

УДК 81.25

ОСОБЕННОСТИ МАШИННОГО ПЕРЕВОДА СТАТЕЙ ИЗ ПЕЧАТНЫХ СМИ

Бикбулатова К.А.

студент,

*Калужский государственный университет им. К.Э. Циолковского,
Калуга, Россия*

Горгарова Я.Ю.

старший преподаватель кафедры французского языка,

*Калужский государственный университет им. К.Э. Циолковского,
Калуга, Россия*

Аннотация

В статье обосновывается понятие «автоматического (машинного) перевода», рассматриваются исторические предпосылки его возникновения и анализируются преимущества и недостатки работы существующих систем машинного перевода (основанных на правилах, статистических, нейронных, гибридных). В ходе исследования была изучена специфика машинного перевода статей из печатных СМИ, осуществленного при помощи системы Яндекс.Переводчик. Переводческий анализ статьи «*Une association saisit la justice après le nouveau décès d'une orque au Marineland d'Antibes*» из газеты *Le Figaro* позволяет сделать выводы о качестве машинного перевода, а также выявить ключевые особенности и проблемы, возникающие при переводе публицистических текстов с французского языка на русский, связанные, в первую очередь, с переводом многозначных слов, специализированной терминологии и имен собственных.

Ключевые слова: машинный перевод, машинный перевод на основе правил, статический машинный перевод, нейронный машинный перевод, гибридный машинный перевод, тексты СМИ.

***PARTICULARITIES OF MACHINE TRANSLATION OF ARTICLES FROM
PRINTED MEDIA***

Bikbulatova K.A.

student,

Kaluga State University named after K.E. Tsiolkovsky,

Kaluga, Russia

Gorgarova Y.Y.

Senior Lecturer, Department of French Language,

Kaluga State University named after K.E. Tsiolkovsky,

Kaluga, Russia

Abstract

The article substantiates the concept of «automatic (machine) translation», examines its historical background and analyses the advantages and disadvantages of existing machine translation systems (rule-based, statistical, neural, hybrid). In our research we studied the specifics of machine translation of articles from the print media using the Yandex.Translator system. The translation analysis of the article «*Une association saisit la justice après le nouveau décès d'une orque au Marineland d'Antibes*» from the newspaper *Le Figaro* allows us to draw conclusions about the quality of machine translation, as well as to identify the key features and problems arising when translating publicistic texts from French into Russian, mainly related to the translation of polysemous words, specialised terminology and proper names.

Keywords: machine translation, rule-based machine translation, statistical machine translation, neural machine translation, hybrid machine translation, media texts.

В области перевода и переводоведения происходят постоянные изменения. В первую очередь, они связаны с применением в переводческой практике систем автоматизированного, или машинного, перевода, которые способны осуществлять перевод слов, предложений и даже целых текстов. Однако это не позволяет полностью избежать ошибок и неточностей при осуществлении перевода. В этой связи актуальным становится вопрос об интеграции функциональных возможностей программ машинного перевода с работой реального переводчика в целях обеспечения более высокого качества переводимых текстов.

В Большой российской энциклопедии автоматический перевод (машинный перевод) определяется как «перевод текстов с одного естественного языка на другой с помощью компьютера» [1].

Термином «машинный перевод» (МП) обозначают, с одной стороны, перевод текстов (письменных или устных) с одних естественных языков на другие при помощи ЭВМ, с другой стороны, направление научных исследований, связанных с построением переводческих систем [3].

Мысль о возможности использовать ЭВМ для перевода была высказана еще в 1946 году, сразу после появления первых вычислительных машин. Толчком к активному повсеместному увлечению этой проблематикой послужил известный Джорджтаунский эксперимент (1954 г.). Это был первый в мире публичный эксперимент по переводу, выполненный в Джорджтаунском университете США [4].

В ходе эксперимента компьютеру предлагалось перевести несколько предложений с русского языка на английский. По тем меркам эксперимент прошел успешно, в большей степени благодаря тому, что его условия были

довольно примитивны (был произведен перевод 49 тщательно подобранных простых предложений на базе грамматики из шести правил с использованием словаря на 150 единиц) [6].

