**Т. А. Плавинская**
*Тверской государственный университет, магистрант*
*Научный руководитель: к.ф.н. С. А. Колосов*

# ИСТОРИЯ И РАЗВИТИЕ МАШИННОГО ПЕРЕВОДА В РОССИИ И ЗА РУБЕЖОМ

Возможность автоматизации одной из самых творческих деятельностей человека давно вызывала интерес, но, несмотря на долгий и сложный путь развития, машинный перевод до сих пор представляет сложности, как для изучения, так и для процесса его оптимизации. Термин машинный перевод (МП) можно рассматривать, по крайней мере, в двух смыслах. В узком смысле под машинным переводом понимается процесс перевода некоторого текста с одного естественного языка на другой, реализуемый компьютером полностью или частично. Машинный перевод в широком смысле – область научных исследований, находящаяся на стыке лингвистики, математики, кибернетики, и имеющая целью построение систем, реализующих машинный перевод в узком смысле [Воронович 2013: 3].

## Предыстория

Первые попытки смоделировать действия человека-переводчика были предприняты П. П. Смирновым-Троянским. В 1933 году в СССР было запатентовано механизированное переводное устройство, представленное как «машина для подбора и печатания слов при переводе с одного языка на другой». Чтобы такое устройство функционировало, были нужны два помощника, один из которых владеет ИЯ и задает машине базовую форму каждого слова и её грамматические характеристики, второй владеет ПЯ и, после того, как машина перевела ранее заданные слова на язык перевода, собственноручно придает переведенному тексту литературную форму. Переводное устройство в академических кругах встретили скептически из-за реализации П. П. Смирновым-Троянским слишком примитивного представления о сущности переводческого процесса в переводном устройстве и на долгое время забыли [Марчук 1983: 17].

## Начало пути

С окончанием Второй Мировой войны и появлением первых ЭВМ начинается бурное развитие машинного перевода в США. В середине XX века У. Уивер предложил рассматривать задачу перевода как область применений технологий дешифровки, которой так активно пользовались при распознавании закодированных сообщений фашистской Германии. Задачу автоматизированного перевода У. Уивер видел в том, чтобы свести

язык к некоему общему коду, который будет содержать в себе инвариантное содержание ИЯ и ПЯ. Так возникает очень важная для будущего развития МП концепция interlingva (концепция языка-посредника) [Митренина 2017: 6]. Впоследствии станет очевидно, что разработка таких систем и моделей МП, которые будут переводить со всех языков на все, очень сложный и трудоемкий процесс, а для облегчения выполнения задач машинного перевода английский язык будет выбран в качестве языка-посредника.

RBMT (Rule-based Machine Translation)

Со временем у ученых и разработчиков появляется общее представление о сущности машинного перевода, который обязательно должен включать в себя машинный словарь, алгоритмы анализа и синтеза, программное обеспечение [Марчук 1983: 19]. Так вектор исследований в сфере МП меняется в сторону машинного перевода на основе правил. Данный подход был успешно реализован в результате Джорджтаунского эксперимента. В 1954 году американцы продемонстрировали миру первую действующую систему машинного перевода с русского языка на английский, которая перевела 49 заранее отобранных предложений с русского языка на английский (программа использовала словарь из 250 слов и грамматику, состоящую из шести синтаксических правил). 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. После публикации российским реферативным журналом ВИНИТИ «Математика» о первом успешном испытании системы МП начинается развитие отечественного машинного перевода [Митренина 2017: 7]. Первое поколение систем МП, разрабатываемых в СССР и за рубежом (до 1960-х гг.), базировалось на алгоритмах последовательного перевода. Возможности таких систем определялись доступными размерами словарей, прямо зависящими от объема памяти компьютера. Текст переводился отдельными предложениями, и смысловые связи между ними не учитывались. Такие системы называют системами прямого перевода. Основной задачей разработчики систем МП по всему миру видели в расширении и пополнении машинного словаря и развитии алгоритмов применения грамматических правил. С середины 1960-1990 гг. RBMT-системы подвергались значительным изменениям. Так, в СССР развивались два подхода к созданию систем МП: трансферные и системы-интерлингвы. Трансферные системы работали по следующим принципам: проводился морфологический, лексический и семантико-синтаксический анализ предложения на языке оригинала, создавалось синтактико-семантическое дерево разбора входного предложения, затем производился так называемый «трансфер», т. е. преобразование структуры входного

