

Презентацию подготовила Виктория Фирсанова  
СПбГУ, кафедра математической лингвистики

/01

# Attention Is All You Need, Vaswani et al.

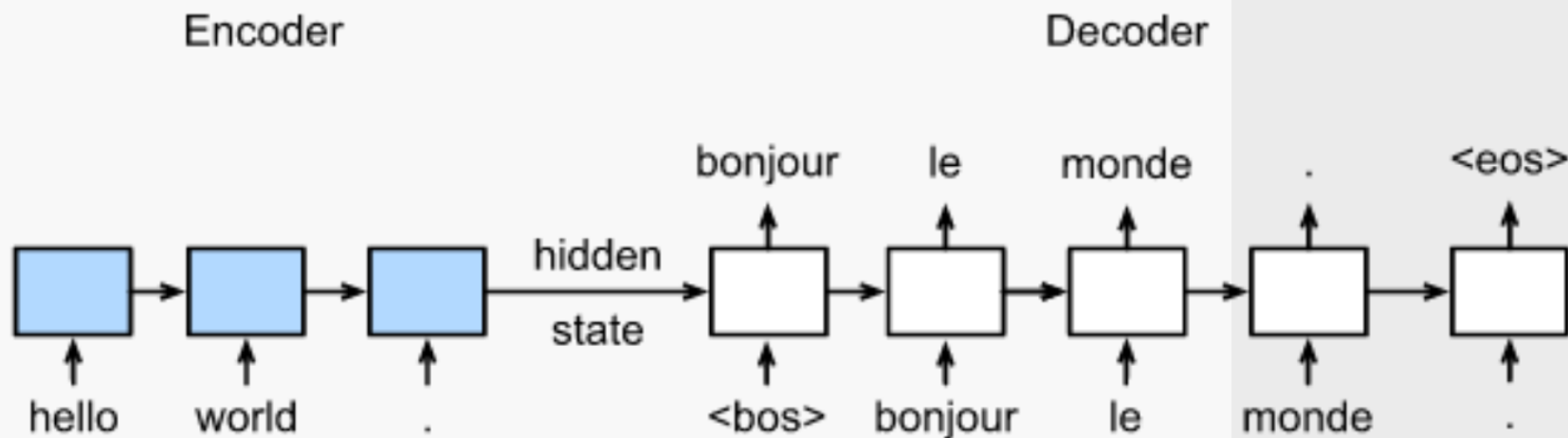


Дискуссионный клуб NLP | 2020

# Обзор

## Проблема

- Большинство моделей для последовательного предсказания:  
RNN/CNN + encoder & decoder