Выбор языка, с которого осуществлялся перевод, был обусловлен начинающейся «холодной войной» и необходимостью в связи с этим анализировать огромные объемы информации, поступающие на русском языке. Для перевода была выбрана достаточно узкая тематика нефтехимии, и все вводимые предложения были очень простыми: «Обработка повышает качество нефти», «Командир получает сведения по телеграфу». При таких условиях эксперимент оказался вполне успешным, более того, появилось мнение, что задача машинного перевода полностью решится в течение ближайших пяти лет. Однако дальнейшие годы исследований показали, что проблема машинного перевода гораздо сложнее и многограннее, чем ожидалось.

Процесс машинного перевода подразделяется на виды в зависимости от подхода к построению алгоритмов работы. Классифицировать машинный перевод можно по следующим параметрам.

По степени участия человека в переводе выделяют:

- FАМТ (Fully-automated Machine Translation) – полностью автоматический машинный перевод;
- НАМТ (Human-assisted Machine Translation) – машинный перевод с участием человека (человек при этом выступает в качестве редактора исходного текста и / или текста перевода);
- МАНТ (Machine-assisted Human Translation) – перевод, выполненный человеком при помощи компьютера (с использованием электронных словарей, автоматических редакторов текста, системы памяти переводов и т. д.) [6].

По особенностям архитектуры системы выделяют: перевод, основанный на правилах, статистический и гибридный перевод [5].

Перевод на основе правил (Rule-based Machine Translation (RBMT)) опирается на бесчисленные встроенные лингвистические правила и множество двуязычных словарей для каждой языковой пары. Одними из наиболее известных разработчиков систем машинного перевода, основанных на правилах, являются PROMT.One (Россия) и SYSTRAN Translate (Франция).

Статистический машинный перевод (Statistical Machine Translation (SMT)) является одним из видов машинного перевода, принцип действия которого заключается в непрерывном сравнении имеющихся баз данных в требуемых языковых парах. Качество полученного перевода во многом будет зависеть от того, как много языковых пар поддерживает система машинного перевода, основанная на статистическом подходе, а также, насколько точно базы данных необходимых языковых пар согласуются друг с другом. Статистический машинный перевод обладает навыком «самообучения». Алгоритм данного типа используется такими популярными сервисами онлайн-перевода как Яндекс.Переводчик (Россия) (до 2017 г.), Google Translate (США) (до 2016 г.), Microsoft Translator (США) (до 2016 г.), SYSTRAN (Франция) (до 2016 г.).

Нейронный машинный перевод использует большую искусственную нейронную сеть. Ее основной чертой является способность самообучаться в пределах одного языка, и применять полученные знания при переводе. Такой подход задействует лишь крошечную часть всей той памяти, которая необходима статистическому машинному переводу, поэтому данная система является намного более ресурсоемкой [2].

Гибридный машинный перевод – это метод машинного перевода, который характеризуется использованием нескольких подходов к машинному переводу в рамках единой системы. Мотивация для разработки систем гибридного машинного перевода проистекает из неспособности какого-либо отдельного метода достичь удовлетворительного уровня точности при переводе. Многие гибридные системы машинного перевода добились успеха в повышении

точности переводов, и существует несколько популярных систем машинного перевода, в которых используются гибридные методы, например, SYSTRAN's hybrid engine и PROMT DeepHybrid.

Для анализа качества машинного перевода статей из печатных СМИ в рамках нашего исследования была выбрана статья «*Une association saisit la justice après le nouveau décès d'une orque au Marineland d'Antibes*», опубликованная на официальном сайте газеты *Le Figaro* 15 апреля 2024 года [8]. Статья рассказывает о том, что организация, специализирующаяся на защите прав животных, подала в суд после того, как одна из косаток умерла в морском тематическом парке «Маринленд», который находится в городе Антиб, во Франции. Организация обвиняет парк в плохих условиях содержания животных и их использовании лишь ради получения финансовой выгоды.

Приведем некоторые примеры переводческих ошибок, допущенных системой машинного перевода Яндекс.Переводчик [7].

Таблица 1 – Пример 1 перевода текста статьи с французского языка на русский, выполненный автоматизированной системой машинного перевода Яндекс.Переводчик

Оригинал	Яндекс.Переводчик
<u>L'association C'est Assez!</u> a engagé une procédure d'urgence auprès du tribunal administratif de Nice.	<u>Ассоциации достаточно!</u> возбудил экстренное разбирательство в административном суде Ниццы.