предложения в соответствии с формальными требованиями языка перевода. На заключительном этапе синтеза формировалось конечное предложение на языке перевода (основанная на правилах система перевода PROMT является классическим примером трансферных систем). На этом этапе в технике МП уже широко применялись как методы морфологического, так и синтаксического анализа, что существенно улучшило качество выходных текстов, однако оставались трудности, связанные с семантикой. В связи с этим следующим этапом развития МП в СССР можно считать 1980-е годы, когда впервые появляются системы семантического типа (системы-интерлигвы). К этому классу относятся системы МП, в основу которых легла теория «Смысл ↔ Текст», разработанная советскими учеными И. А. Мельчуком, Ю. Д. Апресяном и А. К. Жолковским. В основу данных систем легла теория о том, что любое предложение любого языка можно преобразовать в его смысловое представление на так называемом универсальном метаязыке, а из полученного смыслового представления можно синтезировать предложение на языке перевода. Иными словами, с помощью определенного набора правил и словаря с семантическими характеристиками можно преобразовывать текст в смысл и наоборот [Филинов URL].

Переход к SMT (Statistical Machine Translation)

В США в 1980-х г.г. была предложена система перевода, работающая по принципу аналогий. Такая система опиралась на большой набор примеров предложений и их переводов. Так, в США RBMT-системы уступают место системам, основанным на примерах (example-based systems) [Garg, Agarwal 2018: 2]. Большую роль в развитии систем МП сыграло создание высокопроизводительных компьютеров, новых языков программирования, а также известные работы Н. Хомского и ряда других ученых по разработке формальных грамматик для синтаксического анализа, что в дальнейшем привело к формированию новых методов моделирования машинного перевода, основанных на лингвистических корпусах. Объем параллельных корпусов рос вместе с исследованиями в области обработки естественного языка, и в 90-е годы начались разработки статистических систем перевода [Мифтахова 2015: 188]. Данные системы принципиально отличалась тем, что они вообще ничего не знали про правила и про лингвистику. На входе система получала огромное количество заранее переведенных идентичных текстов и анализировала, какие фрагменты предложения часто встречаются вместе в оригинале и в переводе. Однако статистическая модель перевода не давала гарантии, что для слова будет подобрано именно то значение, которое необходимо в условии контекста.

NMT (Neural Machine Translation)

Нейронный машинный перевод – это относительно новый подход к решению проблемы машинного перевода. Функционирование данной системы основано на использовании нейронных сетей, вычислительных моделей, по своей структуре напоминающих строение человеческого мозга, в которых сигнал распространяется через последовательные слои элементов, имитирующих нейроны [Калинин 2017: 71]. Векторное представление слов (word embedding), применяемое в нейронном машинном переводе, как правило, сопоставляет каждому слову вектор длиной в несколько сотен чисел. Векторы, в отличие от простых идентификаторов из статистического подхода, формируются при обучении нейронной сети и учитывают взаимосвязи между словами. Например, модель может распознать, что, поскольку «чай» и «кофе» часто появляются в сходных контекстах, оба эти слова должны быть возможны в контексте нового слова «разлив», с которым, допустим, в обучающих данных встретилось лишь одно из них. Нейронная сеть работает с большими параллельными корпусами, анализирует их с целью обнаружения закономерностей, на которых потом учится, что является главным преимуществом данных систем. Благодаря нейронным машинным сетям существенно улучшилось качество МП. Нейронный машинный перевод не просто ищет и сопоставляет слова и выражения двуязычных корпусов, с его помощью становится возможным глубже проникнуть в связи, существующие между словами, и путем сложного анализа каждого переводимого образца изучить их взаимоотношения для выяснения контекста [Мифтахова, Морозкина 2019:].

