесса и результата работы различных онлайн-систем машинного перевода [4]. На разных сайтах последовательность представления систем может быть различной, однако список их наименований, за редким исключением, почти везде один и тот же. В рейтинг лучших онлайн-перевод-

чиков на основе искусственного интеллекта сегодня входят системы:

- Bing Microsoft Translator;
- Systran;
- Reverso;
- Amazon Translate;
- DeepL;
- Яндекс. Переводчик;
- Google Translate.

Необходимо отметить, что специалисты различают машинный и автоматизированный перевод. Перевод, выполненный только с использованием компьютерной программы, является машинным. Перевод, требующий последующей доработки человеком, – автоматизированный, или CAT-перевод [5]. Поэтому в письменном переводе, выполненном с участием машинных систем, необходимо различать непосредственно машинную и автоматизированную составляющие.

На качество машинного перевода напрямую влияют содержание, стиль переводимого текста. Некоторые специалисты считают, что машинные переводы художественных текстов уступают по качеству переводам научно-технических текстов из-за наличия в первых аргоизмов, поэтизмов, экзотизмов, игры слов и т. д. По их мнению, текст научно-технического содержания может быть более понятен машине, и его перевод требует лишь небольшого постредактирования при условии наличия достаточно объёмной терминологической базы. О.А. Крапивкина и А.А. Мусохранова отмечают, что «язык научной литературы может быть довольно формальным...» [6]. Но как часто и до какой степени?

Возьмём для машинного перевода фразу: «Белки в клетке». Эту короткую фразу можно воспринимать двояко: «Бёлки в клетке» (например, в зоопарке) и «Белки в клетке» (например, организма). Слова «клетка» являются полными омонимами, совпадающими по звучанию и написанию

Использование систем машинного перевода при переводе текстов научно-технического содержания

грамматических форм, а слова «белки» и «белки́» – это омографы. Посмотрим, как эту фразу переведут различные онлайн-системы машинного перевода сначала с рус-

ского на английский (Таблица 1), а затем с русского на французский (Таблица 2). Также добавим фразы с уточняющими словами.

Таблица 1

Перевод фразы различными онлайн-системами машинного перевода с русского языка на английский

Система перевода Фраза	Google Translate	Яндекс Переводчик	DeepL Translate	Reverso	Bing Microsoft Translator
Белки в клетке	Proteins in a cell	Proteins in the cell	Proteins in the cage	Proteins in the cell	Proteins in the cell
Белки в металлической клетке	Squirrels in a metal cage	Proteins in a metal cage	Squirrels in a metal cage	Proteins in a metal cell	Proteins in a metal cell
Белки в клетке организма	Proteins in the cell of the body	Proteins in the body cell	Proteins in the body cell	Proteins in the body cell	Proteins in the cell of the body
Белки́ в клетке	Proteins in a cell	Proteins in the cell	Squirrels in a cage	Proteins in the cell	Whites in a cage
Белки́ в клетке	Proteins in a cell	Proteins in the cell	Proteins in the cell	Proteins in the cell	Proteins in the cell
Лесные белки в клетке	Forest squirrels in a cage	Forest squirrels in a cage	Forest squirrels in a cage	Forest proteins in the cell	Forest squirrels in a cage

Таблица 2

Перевод фразы различными онлайн-системами машинного перевода с русского языка на французский

Система перевода Фраза	Google Translate	Яндекс Переводчик	DeepL Translate	Reverso	Bing Microsoft Translator
Белки в клетке	Protéines dans une cellule	Protéines dans la cellule	Protéines dans la cage	Protéines dans une cellule	Protéines dans la cellule
Белки в металлической клетке	Écureuils dans une cage métallique	Protéines dans une cage métallique	Écureuils dans une cage métallique	Protéines dans une cellule métallique	Protéines dans une cellule métallique
Белки в клетке организма	Protéines dans les cellules du corps	Protéines dans la cellule du corps	Protéines dans les cellules du corps	Protéines dans la cellule du corps	Protéines dans la cellule du corps
Белки́ в клетке	Protéines dans une cellule	Protéines dans la cellule	Écureuils dans une cage	Wapiti en cage	Blancs en cage
Белки́ в клетке	Protéines dans une cellule	Protéines dans la cellule	Protéines dans une cellule	Protéines dans une cellule	Protéines dans la cellule
Лесные белки в клетке	Écureuils forestiers en cage	Écureuils forestiers en cage	Écureuils de forêt en cage	Protéines forestières dans la cellule	Écureuils forestiers en cage