В данном предложении следует отметить несколько ошибок, осуществленных системой машинного перевода. Во-первых, это грамматическая несогласованность форм «*l'association a engagé*» и «*ассоциации возбудил*». Во-вторых, Яндекс.Переводчик не смог определить, что «*C'est Assez!*» – это название организации, защищающей права животных, название которой можно было бы сохранить в тексте в качестве прямого включения. Система правильно интерпретировала употребление существительного с предлогом «*d'urgence*» и

перевела его при помощи прилагательного «экстренный». Однако с учетом правил лексической сочетаемости русского языка считается возможным перевести выражение «*engager une procédure d'urgence*» как «*подать срочный иск*». Таким образом, мы можем предложить следующий перевод первой части предложения: «*Ассоциация «C'est Assez!» подала срочный иск...*» (Таблица 1).

Таблица 2 – Пример 2 перевода текста статьи с французского языка на русский, выполненный автоматизированной системой машинного перевода Яндекс.Переводчик

Оригинал	Яндекс.Переводчик
L'association, qui a saisi <u>le juge des référés</u> vendredi dernier, s'inquiète du sort des autres mammifères vivant <u>sur ce site</u> .	Ассоциация, которая обратилась <u>к судье по апелляциям</u> в прошлую пятницу, обеспокоена судьбой других млекопитающих, обитающих <u>на этом участке</u> .

В представленном предложении можно обратить внимание на тот факт, что система машинного перевода не в полной мере способна работать со специальной юридической терминологией. Так, для выражения «*le juge des référés*» предлагается перевод «*судья по апелляциям*». Однако апелляция подразумевает пересмотр решения судьи вышестоящим судом. Таким образом, наиболее удачным эквивалентом будет выражение «*рассмотреть дело в срочном порядке*». Также лексически неправильно было переведено словосочетание «*sur ce site*» («*на этом участке*»), где можно предложить вариант с использованием приемов добавления и конкретизации: «*на территории парка*» (Таблица 2).

Таблица 3 – Пример 3 перевода текста статьи с французского языка на русский, выполненный автоматизированной системой машинного перевода Яндекс.Переводчик

<i>Оригинал</i>	<i>Яндекс.Переводчик</i>
<p>«Nous saisissons le juge pour demander à la préfecture de mettre en demeure le parc <u>Marineland</u> dans un délai de 8 jours à compter de la décision à intervenir, et de prendre toutes les mesures utiles afin de protéger les deux dernières orques encore en vie <u>au sein de leur établissement</u>», indique <u>Christine Grandjean, la responsable de l'association.</u></p>	<p>«Мы обращаемся к судье с просьбой, чтобы префектура подала в суд на парк <u>Marineland</u> в течение 8 дней с момента принятия решения о вмешательстве и приняла все необходимые меры для защиты двух последних косаток, оставшихся в живых <u>в их поселении</u>», – говорит <u>Кристин Гранджан, директор Департамента охраны окружающей среды. ответственный за ассоциацию.</u></p>

Сравнивая оригинал текста и машинный перевод, можно сделать вывод о том, что переводчик избежал перевода названия парка «*Marineland*», используя прямое включение и, таким образом, снова нарушив единообразие интерпретации имен собственных в тексте статьи. В случае с переводом названия тематического парка и города, где он расположен, Яндекс.Переводчик по неизвестной причине предлагает несколько разных вариантов интерпретации этих имен собственных в заголовке и тексте статьи: «*Маринленд в Антибе*» и «*Маринланд Д'Антиб*». В данной ситуации следует обратиться к устойчивым эквивалентам, которые уже существуют в русском языке и соответствуют тому варианту, который был использован в заголовке статьи («*Маринленд в Антибе*»).