Первый такой переводчик был запущен компанией Google в 2016 году. Но возникает вопрос, что делать с теми редкими входными словами, которые недостаточно часто встречались, чтобы сеть могла построить для них приемлемое векторное представление. В этой ситуации логично совместить оба метода. На данный момент такого гибридного подхода придерживается компания Yandex, которая перешла на использование нейронных сетей в 2017 году. Как только пользователь вводит текст для перевода, Яндекс.Переводчик передает этот текст сразу двум системам: нейронной и статистической. Результат, выдаваемы обеими системами, оценивается алгоритмом, основанным на методе обучения CatBoost. При оценке учитываются десятки факторов – от длины предложения (короткие фразы и редкие слова лучше переводит статистическая модель) до синтаксиса. Алгоритм оценивает оба перевода, выбирает лучший и выдает этот перевод пользователю.

Заключение

Развитие МП тесно связано с историческими и политическими процессами, которые протекали в мире, и с развитием целого ряда наук:

математики, кибернетики, информатики, аналитики и лингвистики. Развитие МП в России и за рубежом протекало достаточно синхронно. Начало исследований в сфере машинного перевода в СССР было обусловлено Холодной войной с США и первыми успешными испытаниями RBMT-системы в ходе Джорджтаунского эксперимента. В основу первых отечественных и зарубежных систем машинного перевода легли словари и правила, которые и определяли качество перевода. Профессиональные лингвисты годами работали над тем, чтобы вывести всё более подробные правила вручную. Работа эта была столь трудоемкой, что серьезное внимание уделялось лишь наиболее популярным парам языков, но даже в рамках них машины справлялись плохо. Выяснилось, что традиционная лингвистика не располагает ни фактическим материалом, ни идеями и представленьями, нужными для построения систем МП, которые использовали бы смысл переводимого текста. Таким образом, люди осознали недостаточность исходных представлений о языке для его полного формального описания, необходимого для создания алгоритмов и систем МП, а перемещение центра тяжести от лексики к синтаксису не гарантировало качественного результата МП. Естественный язык – это очень сложная система, которая плохо поддается формализации, и единственный способ для машины постоянно адаптироваться к изменяющимся условиям – это учиться самостоятельно на большом количестве параллельных текстов.

ЛИТЕРАТУРА

Андреева А.Д., Меньшиков И.Л., Мокрушин А.А. Обзор систем машинного перевода // Молодой учёный. 2013. №12 (59). С. 64-66

Воронович В.В. Машинный перевод: конспект лекций для студентов 5-го курса специальности «Современные иностранные языки»: учеб. пособие. Минск: 2013. 39 с.

Дроздова К.А. Машинный перевод: история, классификация, методы // Филологические науки в России и за рубежом: Материалы 3-й междунар. науч. конф. 20-23 июля 2015 г. СПб., 2015. С. 139-141.

Калинин С.М. Обзор современных подходов к улучшению точности нейронного машинного перевода // Рема. 2017. №2. С. 70-79.

Карцева Е.Ю., Маргарян Т.Д., Гурова Г.Г. Развитие машинного перевода и его место в профессиональной межкультурной коммуникации // Вестник Российского университета дружбы народов. 2016. №3. С. 155-163.

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

Марчук Ю.Н. Проблемы машинного перевода: учеб. пособие. М.: Изд-во «Наука», 1983. 232 с.

Митренина О.В. Назад в 47-й: к 70-летию машинного перевода как научного направления // Вестник новосибирского государственного университета. 2017. № 3. С. 5-12.

Мифтахова Р.Г., Морозкина Е.А. Машинный перевод. Нейроперевод. // Вестник Башкирского университета. 2019. №2. С. 497-502.

Мифтахова Р.Г. Основные факторы улучшения машинного перевода // Вестник Башкирского университета. 2015. С. 188-192.

Новожилова А.А. Машинный системы перевода: качество и возможности использования // Вестник Волгоградского государственного университета. 2014. №3 (22). С. 67-73.

Ревзин И.И., Розенцвейг В.Ю. Основы общего и машинного перевода: учеб. пособие. М.: Изд-во «Высшая школа», 1964. 243 с.

Филинов Е.Н. Машинный перевод. URL: https://computer-museum.ru/histsoft/histmt.htm (дата обращения: 04.05.2021).