

**СЛОЖНОСТИ В АВТОМАТИЧЕСКОЙ ОБРАБОТКЕ И
МАШИННОМ ПЕРЕВОДЕ ТЕКСТА ПРИ НАЛИЧИИ
ОМОНИМИИ**

Самандарова Г.С.

старший преподаватель

Университет Альфраганус

Ташкент, Узбекистан

***Аннотация.** В данной статье рассматривается ряд сложностей, связанных с омонимией и другими особенностями естественного языка при автоматической обработке текста и машинном переводе. Анализируются проблемы, которые могут существенно повлиять на качество перевода и анализ текста. Приводятся примеры основных трудностей, с которыми сталкиваются системы обработки естественного языка (NLP)¹ и машинного перевода, делаются выводы о том, как преодолеть и решить эти проблемы.*

***Ключевые слова:** омонимия, обработка текста, машинный перевод, естественный язык, многозначность, полисемия, алгоритм.*

«Машинный перевод привлекает исследователей и разработчиков массовой потребностью людей в автоматическом переводе и сложностью задачи»². Одной из самых больших проблем в автоматической обработке текста является омонимия, то есть многозначность слов. Омонимы могут

¹ NLP – Natural Language Processing – обработка естественного языка.

² Ганегедара Т. Г. Обработка естественного языка с TensorFlow / пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2020. – 37 стр.

иметь несколько значений в зависимости от контекста, и системы NLP должны уметь правильно интерпретировать эти значения. Машинные алгоритмы часто испытывают трудности в определении контекста, что может привести к неверной интерпретации.

Например, слово "банк" в русском языке может означать как финансовое учреждение ("bank" в английском), так и берег реки ("shore" или "riverbank" в английском). Без учета контекста машинный перевод может выбрать неправильное значение.

«Омонимия значительно усложняет машинный перевод, поскольку одно и то же слово может иметь несколько значений, и правильный перевод зависит от контекста, в котором это слово используется»³. Кроме омонимии, проблемы возникают из-за полисемии, когда одно слово имеет несколько связанных значений. Это особенно затрудняет работу машинных переводчиков, поскольку им необходимо не просто выбрать одно из значений, но и правильно интерпретировать нюансы, которые могут изменять смысл предложения.

Например, слово "run" в английском языке может означать "бежать", "управлять" (компанией), "работать" (о машине), "издавать" (журнал) и т.д. Выбор правильного значения без учета контекста может привести к неадекватному переводу.

Многие языки имеют сложные синтаксические структуры, которые трудны для автоматического анализа. Многозначные слова могут выполнять разные синтаксические функции (например, быть как существительными, так и глаголами), и алгоритмам нужно правильно

³ Самандарова Г. С. ВАЖНОСТЬ ИЗУЧЕНИЯ ОМОНИМОВ ДЛЯ ПЕРЕВОДА В ВЕКЕ ВЫСОКИХ ТЕХНОЛОГИЙ, Межвузовский международный конгресс, Высшая школа, – Москва: Издательство Инфинити, 2024. – 62 с.

определять эти функции на основе контекста. Ошибки в синтаксическом анализе могут привести к неправильному переводу или генерации текста.

Например, в предложении "Она видит утку, которая плывет в реке" слово "утка" может быть и существительным, и глаголом в зависимости от контекста. Машинный переводчик должен правильно интерпретировать синтаксическую роль слова "утка", чтобы избежать ошибок.

Языки часто содержат культурные особенности, идиомы, метафоры и фразеологизмы, которые трудно перевести на другой язык с сохранением смысла. Автоматические системы перевода часто сталкиваются с трудностями при интерпретации таких выражений, так как они требуют глубокого понимания культурного контекста.

Например, английское выражение "kick the bucket" означает "умереть", но дословный перевод на русский "пнуть ведро" будет бессмысленным. Машинные переводчики часто делают такие ошибки, не распознавая идиоматический характер фразы.

Некоторые предложения могут быть неоднозначными и иметь несколько возможных интерпретаций. Машинные алгоритмы часто не могут определить правильную интерпретацию, особенно если фраза сложная или содержит множество омонимов.

Например, фраза "I saw the man with a telescope" может означать как "Я видел человека с телескопом", так и "Я видел человека через телескоп". Без контекстуальной информации автоматический переводчик может выбрать неправильный вариант.