Лексически неправильно было переведено словосочетание «*au sein de leur établissement*» – «*в их поселении*». В конкретном случае данное выражение может быть опущено ввиду того, что оно не несет дополнительной смысловой нагрузки. При переводе имени и фамилии правозащитницы мы бы рекомендовали использование транскрипции («*Кристин Гранджан*»), а при передаче на русский язык ее должности – подбор соответствующего эквивалента, например, «*представитель ассоциации*». В последнем случае не совсем понятно

использование Яндекс.Переводчиком формулировки «*директор Департамента охраны окружающей среды*», которая лишь запутывает читателя. В данном предложении необходимо также обратить внимание на несоблюдение системой машинного перевода правил пунктуации русского языка (Таблица 3).

Таблица 4 – Пример 4 перевода текста статьи с французского языка на русский, выполненный автоматизированной системой машинного перевода Яндекс.Переводчик

Оригинал	Яндекс.Переводчик
Cette dernière demande un transfert des mammifères dans l'un des bassins des dauphins, en meilleur état selon elle.	Последняя просит перевести млекопитающих в один из бассейнов с дельфинами, по ее словам, в лучшем состоянии.

При переводе предложения на русский язык был нарушен порядок слов, поэтому мы рекомендуем следующий вариант перевода с применением приема добавления: «*Она просит перевести млекопитающих в один из бассейнов с дельфинами, который, по ее мнению, находится в лучшем состоянии*» (Таблица 4).

Анализ статей из печатных СМИ позволил выделить основные особенности и проблемы, с которыми сталкиваются системы машинного перевода, в частности связанные с переводом многозначных лексических единиц, специализированной лексики из отдельных областей знания, имен собственных, а также вызванные несовпадением синтаксиса и пунктуации двух языков.

Таким образом, мы можем сделать вывод о том, что дальнейшее развитие технологий машинного перевода требует интеграции современных методов и подходов к осуществлению переводческой деятельности, а также учета специфики переводимых текстов из различных источников, включая статьи из печатных СМИ. Важно продолжать исследования в данной области с целью совершенствования качества и точности машинного перевода и сделать его еще

более доступным и эффективным инструментом для обмена информацией и коммуникации между людьми разных культур и языковых общностей.

Библиографический список:

1. Богуславский, И.М. Автоматический перевод // Большая российская энциклопедия. Электронная версия (2016) [Электронный ресурс]. – Режим доступа – URL: <https://old.bigenc.ru/linguistics/text/1799594> (дата обращения: 09.09.2024).

2. Головкин, Д.Р. Особенности и виды машинного перевода / Д.Р. Головкин // Вестник Московского информационно-технологического университета – Московского архитектурно-строительного института. – 2020. – № 4. – С. 24-30.

3. Кочеткова, Н.С. Особенности машинного перевода / Н.С. Кочеткова, Е.В. Ревина // Филологические науки. Вопросы теории и практики. – 2017. – № 6-2 (72). – С. 106-109.

4. Мироненко, А.Б. Проблемы машинного перевода / А.Б. Мироненко, П.А. Омельченко // Актуальные проблемы авиации и космонавтики. – 2016. – № 12. – С. 1193-1195.

5. Молчанов, А. Статистические и гибридные методы перевода в технологиях компании PROMT / А. Молчанов // Control Engineering Россия. – 2013. – № 4 (46). – С. 68-72 [Электронный ресурс]. – Режим доступа – URL: <https://controlengrussia.com/innovatsii/statisticheskie-i-gibridny-e-metody-perevoda-v-tehnologiyah-kompanii-promt/> (дата обращения: 20.09.2024).

6. Щекина, Л.Э. Особенности онлайн-перевода с французского языка на русский (на примере работы Яндекс. Переводчик). – 2020 [Электронный ресурс]. – Режим доступа – URL: <https://www.rgph.vsu.ru/ru/science/sss/reports/2/schekina.pdf> (дата обращения: 11.09.2024).

7. Яндекс.Переводчик: официальный сайт. – URL:
https://translate.yandex.ru/?utm_source=main_stripe_big

8. Une association saisit la justice après le nouveau décès d'une orque au Marineland d'Antibes // Le Figaro: официальный сайт. – 2024 [Электронный ресурс]. – Режим доступа – URL: <https://www.lefigaro.fr/actualite-france/une-association-saisit-la-justice-apres-le-nouveau-deces-d-une-orque-au-marineland-d-antibes-20240415> (дата обращения: 15.11.2024).

Оригинальность 82%