

МАШИННЫЙ ПЕРЕВОД (АВТОМАТИЧЕСКИЙ ПЕРЕВОД)

Зокир Ахмедов

*Кафедра Теория перевода, Переводческий факультет английского языка
Узбекский государственный университет мировых языков*

Машинный перевод – процесс перевода текстов (письменных, а в идеале и устных) с одного естественного языка на другой с помощью специальной компьютерной программы. Так же называется направление научных исследований, связанных с построением подобных систем.

Вместо «машинный» иногда употребляется слово автоматический, что не влияет на смысл. Однако термин автоматизированный перевод имеет совсем другое значение – при нём программа просто помогает человеку переводить тексты.

Автоматизированный перевод предполагает такие формы взаимодействия:

Частично автоматизированный перевод: например, использование переводчиком-человеком компьютерных словарей.

Системы с разделением труда: компьютер обучен переводить только фразы жёстко заданной структуры (но делает это так, чтобы исправлять за ним не требовалось), а всё, не уложившееся в схему, отдаёт человеку.

В англоязычной терминологии также различаются термины англ. machine translation, MT (полностью автоматический перевод) и англ. machine-aided или англ. machine-assisted translation (MAT) (автоматизированный); если же надо обозначить и то, и другое, пишут M(A)T.

Существуют несколько принципиально разных подходов к построению алгоритмов машинного перевода: основанный на правилах (rule-based), статистический, или основанный на статистике (statistical-based), нейронный машинный перевод (neural machine translation, NMT). Первый подход является традиционным и используется большинством разработчиков систем машинного перевода (ПРОМТ в России, SYSTRAN во Франции, Linguatrec в Германии и др.). Ко второму типу относится популярный сервис Яндекс. Переводчик, Переводчик Google, а также новый сервис от ABBYY. Сейчас большинство систем являются гибридными – сочетая правила, статистику и нейронные сети.

Статистический машинный перевод – это разновидность машинного перевода текста, основанная на сравнении больших объёмов языковых пар. Языковые пары – тексты, содержащие предложения на одном языке и соответствующие им предложения на втором, могут быть как вариантами написания двух предложений человеком – носителем двух языков, так и набором предложений и их переводов, выполненных человеком. Таким образом статистический машинный перевод обладает свойством «самообучения». Чем больше в распоряжении имеется языковых пар и чем точнее они соответствуют друг другу, тем лучше результат статистического машинного перевода. Под понятием «статистического машинного перевода» подразумевается общий подход к решению проблемы перевода, который

основан на поиске наиболее вероятного перевода предложения с использованием данных, полученных из двуязычной совокупности текстов. В качестве примера двуязычной совокупности текстов можно назвать парламентские отчеты, которые представляют собой протоколы дебатов в парламенте. Двуязычные парламентские отчеты издаются в Канаде, Гонконге и других странах; официальные документы Европейского экономического сообщества издаются на 11 языках; а Организация объединённых наций публикует документы на нескольких языках. Как оказалось, эти материалы представляют собой бесценные ресурсы для статистического машинного перевода.

С тех пор, как Уоррен Уивер (1894-1978, американский ученый, математик) распространил свою ныне известную заметку «Перевод», было предложено очень много компьютерных систем для решения проблемы перевода между различными языками. Многие из этих систем свидетельствуют о реальном понимании того, как лучше всего использовать компьютер для поддержки процесса перевода. Они помогают переводу. Другие системы нечувствительны к способностям и недостаткам как человека, так и машины. Несмотря на самые лучшие понимания проблем перевода их создателей, они часто являются препятствием для перевода. Соответствие системы перевода является относительной вещью, связанной с переводимым текстом, потребностями целевой аудитории и требованиями организации, предоставляющей перевод, среди прочих соображений. Из этого следует, что общая система перевода должна включать в себя множество средств перевода, чтобы соответствовать множеству требований к переводу.

АССОРТИМЕНТ МАШИН ДЛЯ ПЕРЕВОДА

Существует множество способов, которыми человек и машина могут разделить ответственность за перевод. Он варьируется от полностью автоматического, высококачественного перевода до человеческого перевода. Промежуточные точки занимают машинный перевод, выполняемый человеком и машинный перевод, осуществляемый человеком. Общеизвестно, что полностью автоматический, высококачественный перевод сегодня не существует. Поэтому, люди не могут описать существующие системы автоматического перевода, так как каждая из систем требует вмешательства человека для получения высококачественного перевода. Такие термины, как «полуавтоматический», «автоматизированный» и «традиционный перевод» были предложены различными наблюдателями для обозначения этого класса систем перевода. Единственным термином, который с какой-то степенью успеха соперничает со стандартным названием «машинный перевод», является «перевод при помощи машинного перевода». К сожалению, многие используют этот термин для обозначения полностью автоматического, высококачественного перевода и машинный перевод, осуществляемый человеком, упомянутых выше. «Полностью автоматический машинный перевод») имеют довольно стандартную интерпретацию, но едва ли они понятны новичку в области технологии перевода.

