

ИССЛЕДОВАНИЕ МЕТОДОВ АВТОМАТИЧЕСКОГО ПЕРЕВОДА ТЕКСТОВ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

***Аннотация.** Машинный перевод стал неотъемлемой частью цифровой коммуникации: он применяется в онлайн-сервисах, корпоративных платформах, автоматизации технической документации и образовательных системах. Развитие искусственного интеллекта привело к смещению парадигмы — от статистических алгоритмов и правил к нейронным моделям, способным обучаться на больших корпусах данных и адаптироваться к контексту текста.*

Цель статьи — исследовать методы автоматического перевода, определить факторы, влияющие на эффективность машинного перевода, и провести практическую реализацию нейронной модели, способной переводить выражения с английского языка на русский.

***Ключевые слова.** машинный перевод, нейронный машинный перевод, рекуррентные нейронные сети, LSTM, Transformer, encoder–decoder, attention-механизм, параллельный корпус, токенизация, эмбединги.*

Введение

Машинный перевод (МП) — это автоматическое преобразование текста с одного языка на другой с помощью программных алгоритмов, основанных на машинном обучении и обработке естественного языка. Применение МП становится повсеместным: он используется в глобальных платформах коммуникации, службах поддержки клиентов, переводе образовательного контента и научной документации.

Развитие нейронных алгоритмов позволило улучшить качество перевода, увеличить точность восприятия контекста и сделать текст более

естественным. Тем не менее эффективность МП зависит от множества факторов: структуры фразы, сложности входного текста, контекста и уровня владения пользователем языком.

Цель исследования:

разработать и протестировать модель машинного перевода на основе нейронных сетей и систематизировать существующие подходы к автоматическому переводу.

Задачи:

1. Изучить основные модели машинного перевода.
2. Реализовать нейросетевую модель для перевода английских выражений на русский.
3. Провести анализ результатов модели.

Теоретические основы машинного перевода

Терминология

Термин	Определение
Машинный перевод (Machine Translation, MT)	Автоматический перевод текста между языками с помощью алгоритмов.
Encoder	Часть нейросети, кодирующая входное выражение в скрытое представление.
Decoder	Компонент, генерирующий перевод на основе информации, полученной от encoder'а.
LSTM (Long Short-Term Memory)	Тип рекуррентной нейронной сети, способный обрабатывать последовательность и учитывать контекст.
Dataset (корпус данных)	Набор параллельных предложений на двух языках для обучения модели.

Современный тренд — переход от LSTM к Transformer (GPT/DeepL), так как архитектура трансформеров лучше воспринимает контекст.

Машинный перевод как интеллектуальная система

МП как лингвистический процессор

Система анализирует структуру предложения и генерирует перевод, опираясь на статистические закономерности.

МП как средство коммуникации

Перевод становится частью взаимодействия человека и системы: пользователь корректирует выход, формирует стиль.

Эксперимент: реализация нейросети для машинного перевода

Подготовка данных

Модель обучена на корпусе пар выражений «English → Russian». Данные очищались, токенизировались и преобразовывались в векторы.

144486	Please begin immediately.	Пожалуйста, начните немедленно.	CC-BY 2.0 (Fr
144487	Please behave yourselves.	Пожалуйста, ведите себя прилично.	CC-BY 2.0
144488	Please behave yourselves.	Ведите себя прилично, пожалуйста.	CC-BY 2.0
144489	Please buy me three pens.	Купи мне три ручки, пожалуйста.	CC-BY 2.0 (Fr
144490	Please buy me three pens.	Купите мне три ручки, пожалуйста.	CC-BY 2.0
144491	Please call an ambulance.	Вызовите скорую, пожалуйста.	CC-BY 2.0 (Fr
144492	Please call an ambulance.	Вызови скорую, пожалуйста.	CC-BY 2.0 (France
144493	Please choose one person.	Выберите одного человека, пожалуйста.	CC-BY
144494	Please choose one person.	Выбери одного человека, пожалуйста.	CC-BY 2.0
144495	Please come to my office.	Пожалуйста, приходите ко мне в офис.	CC-BY
144496	Please come to the point.	Изяснитесь удовлетворительнее.	CC-BY 2.0 (Fr
144497	Please come to the point.	Пожалуйста, говорите по существу.	CC-BY 2.0
144498	Please come to the point.	Пожалуйста, переходите к сути вопроса.	CC-BY
144499	Please come to the point.	Пожалуйста, переходите к сути дела.	CC-BY 2.0
144500	Please come to the point.	Прошу вас, ближе к делу.	CC-BY 2.0 (France

Рисунок 1. Скриншот из реализации модели.

Архитектура модели

- **Encoder (LSTM)** получает входную фразу и кодирует её.
- **Decoder (LSTM)** на основе скрытого состояния генерирует перевод.

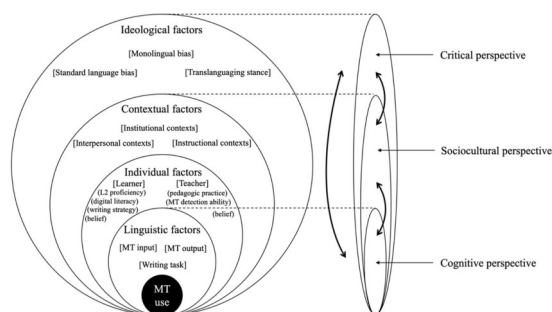


Рисунок 2. Скриншот из реализации модели.

Тестирование модели

Фраза (EN)	Перевод модели	Литературный перевод
------------	----------------	----------------------

Work hard. Dream big.	Работа трудная. Мечта большая.	смысл передан
-----------------------	--------------------------------	---------------

Результаты

Нейросеть:

- успешно распознавала простые фразы,
- допускала погрешности при сложных конструкциях,

У всех конструкций, рассмотренных до этого, значения на выходах определялись тем, что у них на входах в данный момент. Соответственно, от порядка предъявления примеров ничего не зависело. Если же мы попытаемся работать с временными рядами, в них следующее состояние будет зависеть от предыдущего. Для того чтобы смоделировать это следующее состояние с помощью нейронной сети есть два варианта: указать предыдущее состояние к каждому имеющемуся примеру, либо сделать так, чтобы это помнил не пример, а сеть, – это можно обеспечить с помощью наличия обратных связей внутри самой сети. $y(t) = f[x(t), x(t-1), \dots, x(t-n), y(t-1), \dots, y(t-m)]$

Заключение

Машинный перевод прошёл путь от словарных таблиц к сложным нейросетям. Нейронные модели обеспечивают естественность перевода и способность учитывать контекст. Практическая реализация показала потенциал LSTM для перевода коротких фраз, однако масштабирование корпуса данных и переход к архитектуре Transformer значительно повысят качество.

Список литературы

- [1] Arksey H., O'Malley L. Scoping studies: Towards a methodological framework // International Journal of Social Research Methodology. — 2005. — Т. 8, № 1. — С. 19–32.
- [2] Chung E. S., Ahn S. The effect of using machine translation on linguistic features in L2 writing across proficiency levels and text genres // Computer Assisted Language Learning. — 2022. — Т. 35, № 9. — С. 2239–2264.

[3] Garcia I., Pena M. Machine translation-assisted language learning: Writing and grammar // Computer Assisted Language Learning. — 2011. — T. 24, № 5. — С. 471–487.

[4] Hutchins W. J. Machine translation: History of research and applications // Handbook of translation technology. — Routledge, 2023. — С. 128–144.