При переводе исходной фразы разница в переводах и с английского, и с французского языка прежде всего заключается в использовании определённого или неопределённого артикля – a/the, une/la. Однако окончательный вариант артикля может осуществляться только переводчиком, поскольку, как отмечает Е.А. Реферовская, «первой, наиболее нормальной и закреплённой функцией определённого артикля являлось обозначение связи конкретного предмета с конкретной ситуацией. Связь эта выясняется либо из предыдущего контекста, либо уточняется тут же, либо просто вытекает из самой ситуации...» [7, с. 27]. По мнению Е.А. Реферовской, артикль выражает субъективное отношение говорящего к предмету своего высказывания.

Кроме того, система DeepL Translate и в английском, и во французском вариантах вместо cell/cellule предлагает слова cage/cage. Однако данная система всегда предлагает пользователю несколько вариантов перевода, и в этих вариантах перевода мы видим и такие, как “proteins in a cage”, “proteins in the cell”, “proteins in a cell”.

Со второй фразой успешно справились лишь Google Translate и DeepL Translate. Только эти две системы «поняли», что в металлических клетках могут находиться именно звери, а не органические соединения.

С третьей фразой справились все системы. Но при этом при переводе с русского на французский Google Translate и DeepL Translate слово «клетка» переводят во множественном числе.

Четвёртую фразу правильно восприняла только система DeepL Translate. Система Bing Microsoft Translator в данном случае «поняла» слово «белки» в значении «что-то, имеющее белый цвет». Система Reverso при переводе на французский язык перевела слово «белки» – “wapiti” (лось).

В пятом случае, как и в первом, все системы верно интерпретировали фразу, и разница в переводах заключается только в употреблении определённого/неопределённого артикля.

При переводе последнего варианта явную ошибку выдала только система Reverso, переведя выражение «лесные белки» как “forest proteins” и “protéines forestières”. Также отметим, что при переводе на французский язык DeepL Translate, в отличие от других систем, перевела прилагательное «лесной» не как “forestier” (имя прилагательное), а как «de forêt» (предлог “de” + имя существительное).

Отметим, что при переводе с русского языка на французский выражений, отражающих посессивные отношения, всегда наблюдается большая разница между вариантами, предлагаемыми различными системами. Возьмем, к примеру, словосочетание «зубная коронка». Системы машинного перевода дают нам следующие варианты: “couronne de la dent” (Google Translate), “couronne” (Яндекс Переводчик), “couronne dentaire” (Bing Microsoft Translator), “couronne des dents” (Reverso), “couronne dentaire” (DeepL Translate). Конечно, во французском языке категория посессивности может выражаться как конструкцией «имя существительное + имя прилагательное», так и «имя существительное + предлог de + имя существительное». Наш опыт работы позволил нам сделать вывод, что только кропотливый анализ большого количества аутентичного материала помогает определиться с выбором наиболее подходящего слова или выражения. Поэтому и в данном случае окончательный выбор остаётся за переводчиком.

Занимаясь на протяжении нескольких лет письменными переводами текстов преимущественно медицинской тематики с русского языка на английский и французский,

Использование систем машинного перевода при переводе текстов
научно-технического содержания

и используя онлайн-системы машинного перевода, мы считаем, что основные недостатки переводов, выполненных с их использованием, сводятся к следующим:

- несоответствие порядка слов в предложении, свойственного одному языку, другому;
- подмена специальных терминов общепотребительной лексикой или терминами из другой предметной области;
- искажение смысла понятий, отражающих реалии организации отечественного здравоохранения и других сфер деятельности;
- ошибки в согласованиях (например, прилагательные согласуются не с теми именами существительными).

Поэтому, прежде чем приступить непосредственно к переводу, при письменном переводе с использованием машинных систем необходимо провести большую подготовительную работу:

- 1) вдумчиво прочитать текст для перевода (если текст сложный, то необходимо прочесть его несколько раз);
- 2) составить глоссарий для каждого переводимого текста;

3) поделить текст на небольшие фрагменты (определить единицу переводческого процесса), но не без отрыва от контекста;

4) провести предварительную обработку переводимого отрывка (очень много ошибок допускается системами при переводе сложносочиненных и сложноподчиненных предложений, с причастными и деепричастными оборотами. Поэтому иногда целесообразно разбить длинное предложение на несколько более коротких).

