

УДК 81'322.4

МАШИННЫЙ ПЕРЕВОД: ИСТОРИЯ, КЛАССИФИКАЦИЯ, ПРИЧИНЫ РАЗВИТИЯ

Д. Р. Быченко¹⁾, О. С. Полетаева²⁾

¹⁾ Белорусская государственная академия авиации,
ул. Уборевича, 77, 220096, г. Минск, Беларусь, d.bychanok@outlook.com

²⁾ Белорусская государственная академия авиации,
ул. Уборевича, 104, 220096, г. Минск, Беларусь, volhapaletayeva@gmail.com

Данная статья посвящена актуальному в наши дни вопросу машинного перевода, который применяется как в обычной жизни, так и профессиональной деятельности. В статье рассматривается классификация машинного перевода, его преимущества и недостатки, причины развития, история становления. Выявлены сложности для машинного перевода и представлены правила для его оптимизации.

Ключевые слова: машинный перевод; компьютер; онлайн-переводчик; человек; интернет; кибернетика.

MACHINE TRANSLATION: HISTORY, CLASSIFICATION, CAUSES OF DEVELOPMENT

D. R. Bychanok²⁾, V. S. Paletayeva¹⁾

¹⁾ Belarusian State Aviation Academy,
77 Uborevich St., 220096, Minsk, Belarus, d.bychanok@outlook.com

²⁾ Belarusian State Aviation Academy,
104 Uborevich St., 220096, Minsk, Belarus, volhapaletayeva@gmail.com

This article is devoted to the current issue of machine translation, which is used both in everyday life and professional activities. The article discusses the classification of machine translation, its advantages and disadvantages, the reasons for its development, and the history of its formation. The difficulties for machine translation are revealed and the rules for its optimization are presented.

Keywords: machine translation; computer; online translator; human; Internet; cybernetics.

С давних времен люди задумывались о том, как создать единый язык, который смог бы разрушить языковой барьер и позволил бы легко общаться с каждым человеком на Земле. Для этого необходимо было создать технологию, легко понимающую иностранные языки.

Первые идеи зародились в XVII веке, когда Рене Декарт и Готфрид Лейбниц независимо друг от друга предлагали создать универсальный

язык как новую основу для логического мышления, а также для устранения взаимного непонимания, которое происходит из-за нелогичности языков.

В 1668 году британский священник Джон Уилкинс выпустил трактат «Опыт о подлинной символической и философском языке», где представил свой подход к созданию универсального языка ученых и философов, который мог бы прийти на смену латыни. Однако его предложение не встретило одобрения среди лингвистов [1].

После этого ученые вернулись к разработке единого международного языка только в XIX–XX веке, так появилось на свет Эсперанто, однако первые попытки создания машинного перевода (МП) были предприняты только через несколько десятков лет.

Так, в 30-е годы XX века французский ученый Жорж Арцруни создал двуязычный автоматический словарь с помощью перфокарты, а советский инженер Петр Троянский «машину для подбора и печатания слов при переводе с одного языка на другой». Его идеи оставались неизвестными до 1950-х годов.

Концепция машинного перевода была сформулирована в конце 1940-х годов Уорреном Уивером, который в письме коллеге Норберту Винеру предложил рассмотреть вопрос перевода с одного языка на другой в качестве новой области применения технологий декодирования. После этого в 1949 году Уивер доказал возможность реализации машинного перевода в меморандуме, в котором описал концепцию *interlingua*:

- исходное предложение переводится через язык-посредник (английский)
- результат преобразуется на нужный язык.

7 января 1954 года был успешно проведен Джорджтаунский эксперимент. Он заключался в переводе 60 предложений с русского на английский язык с помощью компьютера IBM 701. Этот эксперимент дал мощный толчок в развитии, однако проект развития данного направления оказался очень сложным и дорогостоящим. Так, в 1964 году группа американских ученых провела оценку технологий машинного перевода и приняла решение о прекращении финансирования проекта, через 2 года он полностью был остановлен и только через 20 возобновлен.

