# ЛИНГВИСТИЧЕСКИЕ ПРОБЛЕМЫ МАШИННОГО ПЕРЕВОДА

# Ткачёва Алиса Алексеевна

Студентка Казанского инновационного университета им. В. Г. Тимирясова (ИЭУП), г. Казань, Россия

e-mail:alisa200585@gmail.com

Машинный перевод (МП) - выдающееся достижение современного мира, но на пути к совершенству ему приходится сталкиваться с препятствиями. Одним из самых сложных камней преткновения является проблема лексической многозначности, или полисемии. Эта проблема не только препятствует точности перевода, но и затрагивает основы понимания языка, как машинами, так и людьми.

Полисемия, от греческого «polysemos» (многозначный), означает, что языковая единица (чаще всего слова) имеет несколько значений.

Существует лексическая полисемия, которую можно описать как явление, при котором одно и то же слово служит для выражения разных объектов или направлений. (Например, «ключ» — это инструмент для открывания замка или музыкальный инструмент, «брак» - может означать супружество или испорченные продукты). Грамматическая полисемия, когда формы разных лексем совпадают (например, «бег» — это процесс или вид спорта). Выяснить, какое значение слова уместно в конкретном контексте, - сложная задача, которую человек может легко решить, интуитивно полагаясь на свой опыт и знание мира. Для машины это серьезный вызов.

Современные программы MП используют целый ряд методов для решения проблемы лексической многозначности:

- Контекстный анализ: алгоритмы анализируют окружающие слова и предложения, выявляя ключевые слова и синтаксические связи. Более продвинутые системы учитывают не только непосредственный контекст, но и более широкий параграф, весь текст. Однако даже продвинутый контекстный анализ не всегда четко указывает на нужный смысл, особенно в случае неоднозначных или идиоматических выражений

- Встроенные словари и онтологии: Системы МП используют обширные базы данных, содержащие информацию о словах, включая их различные значения, синонимы, антонимы и семантические связи. Онтологии - формализованные описания предметных областей - помогают системам лучше понять взаимосвязи между понятиями.

- Машинное обучение: Современные системы МП активно используют методы машинного обучения, в частности глубокого обучение. Нейронные сети обучаются на огромных текстовых массивах, учатся распознавать контекст и выбирать наиболее вероятное значение слова. Однако здесь есть свои ограничения: нейронные сети могут «заучивать» статистические корреляции, которые не всегда отражают истинное семантическое значение.

Мы взяли 40 наиболее употребительных идиом современного английского языка [100 IDIOMS YOU MUST KNOW FOR SSC CGL, 2019: 2-7]. и перевели их в четырех системах машинного перевода (Яндекс переводчик, Google Translate, Deepl, Reverso). Анализ работы МП с фразеологическими единицами позволил нам выделить наиболее часто встречающиеся ошибки, допускаемые автоматическим переводом, а также оценить общую производительность различных систем перевода. Мы пришли к выводу, что наиболее качественным в контексте понимания и перевода идиом наилучший результат продемонстрировал «Яндекс переводчик» точность - 19 из 40, со своим функции «пояснения» с помощью специального бота. На втором месте следует «Reverso», который также сгенерировал 17 правильных значений. Google Translate – 14, наименее эффективным из онлайн-переводчиков оказался DeepL с точностью 8 из 40, который часто переводил идиоматические выражения дословно, искажая суть исходного текста.

**Литература**

1. 100 IDIOMS YOU MUST KNOW FOR SSC CGL. - oliveboard, 2019.