«Автоматический перевод» относится к любой системе, в которой компьютер производит перевод полных предложений, полностью основываясь на своих собственных ресурсах (алгоритмах, грамматиках и словарях). Тот факт, что тексты, переведенные с помощью автоматической системы

перевода, могут подвергаться предварительному или постредактированию, не делает саму систему перевода менее автоматической.

В 1979 году в Бельгии собрался международный комитет экспертов по машинному переводу под эгидой Международной федерации документации (FID) и Международной ассоциации прикладной лингвистики (AILA). Дональд Уокер и Ханс Карлгрен сообщили об заключении, сделанном комитетом: «Ожидаются обнадеживающие разработки в области усовершенствованных комбинаций машинного и человеческого сотрудничества, а не попыток полной автоматизации. Простое постредактирование машинного вывода не кажется реальным способом получения адекватного перевода».

В 1980 году Мартин Кей опубликовал свое замечательное эссе «Должное место людей и машин в языковом переводе». В нем он заявил: «Потребность в переведенных текстах не будет удовлетворена исследовательской программой, которая посвящает все свои ресурсы отдаленному идеалу, и лингвисты и специалисты по компьютерам будут лишены должного вознаграждения за свои труды, так как, они должны пообещать достичь идеал к какому-то определенному времени. Здоровым климатом для полного, качественного, машинного перевода будет тот, в котором различные, хотя и связанные между собой цели преследуются с одинаковой энергией ради интеллектуальной и практической пользы, которую они могут принести.

В 1981 году на семинаре «Прикладная компьютерная лингвистика в перспективе» комиссия по машинному переводу под председательством Мартина Кея основывала свои рекомендации на следующем: принимаются меры по...

Единственная надежда на всеобъемлющее решение, похоже, связана с технологиями. Но это не означает, что есть только одно решение, а именно машинный перевод, в классическом смысле полностью автоматическая процедура, которая переносит текст с одного языка на другой с вмешательством человека только в окончательной редакции. На самом деле существует множество способов применения технологий, с полностью автоматическим переводом на одном полюсе и оборудованием для обработки текстов, и диктофонами на другом.

В 1982 г. на конференции COLING 82, проходившей в Праге, Алан Мелби еще раз поднял этот вопрос: «Сейчас в компьютерной лингвистике вполне респектабельно разработать компьютерную систему, которая является ИНСТРУМЕНТОМ, используемым человеком-экспертом для доступа к информации, полезной для получения диагноза или другого заключения. Тогда, возможно, пришло время рассмотреть возможность того, что разработка системы машинного перевода, которая включает в себя сложную лингвистическую обработку, предназначена для использования в качестве инструмента для человека-переводчика.

В настоящее время, кажется, что точка еще не поставлена. За исключением прекрасной работы над банками терминов, которая ведется в ряде мест по всему миру, похоже, никто особо не заинтересован в том, чтобы сосредоточиться на системах перевода, ориентированных на человека. Насколько мне известно, работа, проводившаяся в Институте переводоведения университета Бригама Янга в 1970-х годах, была

единственной крупной исследовательской работой, посвященной интерактивному переводу. Его ответвление, ALPS, по-видимому, является одним из единственных коммерческих предприятий, которое занимается этим направлением развития. Мы в ALPS смотрим на эту ситуацию со смешанными чувствами: приятно стоять в стороне от толпы, тем более что мы чувствуем себя на твердой почве, но мы постоянно удивляемся тому, что никто не пытался оспорить нашу позицию.

