Р. П. Ангар
Тверской государственный университет, 3 курс
Научный руководитель: к.ф.н. В. М. Мельникова

НЕОДНОЗНАЧНОСТЬ В МАШИННОМ ПЕРЕВОДЕ

Современный мир почти невозможно представить без людей, осуществляющих процесс перевода. В эпоху глобализации представители одной языковой культуры контактируют с представителями другой языковой культуры, получают и передают информацию. В связи с этим растет потребность не только в людях-переводчиках, но и в машинном переводе (далее МП). В такой функции, которая позволила бы человеку, не владеющим иностранным языком, понимать иностранный текст и общаться с представителем иной языковой культуры без посредников. С появлением МП, появляются задачи, от решения которых зависит качество машинного перевода. Одной из них является разрешение лексической неоднозначности.

Под лексической неоднозначностью подразумевается наличие у лексемы более одного значения. При машинном переводе должно быть выбрано одно из них, подходящее по контексту. Поскольку искусственный интеллект (далее ИИ) не владеет каким-либо языком и в принципе не наделен человеческим разумом, он часто не в состоянии выбрать нужное значение слова из-за нехватки, неопределённости или множественности контекста, или же, с другой стороны, контекста может быть достаточно, но перевод, тем не менее, похож на несвязанный между собой набор предложений. Так как для выбора того или иного значения слова ИИ не способен обратиться к «картине мира» или же воспользоваться экстралингвистическими знаниями, то есть не связанными с лингвистикой знаниями. Следует также уточнить, что, говоря о «лексической неоднозначности» в машинном переводе, термин «лексическая омонимия» становится неактуальным, поскольку ИИ не сможет отличить омонимию от лексической неоднозначности. «Молния» в смысле ‘искровой разряд’ от «молнии» в значении ‘застёжка’ отличить можно только в том случае, если в компьютер будет заложена чрезвычайно сложная картина мира или строго определены границы подъязыка.

Однако, нельзя не заметить, что современные системы машинного перевода (далее СМП) не перестают совершенствоваться. Если раньше перевод технических, юридических, медицинских и других текстов оставлял желать лучшего, то в настоящее время перевод таких специализированных текстов стал намного лучше. Говоря о качестве перевода более ранних СМП, одним из самых известных примеров неудачного перевода является перевод текста документации к драйверу

мыши, более известный как «Гуртовщики Мыши». Еще один не менее известный случай – обращение известной поп-исполнительницы Мадонны к российским фанатам на страницах своего блога. Обращение начиналось так: *«Спасибо 50.000 ecstatic русских вентиляторов присутствовали на моих исповедях показывают в moscow! Выставкой был взрыв! Я нахожусь на дороге back to гостиница теперь и я как раз хотел препятствую вам знать что все одобрено. Штат обеспеченностью сделал чудесную работу делать уверенной выставка был безопасен! Спасибо Россия!»* [Lenta.ru URL].

Такие слова как «fans», «show» и «security staff» программа перевела как «вентиляторы», «выставка» и «штат обеспеченностью» соответственно. Некоторые слова («ecstatic», «moscow», «back to») остались в исходном виде. Попробуем перевести первоначальное обращение Мадонны с помощью одной из современных систем машинного перевода.

Таблица 1.

Оригинал	Машинный перевод
Thanks to the 50.000 ecstatic Russian fans who attended my Confessions show in Moscow! The show was a blast! I'm on the way back to the hotel now and I just wanted to let you know that everything is ok. The security staff did a wonderful job of making sure the show was safe! Thanks Russia!	Спасибо 50 000 восторженных российских фанатов, которые посетили мое шоу "исповедь" в Москве! Шоу было просто потрясающим! Я сейчас возвращаюсь в отель и просто хочу, чтобы вы знали, что все в порядке. Сотрудники Службы безопасности проделали замечательную работу, чтобы убедиться, что шоу было безопасным! Спасибо России!

Как мы видим, современный машинный перевод выглядит куда более логичным и понятным. Вся морфологическая информация, т.е. согласование слов, падежи, перевод глаголов и многие другие трудности в данном варианте перевода полностью решены.

