**Проблема преодоления омонимии и полисемии лексических единиц при нейронном и гибридном машинном переводе**

***Лотарева И.П.***

*студент, магистр*

*Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М. В. Ломоносова», Высшая школа перевода, Москва, Российская Федерация*

[*i\_p\_lotareva@mail.ru*](mailto:i_p_lotareva@mail.ru)

Нейронный машинный перевод (НМП), – «вид машинного перевода с использованием искусственной нейронной сети» [6, 2023, c. 38]. Гибридный машинный перевод (ГМП) основан на сочетании «статистического метода с нейронными сетями» [7, 2019, эл. ресурс], при котором программа каждый раз может обратиться или к статистической, или к нейросетевой технологии, и обучается в т. ч. на параллельных текстах. При исключительно НМП обучение идёт с нуля. Нейросеть – «совокупность нейронов, связанных между собой и функционирующих по определенным правилам» [8, 2024, с. 16], реализация механизма ИИ. Речь идёт об искусственных нейронах, «которые имитируют нейроны в мозге» [1, 2023, с. 13]. ИИ для программ – не полноценный искусственный разум, а «способность системы…интерпретировать внешние данные, извлекать сведения из анализа таких данных и использовать эти сведения для достижения конкретных целей в решении задач при помощи гибкой адаптации» [8, 2024, с. 11], в т.ч. обучаться, анализировать и фиксировать соотносящиеся данные. Слои искусственных нейронов сгруппированы по роли в обработке данных: «входной, скрытые и выходной» [1, 2023, с. 13]. Основная обработка проходит в скрытых слоях. При этом все искусственные нейроны связаны с нейронами последующих слоёв «через веса, …» определяемые и настраиваемые «…в процессе обучения» [1, 2023, с. 13]. «Вес» является рассчитанной вероятностью проявления значений – в некоторой данной последовательности, и в зависимости от контекста.

Для НМП и ГМП требуется «рекуррентная нейронная сеть», такая нейронная сеть, в которой связи элементов «образуют направленную последовательность» [4, 2018, с. 74], подходящую для линейной «обработки сигналов и текстов» [2, 2019, эл. ресурс]. НМП с активной ролью людей входит в объём понятия «цифровой перевод», по определению Н. К. Гарбовского и О. И. Костиковой, это система «взаимодействия когнитивно-коммуникативной деятельности переводчика-человека…» с ИИ [3, 2019, эл. ресурс]. Тем не менее, даже цельная двусторонняя проверка в рекуррентной нейронной сети может не гарантировать верное разрешение неоднозначности, т. е. омонимии или полисемии той или иной единицы. При полисемии «у единицы языка» возможно более одного значения [5, 2020, с. 175]. «Омонимия» – явление, при котором наблюдается «звуковое совпадение различных языковых единиц, значения которых не связаны друг с другом» [9, 1990, эл. ресурс], совпадение внешних форм.

Для эксперимента использовались программы «Яндекс. Переводчик» и «Google Translate» с ГМП, «DeepL» как программа НМП. В научно-технической исследовательской работе «Environmental impact of winter maintenance with salt» содержится ряд примеров неоднозначности. В качестве потенциально трудных для НМП и ГМП случаев омонимии мы выбрали лексические единицы «molasses» [10, 2021, 16], «deposit» [10, 2021, 45]. Яндекс. Переводчик», «Google Translate» и «DeepL» определили нужную область при наличии контекста в предложении, и подобрали подходящие теме варианты («меласса», «залежи»). При полисемии мы обнаружили больше ошибок: «рассол» вместо «раствора соли» для brine; «выпаривание» вместо принятого «вываривания» для evaporation (evaporated salt), «разбрасывание» вместо «нанесения» соли или «обработки» для spreading, и т. д. В данной связи возникает практическая проблема – необходимость проводить каждую правку вручную без сохранений, при этом предложенные переводы остаются в сети в качестве списка второстепенных вариантов. Как показал эксперимент, неспособность программ преодолеть неоднозначность при полисемии не зависит от объёма введённого англоязычного текста: от двух предложений до максимально доступного (Яндекс: 10000 знаков, Google: 5000 знаков, и DeepL – 1500 знаков, около страницы). Изменения за границей абзаца (в том числе преднамеренно неверно отделённого, не авторского) также не имели влияния на готовый текст.