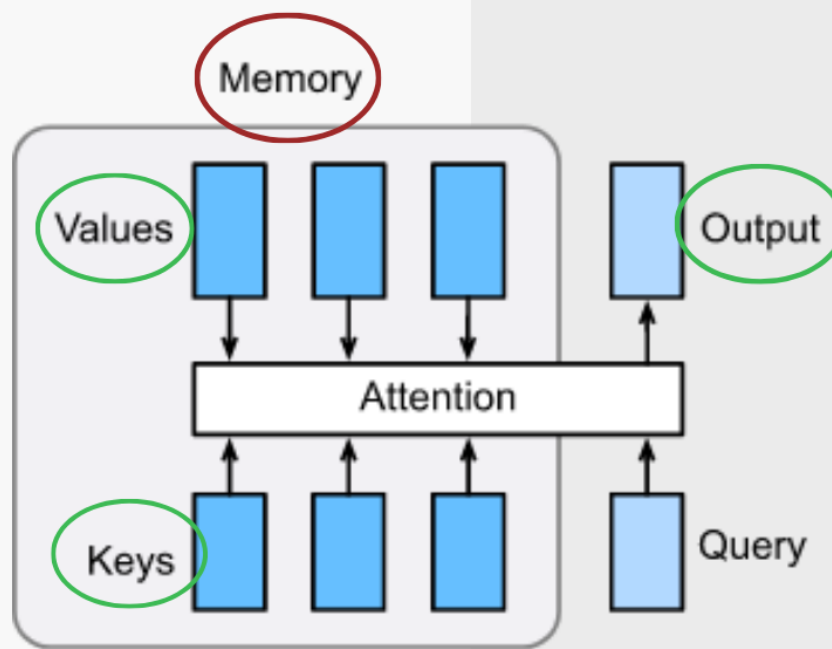


/02

# Обзор

## Проблема

- Иногда кодер и декодер соединены с помощью механизма внимания (attention mechanism)
- Слой внимания:  
запрос -> ключ-значение