Но, несмотря на это, в 1968 году в Сан-Диего Питер Том открыл первую в мире компанию, разрабатывающую программы для машинного перевода – Systran. Специалисты Systran использовали Rule-based Machine Translation, первые переводчики компании оказались востребованы на международном рынке.

После Джорджтаунского эксперимента Советский Союз начал активно заниматься вопросом машинного перевода. Академия Наук СССР в 1955 году создала 2 исследовательские группы: первую при Математическом Институте имени В. А. Стеклова во главе с Алексеем Ляпуновым, вторую – при Институте точной механики и вычислительной техники во главе с Дмитрием Пановым. Свои первые результаты две группы опубликовали спустя год исследований на компьютере БЭСМ. Еще через год вторая группа представила первый советский компьютерный переводчик из 23000 слов словаря. В это же время Ольга Кулагина и Игорь Мельчук вместе с Ляпуновым работали над созданием еще одного переводчика.

В конце 1956 года заведующий кафедрой перевода Московского государственного педагогического института иностранных языков Виктор Розенцвейг основал Объединение по машинному переводу – неформальный центр общения лингвистов и математиков. В 1957 году ученые, входившие в это Объединение, начали выпускать бюллетень «Машинный перевод и прикладная лингвистика» [1].

В мае 1958 года в Москве Розенцвейгом была организована первая советская конференция по машинному переводу. В ней приняли участие 340 специалистов 79 организаций. Итогом конференции стало издание приказа «О развитии научных исследований в области машинного перевода».

В 1972 году специалисты института «Информэлектрон» под руководством Юрия Апресяна начали разрабатывать систему машинного перевода – ЭТАП (Электротехнический автоматический перевод), которая основывалась на модели языка «Смысл – Текст».

В 1971 году в Ленинграде была создана научно-исследовательская лаборатория инженерной лингвистики, инициатором создания стал профессор Раймонд Пиотровский. Специалисты лаборатории занимались разработкой технологий для автоматической обработки текста. В 1991 году под руководством Светланы Соколовской бывшие сотрудники лаборатории основали компанию PROMT (PROject Machine Translation), которая в будущем вышла на международный уровень. В 2000 году появилась компания «Яндекс», которая успешно разработала свою технологию машинного перевода и также успешно вышла на международный уровень.

С появлением интернета и различной техники машинный перевод стал неотъемлемой частью жизни людей XXI века. Сегодня он активно используется в повседневной жизни: дома, на учебе, работе, в путешествиях. Он помогает справляться с рядом задач, но в первую

очередь – преодолеть языковой барьер, что в разы упростило жизнь человечества.

Машинный перевод (англ. Machine Translation) – процесс письменного перевода текстов с одного естественного языка на другой с помощью специализированного программного обеспечения.

Существуют следующие типы МП: на основе правил; статистический; гибридный (правила + статистика); нейронный.

МП на основе правил (Rule-based machine translation, RBMT) основывается на применении различных языковых правил и двуязычных словарей для каждой языковой пары.

Статистический МП (Statistical machine translation, SMT) использует статистические переводческие модели, основывающиеся на анализе одно- и двуязычных корпусов текстов. Построение таких моделей – достаточно быстрый процесс, но технология сильно зависит от существующего корпуса текстов (уже переведенных кем-то).

Гибридная модель может обрабатывать тексты разных тематик и направлений перевода, но требует значительных инвестиций и серьезной настройки.

При нейронном МП используется искусственная нейронная сеть, которая является ресурсоемкой. Она может самообучаться в пределах одного языка, а также работать по аналогии с машинным переводом на основе примеров для более точного результата.

Также МП можно разделить на виды: письменный и устный. Письменный дается машине намного легче, так как устный требует распознавания и преобразования устной речи в письменную.

На сегодняшний день машинный перевод в той или иной степени требует вмешательства человека. Существует несколько форм взаимодействия человека с электронной машиной.

