

Секция «Иностранные языки и регионоведение»

Машинный перевод: история и современность

Кенжаев Азиз Джамшедович

Студент

*Южный федеральный университет, Институт филологии, журналистики и
межкультурной коммуникации, Ростов-на-Дону, Россия*

E-mail: sensational.aziz@gmail.com

Машинный перевод – это перевод текста с одного естественного языка на другой естественный язык с помощью компьютерной программы. В прошлом столетии, когда вышел в свет первый машинный перевод, и когда тема машинного перевода вызвала большой ажиотаж, было разработано несколько видов машинного перевода: непосредственный перевод; косвенный перевод; интерлингва (Interlingua); трансфер (Transfer).

Датой рождения машинного перевода можно считать 1947 год, год, когда специалист по криптографии Уоррен Уивер в своем письме дешифровщику Норберту Винеру впервые поставил задачу машинного перевода, сравнив ее с задачей дешифровки. (William John Hutchins, 1986, стр. 21-29) Тот же Уивер в 1949 г. составил Меморандум о теоретической возможности. В 1952 году логик и математик Й. Бар-Хиллел создал первую конференцию по машинному переводу.

Самый первый результат плодотворного труда специалистов был продемонстрирован в 1954 году. Компания IBM совместно с Джорджтаунским Университетом провели первый эксперимент: GAT (Джорджтаунский автоматизированный переводчик), система, включающая в себя словарь всего из 250 слов и грамматику из 6 синтаксических правил, осуществила перевод 49 заранее отобранных предложений. (Sergei Nirenburg and Yorick Wilks, 1986г. стр. 3)

Система GAT была предназначена для перевода текстов советских физиков и ядерщиков с русского языка на английский. В этом же году эксперимент по машинному переводу был осуществлен и в СССР И.К. Бельской (лингвистическая часть) и Д.Ю. Пановым (программная часть) в Институте точной механики и вычислительной техники Академии наук СССР. Система переводила текст с английского языка на русский. В том же году началось ее программирование для тогдашних ЭВМ ЕС (единая система) 1040, ЕС 1065. (Ю. Н. Марчук, 2000, стр. 177)

В 1960 году был выпущен доклад Специального комитета по прикладной лингвистике (ALPAC) Национальной академии наук США под названием «Черная книга машинного перевода», в котором говорилось, что в ближайшем будущем создание высококачественной системы машинного перевода невозможно. В 1961 г. благодаря Государственному спонсированию был создан проект METAL для перевода текстов с немецкого языка.

В 1965 году в Монреальском университете был создан проект TAUM (Автоматизированный перевод), спонсируемый Канадским правительством. Этот проект был единственным, основанным на способе обработки типа трансфер.

Петер Тома, один из специалистов, работавших над системой GAT, разработал систему SYSTRAN, которая в 1970-м году заменила систему Mark II, произведенную компанией IBM, в Департаменте иностранных технологий Военно-воздушных сил США.

Корпорация SYSTRAN была более успешной в машинном переводе и в 1974 году НАСА выбрала SYSTRAN для перевода материалов, связанных с совместной Советско-Американской программой Аполло-Союз. (Makoto Nagao, 1989г., стр. 108-110)

В том же году была разработана 80,000 строчная, 14-слойная программа FORTRAN, работающая на ЭВМ типа CDC 6600. FORTRAN был переписан так, что процесс составлял всего три важных этапа: анализ, трансфер и синтез. Вся программа была составлена из 4,000 строк кодов, которые были внесены в три слоя, что позволило уменьшить загруженность ЦП в 10 раз.

В 1976 году Комиссия Европейского сообщества приобретает англо-французскую версию SYSTRAN для оценки способностей системы и ее потенциальном использовании. (Makoto Nagao, 1989г., стр. 110-112) В 1996 году SYSTRAN предложила перевод небольших текстов онлайн, совершенно бесплатно. Затем вышла Alta Vista Babelfish, которая получала до 500,000 запросов в день и впоследствии была выкуплена корпорацией Yahoo.

Франц-Джозеф Ох (будущий глава Департамента развития перевода Google), одержал победу в гонке машинного перевода в 2003 году и создал онлайн-переводчик Google Translate. Затем, в 2007 г. вышел MOSES, свободный движок статистического перевода.

Гонка за качество перевода изменила представление о способе перевода. Система машинного перевода строится на основе лингвистического описания двух естественных языков, формальных грамматик и алгоритмов перевода. Качество зависит от объема базы данных и глубины описания лингвистических языков.

Технология статистического машинного перевода (СМП) основана на поиске наиболее вероятного перевода предложения с использованием данных, полученных из двуязычной совокупности текстов. Такие системы перевода строятся на основе сравнения больших объемов корпусов параллельных текстов. Корпус параллельных текстов – это тексты, содержащие предложения на одном языке и соответствующие им предложения на втором. СМП имеет особенность самообучения. Разработчики машинного перевода на основе СМП: PROMT, Google, Microsoft.

Каждая технология перевода имеет свои недостатки и преимущества. Например, машинный перевод, основанный на правилах, дает синтаксическую точность, но требует постоянного пополнения словаря. Статистический машинный перевод дает гладкость перевода, но не умеет справляться с морфологией и синтаксисом.

В связи с тем, что данные системы достигли своего предела развития, разработчики рассчитывают на прорыв за счет создания некоего гибрида. У компании SYSTRAN уже имеется проект гибридного машинного перевода.

Помимо систем существуют еще программы, принцип которых основан на технологии Translation memory (память перевода). Память перевода – это двуязычные базы часто встречающихся предложений. В основе технологии лежит принцип «не переводить дважды один и тот же текст». Память перевода чаще используется как помощь переводчику, такие программы известны как CAT (Computer assisted translation – перевод с помощью компьютера).

В настоящее время системы/программы машинного перевода показывают хороший результат и значительно облегчают труд человека-переводчика. Однако до сих пор совершенства достигнуть не удалось, и гонка по созданию идеального электронного переводчика продолжается.

Литература

1. Марчук Ю. Н. Основы компьютерной лингвистики. Москва, 2000г.
2. Jonathan Slocum A Survey of Machine Translation: its history, current status and future prospects // Austin, Texas. Microelectronics and Computer Technology Corporation. 1985г.
3. Makoto Nagao Machine translation: How far can it go? Kyoto University, Translated by Norman Cook, Oxford University Press, 1989г.
4. Sergei Nirenburg and Yorick Wilks Machine transition in 50, 1986г.
5. William John Hutchins Machine translation: Past, present, future. Ellis Horwood, 1986г.