# Обзор

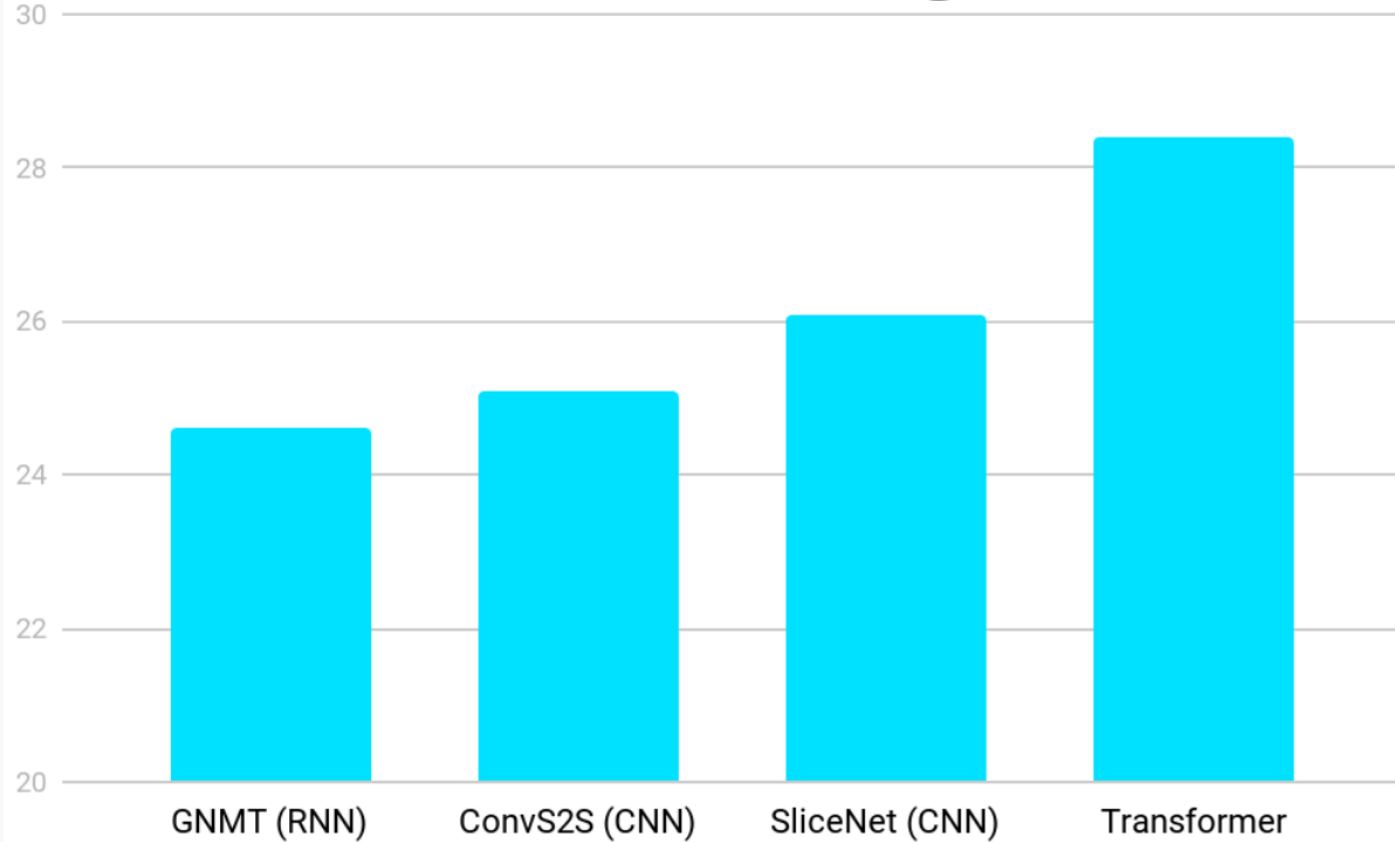
## Решение

- Трансформер:
  - в основе - только механизмы внимания;
  - более гибкие, чем RNN и CNN;
  - быстрее обучаются;
  - не боятся ни больших, ни ограниченных датасетов