Языки с развитой морфологией (русский, финский, венгерский, арабский) представляют особую трудность для автоматической

обработки по сравнению с аналитическими языками (английский, китайский). Слова могут изменять форму в зависимости от падежа, числа, рода и других грамматических характеристик. Это создает дополнительные сложности для автоматической обработки текста, так как алгоритмам нужно уметь правильно разбирать и интерпретировать такие формы. Ошибки в морфологическом анализе могут привести к неправильным переводам или несогласованности в тексте. Морфологическая сложность усиливает лексическую неоднозначность, поскольку одна словоформа может соответствовать: нескольким лексемам, нескольким грамматическим интерпретациям, нескольким семантическим ролям. В машинном переводе системы сталкиваются с необходимостью: корректно интерпретировать морфологию исходного языка, сгенерировать грамматически и семантически согласованные формы целевого языка. «Для морфологически богатых языков это особенно сложно, так как ошибка в определении рода, падежа или вида глагола приводит к неестественному или неверному переводу⁴».

Многие современные системы машинного перевода, такие как Google Translate или DeepL, основаны на методах машинного обучения и нейронных сетей. Однако эти системы часто зависят от большого объема данных и могут не учитывать контекст в широком смысле.

«Существует ряд проблем решения омонимии при машинном переводе.

1. Недостаток контекста: если в предложении слишком мало информации, модели могут ошибиться. Например, фраза «Он смотрел на

⁴ Philipp Koehn. Statistical Machine Translation. Cambridge, UK, 2010. p.201

банк» может быть двусмысленной даже для человека без дополнительных данных.

2. Различия в грамматике и структуре: иногда контекстный анализ затрудняется из-за грамматических или синтаксических различий между исходным и целевым языками, что приводит к неверному выбору значения слова.

3. Трудности с малораспространёнными языками: для языков, которые не имеют больших корпусов текстов, обучение машинному переводу может быть затруднено, что приводит к большим ошибкам при разрешении омонимии»⁵.

Хотя они и могут обрабатывать текст на основе обучающего корпуса, они часто не справляются с редкими или неизвестными контекстами.

Например, если система не была обучена на конкретном омониме или сочетании слов, она может выбрать неправильный перевод. Это особенно заметно в случае новых слов или сленга.

Сложные грамматические конструкции, такие как вложенные предложения или редкие формы слов, могут представлять проблемы для машинных переводчиков. Такие конструкции могут содержать омонимы или многозначные слова, что еще больше усложняет перевод.

Например, предложения с большой вложенностью или многозначными глаголами могут быть неправильно интерпретированы, если алгоритмы не способны правильно разбирать каждую часть предложения.

⁵ Самандарова Г. С. ВАЖНОСТЬ ИЗУЧЕНИЯ ОМОНИМОВ ДЛЯ ПЕРЕВОДА В ВЕКЕ ВЫСОКИХ ТЕХНОЛОГИЙ, Межвузовский международный конгресс, Высшая школа, – Москва: Издательство Инфинити, 2024. – 66с.

В устной речи часто встречаются неоднозначности, такие как омонимия, многозначность и идиомы. Машинные системы обработки речи сталкиваются с дополнительными трудностями, поскольку им нужно учитывать не только лексические, но и фонетические аспекты. Ошибки в распознавании речи могут привести к неправильной интерпретации омонимов и фразеологизмов.

Омонимия и связанные с ней проблемы многозначности создают значительные трудности для автоматической обработки текста и машинного перевода. Эти трудности обусловлены сложностями в синтаксическом, морфологическом и семантическом анализе, необходимостью учета контекста, культурных особенностей и стилистических эффектов. Для повышения качества перевода и автоматической обработки текста необходимы улучшенные алгоритмы, учитывающие многозначность слов и контекста, а также более продвинутые методы машинного обучения, способные справляться с различными языковыми явлениями.

СПИСОК ЛИТЕРАТУРЫ

1. Виноградов В. В. «Об омонимии и смежных явлениях» Вопросы языкознания. 1960. № 5. С. 3
2. Мучник Б. С. «Многозначность и многовыраженность», Русский язык в школе. 1991. №1. С. 27 — 31.
3. Ганегедара Т. Г. Обработка естественного языка с TensorFlow / пер. с англ. В. С. Яценкова. – М.: ДМК Пресс, 2020. – 37 стр.
4. Самандарова Г. С. ВАЖНОСТЬ ИЗУЧЕНИЯ ОМОНИМОВ ДЛЯ ПЕРЕВОДА В ВЕКЕ ВЫСОКИХ ТЕХНОЛОГИЙ, Межвузовский

FRONTIERS OF KNOWLEDGE AND INTERDISCIPLINARY DISCOVERY
international scientific conference

международный конгресс, Высшая школа, – Москва: Издательство Инфинити, 2024. – 66с.

5. Philipp Koehn. Statistical Machine Translation. Cambridge, UK, 2010.
p.201