Омонимия и полисемия по-прежнему могут создавать значительное препятствие для машинного перевода. При явлении омонимии программа считывает один из различных лексико-семантических вариантов в зависимости от контекста, поэтому она чаще справляется с переводом омонимов. При явлении полисемии для адекватности перевода особенно нужен человеческий интеллект, так как значимым может быть проявление оттенков значения. Как показало исследование, разрешение неоднозначности актуально при работе с текстами научно-технической тематики, особенно при обращении к машинному переводу с нейросетевой технологией.

Литература

1. Антипко, А. В. Нейросеть: определение, принцип работы, область применения. Специалисты по нейросетям / А. В. Антипко. — Текст: непосредственный // Молодой ученый. — 2023. — № 33 (480). — С. 13–14. — [Электронный ресурс]. URL: https://moluch.ru/archive/480/105505/ (дата обращения: 12.02.2025).
2. Воронцов, К. В. Курс «Машинное обучение». 2019. [Электронный ресурс]. URL: [Воронцов. Курс «Машинное обучение» 2019 (Школа анализа данных) — Журнал «Я Robot»](https://ya-r.ru/2020/05/07/vorontsov-kurs-mashinnoe-obuchenie-2019-shkola-analiza-dannyh/) (дата обращения: 16. 02. 2025).
3. Гарбовский Н. К., Костикова О. И. Интеллект для перевода: искусный или искусственный? // Вестник Московского университета. Серия 22. Теория перевода. 2019. №4. [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/intellekt-dlya-perevoda-iskusnyy-ili-iskusstvennyy (дата обращения: 31.10.2024).
4. Гафаров, Ф. М. Искусственные нейронные сети и приложения: учеб. пособие / Ф.М. Гафаров, А.Ф. Галимянов. – Казань: Изд-во Казан. ун-та, 2018. – 121 с.
5. Ермолаева, А. А. Полисемия в контексте нейронного машинного перевода / А. А. Ермолаева. — Текст: непосредственный // Молодой ученый. — 2020. — № 32 (322). — С. 175–177. — URL: https://moluch.ru/archive/322/73064/ (дата обращения: 01.03.2025).
6. Зайцева, М. А. О типологии ошибок гибридной системы машинного перевода «Яндекс Переводчик» / М. А. Зайцева. — Текст: непосредственный // Молодой ученый. — 2023. — № 45 (492). — С. 38–46. — [Электронный ресурс].URL: https://moluch.ru/archive/492/107574/ (дата обращения: 12.02.2025).
7. Мифтахова, Р. Г. МАШИННЫЙ ПЕРЕВОД. НЕЙРОПЕРЕВОД / Р. Г. Мифтахова, Е. А. Морозкина // Вестник Башкирск. ун-та. – 2019. – №2. [Электронный ресурс] URL: <https://cyberleninka.ru/article/n/mashinnyy-perevod-neyroperevod> (дата обращения: 16.04.2024).
8. Урбанович, П. П. Нейросетевые технологии в криптографических приложениях: монография / П. П. Урбанович, М. Д. Плонковски, М. Долецки. – Минск: БГТУ, 2024. – 223 с. ISBN 978-985-897-160-1.
9. Ярцева, В. Н. Лингвистический энциклопедический словарь. — М.: Советская энциклопедия.Гл. ред. В. Н. Ярцева. 1990.— 682,[3] с.: ил.; 27 см.
10. Götzfried, Franz ENVIRONMENTAL IMPACT OF WINTER MAINTENANCE WITH SALT / Stéphanie Gaudé, Dr. Franz Götzfried – 2021, 56 p.