Приведём еще несколько примеров неудачных машинных переводов: фразу «our cat gave birth to three kittens — two whites and one black» онлайн-переводчик «ПРОМТ» (версия 7.0, 2007) трансформировал в «наш кот родил трёх котят — двух белых и одного афроамериканца». Если «афроамериканца» ещё можно было сделать «чёрным», немного изменив контекст и написав «black kitten», то «коту» так и не удавалось сменить пол (например, female cat переводился как «самка кот»). Попробуем перевести эту фразу с помощью современного онлайн-переводчика «PROMT.One».

Таблица 2.

Оригинал	Машинный перевод (PROMT.One)
Our cat gave birth to three kittens — two whites and one black	Наша кошка родила трех котят — двух белых и одного черного

Как видно, абсолютно вся информация английского предложения передана верно и не требует никаких исправлений.

В большинстве случаев подобные ошибки были связаны с тем, что программа не распознавала контекст фразы и переводила термины буквально, к тому же, не отличая имён собственных от обычных слов. Тот же переводчик ПРОМТ трансформировал «Лев Толстой» в «Lion Thick» («толстый лев»), «bra-ket notation» в «примечание Кети лифчика» («бра и кет»), «Lie algebra» — в «алгебру Лжи» («алгебра Ли»), «eccentricity vector» — в «вектор оригинальности» («вектор эксцентриситета»), «Shawnee Smith» в «индеец племени шони Смит» и т. п. Тем временем переводчик Google, наоборот, слово «rice» часто принимал за фамилию госсекретаря США (Condoleezza Rice) [Википедия URL].

Можно сказать, что проблема разрешения лексической неоднозначности в машинном переводе сводится к поиску наиболее подходящего перевода слова. Эта проблема затрагивается почти в каждой модели перевода. В ней есть несколько аспектов [Марчук 2016]:

1) взаимодействие предметных словарей;
2) взаимодействие предметных словарей и словарей общеупотребительных слов.

Проблема взаимодействия словарей в машинном переводе является крайне трудно разрешимой. Каждый научный текст может быть полон лексики, выходящей за границы какого-либо определенного стиля, превалирующем в этом тексте. Так, например, в юридическом тексте, уже наполненном своей специализированной лексикой, может содержаться описание какого-либо процесса или явления связанного с медициной, в котором будут присутствовать медицинские термины, не входящие в состав традиционной юридической лексики.

Когда машинный перевод только начинал зарождаться, проблема разрешения неоднозначности уже была сформулирована как отдельная задача [Weaver URL]. Так, в работе [Киселёв 1984] была разработана идея использования специальных словарей-конкордансов, составленных для определённых предметных областей. Цель такого словаря заключается в выявлении характерных наборов детерминант с последующим использованием этих наборов в алгоритмах разрешения неоднозначности по контексту. Однако же, не остался незамеченным тот факт, что в реальных текстах практически не встречается лексика строго обособленной предметной области.

Таким образом, можно сделать вывод о том, что следующая последовательность действий будет эффективна при разрешении неоднозначности:

1) статистический анализ лексики в тексте, позволяющий определить преобладающие и второстепенные в данном тексте предметные области;
2) составление словаря-конкорданса по материалам обрабатываемых текстов. Каждое слово будет представлено в контексте своего употребления. Конкорданс может быть составлен на основе параллельных текстов, т.е. текстов оригинала и перевода;
3) уже на основе конкорданса составляется контекстологический словарь, в котором каждое значение неоднозначного слова сопоставляется с набором контекстных детерминант, с помощью которого разрешается неоднозначность.

Предполагается, что выполнение такой последовательности действий позволит свести степень неоднозначности текста к минимуму и облегчить процесс понимания.

ЛИТЕРАТУРА

Википедия [Электронный ресурс]. URL: https://ru.wikipedia.org/wiki/Машинный_перевод

Киселев А.Н. Исследование и алгоритмическое разрешение омографии при машинном переводе методом схемного анализа по структурным типам: дис. … канд. филол. наук. М.: Военный Краснознаменный ин-т, 1984. 125 с.

Марчук Ю.Н. Контекстное разрешение лексической многозначности // Вестник Московского государственного областного университета. Серия: Лингвистика. 2016. № 1. С. 26-32.

Lenta.ru [Электронный ресурс]. URL: https://lenta.ru/news/2006/09/13/madonna/

W. Weaver. The “Translation” memorandum. In Machine Translation of Languages: Fourteen Essays, ed. by Locke, W.N. and Booth, A.D. Cambridge, MA: MIT Press. 1949. [Электронный ресурс]. URL: http://www.hutchinsweb.me.uk/MTNI-22-1999.pdf