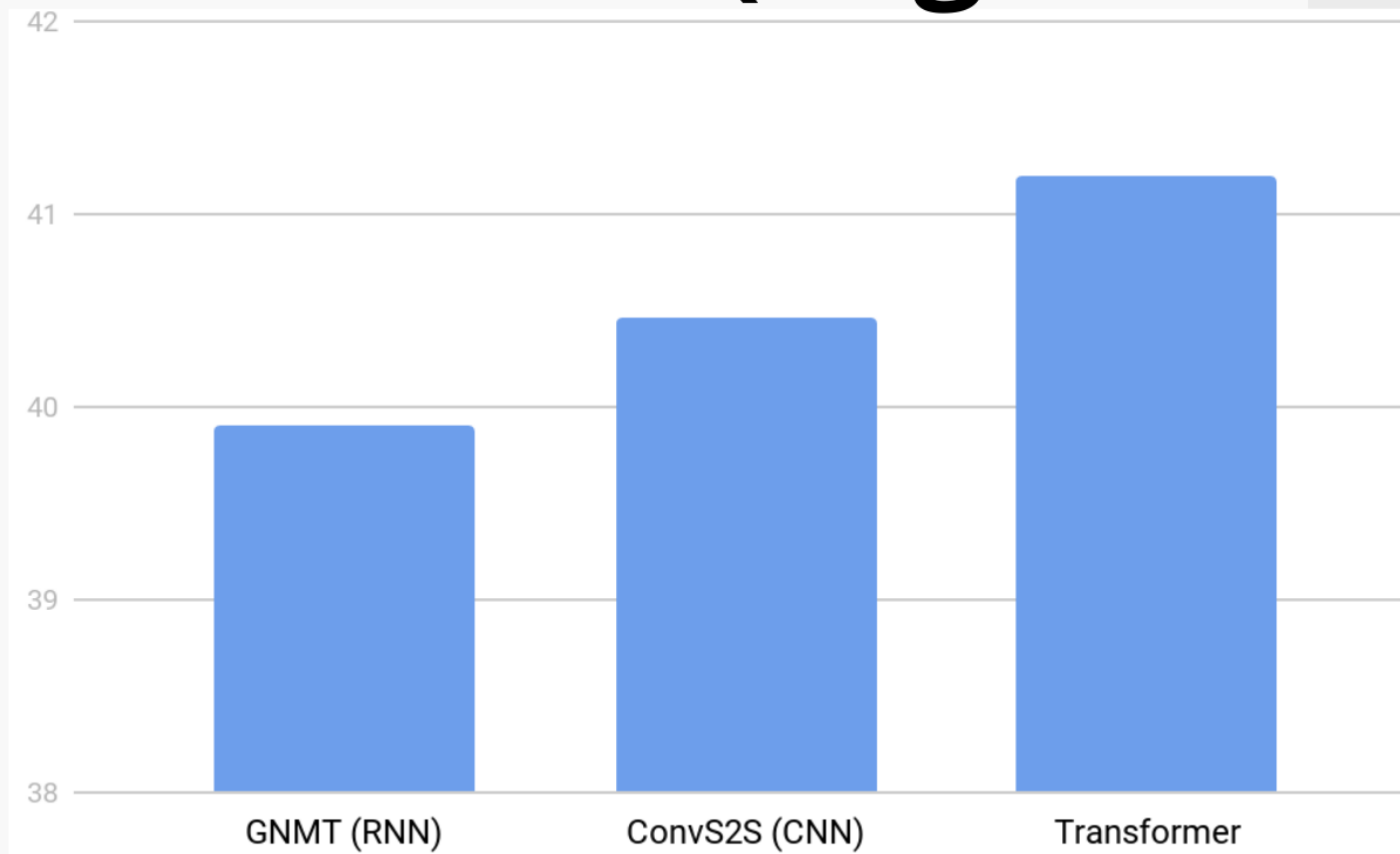
## Идея

- RNN + attention лучше справляется с длинными последовательностями, несмотря на способности LSTM
- А что если использовать внимание без RNN?

# NMT: BLEU (Eng - German)



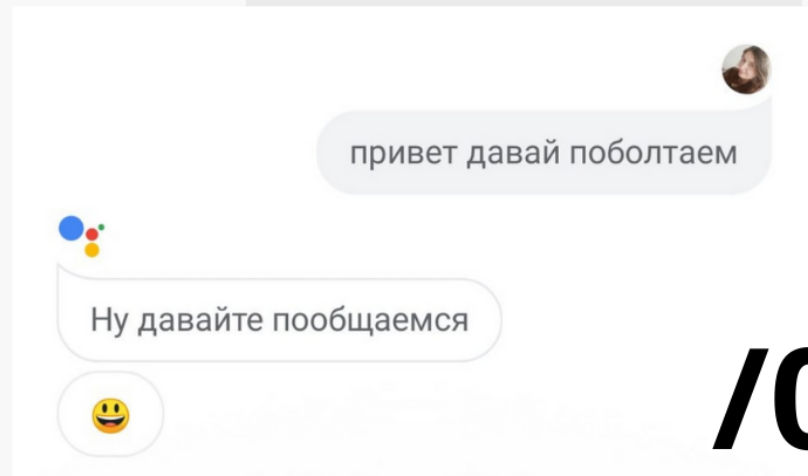
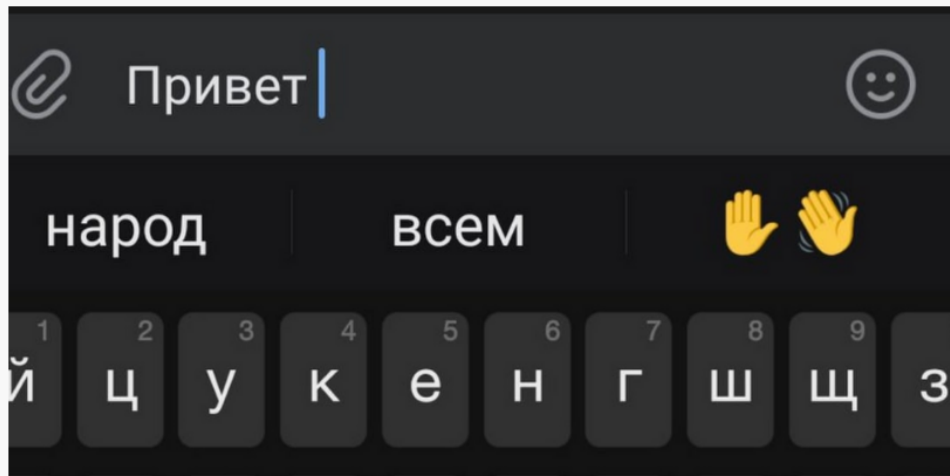
# NMT: BLEU (Eng - French)



# Задачи

Создание систем, которые требуют "*понимания языка*":

- построение языковых моделей;
- машинный перевод;
- построение вопросно-ответных систем



# Архитектура

- Нерекуррентная глубокая модель
- Состоит из последовательности блоков трансформера, в основе которых лежит механизм внимания
- Функция потерь - кросс-энтропия
- Функция активации - softmax
  - Стандартные параметры для моделей NLP
  - *Обработка языка всегда подразумевает многомерную обработку, т.к. лингвистические данные включают обычно одновременно множество переменных*