- С предредактированием. Человек редактирует текст перед помещением его в буфер.

- С интерредактированием. В процессе перевода может видоизменять текст.

- С постредактированием. Редактирует готовый текст на выходе.

- Смешанная форма взаимодействия. Означает, что человек вмешивается в процесс перевода на разных этапах [2].

Также выделяют 2 типа перевода текстов по степени автоматизации:

- частично автоматизированный (подразумевает использование человеком электронных словарей);

- система перевода с разделением труда, т. е. ЭВМ переводит только определенные структуры текста, с остальным текстом работает человек.

Машинный перевод имеет как преимущества, так и недостатки.

К преимуществам машинного перевода относят:

- высокую скорость;
- низкую стоимость;
- универсальность;
- доступность;
- конфиденциальность;
- перевод в режиме онлайн.

Несмотря на все преимущества, машинный перевод имеет большой недостаток – низкое качество. Сервисы, отвечающие за перевод, в большинстве случаев переводят тексты дословно, не учитывая контекст, что сильно искажает суть содержимого. Кроме того, в словарной базе могут отсутствовать некоторые слова. ЭВМ тяжело переводить громоздкие предложения с различными оборотами. Часто не учитываются правила грамматики, порядок слов в предложениях, логико-смысловые акценты.

К основным сложностям при переводе относят:

- сокращения слов;
- прямую речь;
- слова через дефис;
- инициалы;
- идиомы и фразеологизмы;
- омонимы;
- опечатки;
- сленг.

Но даже при наличии довольно существенного недостатка, МП набирает обороты. Человеческий перевод не поспевает за скоростью генерации контента. Мы живем в эпоху взрыва цифрового контента, характеризующийся следующими параметрами:

- 2,4 из 7,1 млрд. людей на планете пользуются Интернетом;
- каждую минуту в мире создается 571 новый веб-сайт;
- каждую минуту пользователи Facebook делятся со своими подписчиками 684 478 единицами контента;
- каждый день производится 2,5 квинтиллионов байтов новых данных;
- Объем продаж онлайн превысит 3,5 трлн. долл. США в течение ближайших пяти лет. Это 9,3 % от всех розничных продаж в мире. Данная цифра через год увеличится до 15,4 %.

Еврокомиссия запустила и финансирует проект Moses Core, направленный на поощрение разработку и использование машинного перевода с открытым исходным кодом.

Подводя итоги, можно сказать, что машинный перевод активно используется, несмотря на свою неидеальность. Особенно востребованным в настоящий момент является перевод технической, финансовой и юридической документации. Поскольку эти тексты характеризуются большим процентом содержания штампов, клишированных выражений, повторяющихся фраз, их перевод не представляется возможным без помощи САТ-инструментов (Computer–Aided (Assisted) Translation). С развитием технологий машинный перевод будет модернизироваться и совершенствоваться и непременно достигнет уровня, не требующего вмешательства человека. А на сегодняшний день, чтобы добиться корректного машинного перевода, стоит придерживаться следующих правил:

- максимально упрощать предложения, заменяя сложные слова синонимами. Сложные предложения, предложения с причастными оборотами, инфинитивными группами следует разбить на несколько отдельных, простых предложений;
- стараться не использовать слова, которые являются многозначными (в противном случае велика вероятность разночтений);
- проверить текст, который требуется перевести, на наличие ошибок и опечаток [3].

Библиографические ссылки

1. История машинного перевода [Электронный ресурс]. URL: <https://www.promt.ru> (дата обращения: 17.02.2024).
2. Машинный перевод: современные технологии и топ-10 полезных инструментов [Электронный ресурс]. URL: <https://4brain.ru> (дата обращения: 13.02.2024).
3. Преимущества и недостатки машинного перевода на примере популярных онлайн переводчиков [Электронный ресурс]. URL: <https://scipress.ru> (дата обращения: 10.02.2024).