ПРАВИЛЬНОЕ ПРИМЕНЕНИЕ МАШИНЫ ДЛЯ ПЕРЕВОДА

Даже если неясно, где лучше всего применять каждое из машинных средств перевода, должно быть очевидно, что у каждого есть свое место. Ни одна система не подходит для всех приложений. Один размер, увы, не подходит всем. Почему же тогда некоторые люди настаивают на том, что автоматический перевод не подходит нам? Почему другие продолжают действовать так, как будто это единственно возможный выбор? Поучительно спросить, какова мотивация обращения к помощи компьютера с переводом в первую очередь? Интерес в первую очередь практический или чисто академический? Некоторые люди, кажется, считают, что что-то менее амбициозное, чем полностью автоматический машинный перевод, не стоит того, чтобы прибегать к более синергетическому использованию человеческого и машинного вклада – это отговорка или какой-то обман. Мартин Кей приводит пример неправильно разработанной и неправильно примененной технологии: «Был долгий период – насколько мне известно, он еще не закончился, – когда в недрах американского правительственного учреждения каждую ночь разыгрывалась следующая комедия с целью перевода иностранных текстов на английский язык. Отрывки невинной прозы, над которыми хотели произвести эту тонкую и сложную операцию, были подвергнуты процессу вивисекции руками ничего не понимающего электронного монстра, превратившего их в запинающиеся потоки словесных обломков. Затем они были переданы в чуть более нежные руки для ремонта. Но ущерб был нанесен. Простых инструментов, которые могли бы сделать ремонтные работы (т.е. перевод) проще и эффективнее, не было, по-видимому, из-за ненасытного аппетита монстра, который не оставлял ресурсов ни на что другое. На самом деле, те лекарства, которые можно было применить к измученным остаткам этих текстов, применялись цветными карандашами на бумаге, а окончательная копия была сделана действием человеческих пальцев на клавишах пишущей машинки. Короче говоря, из довольно долгого и сложного процесса был выделен один шаг, на котором можно было осуществить автоматизацию. Выбранный шаг был наименее понятным и, совершенно очевидно, наименее подходящим для такого рода лечения. Правительство и бюрократия могут быть проникнуты печальным фатализмом, заставляющим их смотреть в будущее как на обреченное на повторение безрассудств прошлого, но мы, безусловно, можем воспользоваться моментом, чтобы удивиться безрассудствам прошлого и с ностальгией поразмыслить о том, что мир был бы более добрым и рациональным.

Будет ли мир будущего добрее или более рациональным, неизвестно. Несомненно, однако, то, что это будет мир, созданный нами самими и, следовательно, мир, которого мы заслуживаем. Область машинного перевода находится на перепутье. Мы можем разработать системы, которые делают слишком много или слишком мало. Мы можем разработать системы, которые используют особые преимущества человека и компонентов машин, или

системы, которые их игнорируют. Мы можем разработать машинный перевод, или их комбинацию. Выбор за нами. Мы стремимся к разработке гибких систем, позволяющих людям и машинам продуктивно взаимодействовать с помощью набора инструментов, соответствующих требованиям широкого спектра переводческих задач.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ:

1. Ванкуос Б. Автоматический перевод – обзор, разные подходы. Лингвистический журнал, 1, 1976, 127–35.
2. Уокер Д.Э. и Карелгрэн Х. Компьютерные средства в перевод. 25, 1980, 14. Мерл Д. Тенни 113.
3. Кей М. Надлежащее место людей и машин в языковом переводе. Исследовательский документ, Хероx Palo Alto, Исследовательский центр, 1980.
4. Кей М. Машинный перевод. Американский журнал «Компьютерная лингвистика», 8, 1982, 74-8.
5. Мелби А.К. Многоуровневый перевод помогает. В: COLING 82: Материалы 9-ой Международной конференции по вычислительной лингвистике, Прага, Чехословакия, 1982, стр. 215-20.
6. Сейджер Ж.С. Компьютер и многоязычие в ЕС., 24 года, 1979 г., 103-7.
7. Кроллманн Ф. Обработка данных у переводчика. Revue Internationale de la Перевод / Международный переводческий журнал, 1974 г., 84-8.
8. Хейс Д.Г. Машинный (ассистированный) перевод: общие положения и руководство к действию. Американский журнал Компьютерная лингвистика, микрофиша, 46, 1976, 84-8.
9. Буате К., Шателин П. и Дан Фрага П. Настоящие и будущие парадигмы в автоматизированном переводе. В: КОЛИНГ 80; Материалы 8-й Международной конференции по Компьютерной лингвистике, Токио, Япония, 1980, с. 403-6.
10. Ван Слайп Г. Гинет Дж.Ф., Сейтц Ф. и Бенежам Э. Лучший перевод для лучшей коммуникации. Нью-Йорк: Пергамон, 1983.
11. Мастерман М. Основные навыки, которые необходимо для машинного перевода. Снелл Б. (ред.). Перевод и компьютер. Амстердам: Северная Голландия, 1979, стр. 159-80.
12. Википедия. Машинный перевод – https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BF%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4.