Таким образом, несмотря на существенный прогресс и продолжающуюся работу, существующие онлайн-системы при переводе текстов научной направленности можно использовать лишь как инструмент, позволяющий переводчику сэкономить большое количество времени на наборе текста. Кроме того, их применение снижает возможность опечатки, которая может возникнуть при ручном наборе. Использование данных систем для выполнения перевода высокого качества должно сопровождаться пред- и постредактированием.

Литература

1. *Комиссаров В.Н.* Теория перевода (лингвистические аспекты): учеб. для институтов и фак. иностр. яз. М. : Высшая школа, 1990. 253 с. ISBN 5-06-001057-0.
2. *Бреус Е.В.* Теория и практика машинного перевода : курс лекций / Ун-т рос. акад. образования. 2-е изд. М. : Изд-во УРАО, 2003. 103 с. ISBN 5-204-00402-5.
3. *Мифтахова Р.Г., Морозкина Е.А.* Машинный перевод. Нейроперевод // Вестник Башкирского университета. 2019. Т. 24. № 2. С. 497–502. EDN VWVCPH.
4. 7 лучших онлайн-переводчиков на основе ИИ // YADOVERYAUI.NET. 2023. 22 июля. URL: <https://yadoveryaui.net/7-lychshih-onlain-perevodchikov-na-osnove-ii.html> (дата обращения: 10.09.2023).
5. *Березовская Е.А., Ильнер А.О.* Редактирование письменных переводов: теория и практика. Екатеринбург : Изд-во Уральского университета, 2019. 135 с. ISBN 978-5-7996-2622-8. EDN UVCWYP.
6. *Крапивкина О.А., Мусохранова А.А.* Лингво-прагматические аспекты перевода научно-технических текстов // Гуманитарные научные исследования. 2014. № 8. С. 34–37. EDN SMWUXJ. URL: <https://human.snauka.ru/2014/08/7629?ysclid=lolpdczut2553900247> (дата обращения: 10.09.2023).
7. *Реферовская Е.А.* Формирование романских литературных языков. Л. : Наука, 1980.

References

1. Komissarov V.N. (1990) *Teoriya perevoda (lingvisticheskiye aspekty)* [Translation theory (linguistic aspects)] : Textbook. Moscow : Vysshaya shkola Publ. 253 p. ISBN 5-06-001057-0. (In Russian).
2. Breus E.V. (2003) *Teoriya i praktika mashinnogo perevoda* [Theory and practice of machine translation] : Course of lectures. Univ. of the Russian academy of education. 2nd edition. Moscow : Univ. of the Russian academy of education Publ. 103 p. ISBN 5-204-00402-5. (In Russian).
3. Miftakhova R.G., Morozkina E.A. (2019) Machine translation. Neural translation. *Vestnik Baschkirskogo universiteta* [Bulletin of Bashkir University]. Vol. 24. No. 2. Pp. 497–502. (In Russian).
4. 7 best AI-powered online translators. *YADOVERYAUI.NET*. 2023. July 22. URL : <https://yadoveryaui.net/7-lychshih-onlain-perevodchikov-na-osnove-ii.html> (accessed 10.09.2023). (In Russian).
5. Berezovskaya E.A., Il'ner A.O. (2019) *Redaktirovaniye pis'mennykh perevodov: teoriya i praktika* [Editing of Written Translations: Theory and Practice]. Ekaterinburg : Ural University Publ. 135 p. ISBN 978-5-7996-2622-8. (In Russian).
6. Krapivkina O.A., Musokhranova A.A. (2014) Linguistic and pragmatic aspects of scientific and technical translation. *Humanities scientific research*. No. 8. Pp. 34–37. URL: <https://human.snauka.ru/2014/08/7629?ysclid=lolpdczut2553900247> (accessed 10.09.2023). (In Russian).
7. Referovskaya E.A. (1980) *Formirovaniye romanskikh literaturnikh yazikov* [Formation of Romance literary languages]. Leningrad : Nauka Publ. 200 p. (In Russian).