# Блок

## Слой 1: Multihead attention

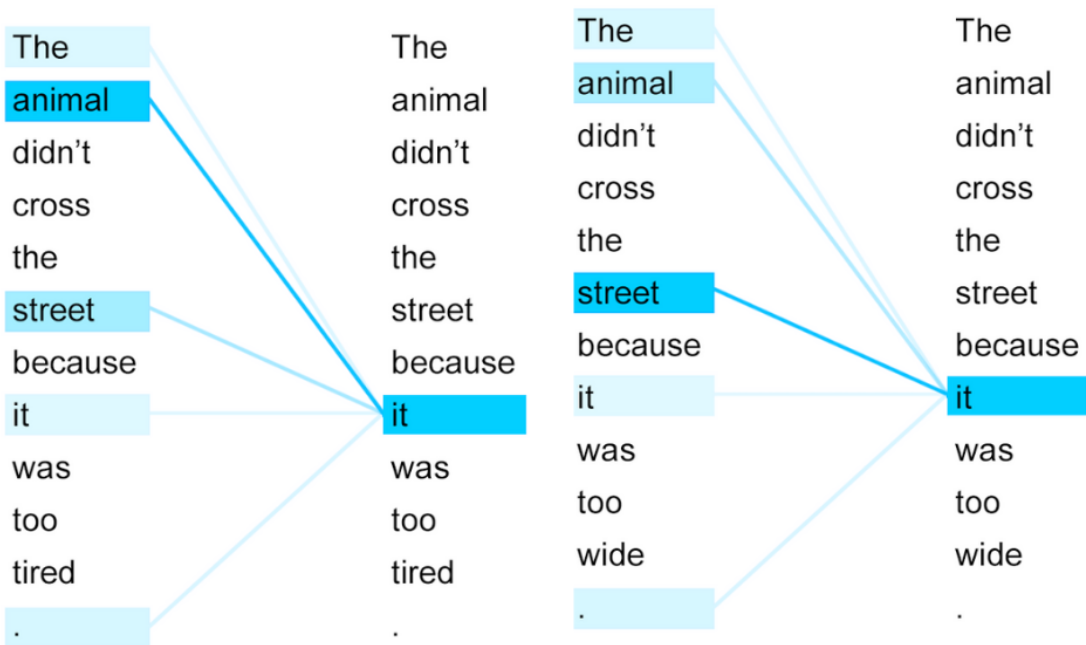
- Обычное внимание учитывает только 1 вариант связи слова с другими словами
- А "необычное" позволяет модели собирать информацию о разных возможных представлениях.

## Слой 2: двухслойная НС прямого распространения

**Новые возможности:** трансформер хорошо справляется с синтаксической разметкой (выделением составляющих) - одной из самых спорных задач NLP.

The animal didn't cross the street because it was too tired.  
L'animal n'a pas traversé la rue parce qu'il était trop fatigué.

The animal didn't cross the street because it was too wide.  
L'animal n'a pas traversé la rue parce qu'elle était trop large.



/10

# Byte pair encoding

- Алгоритм компрессии
- В основе - правила
- Вариант решения проблемы OOV-слов
- Единица анализа - символ (не токен)
- Длина последовательности сжимается - словарь расширяется

Схема: **a**ba cb**a**b -> **X**ba cb**X**b -> **X**ba cb**X**b -> **Y**a cb**Y**

*Самая распространенная пара следующих друг за другом байтов данных (здесь: символов) заменяется новым байтом*

## Список источников:

1. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 6000–6010.
2. Jakob Uszkoreit. 2017. *Transformer: A Novel Neural Network Architecture for Language Understanding*. Google AI Blog. URL: <https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>

**Подготовила Виктория Фирсанова,**  
 СПбГУ, кафедра математической лингвистики  
 E-mail: vifirsanova@gmail.com