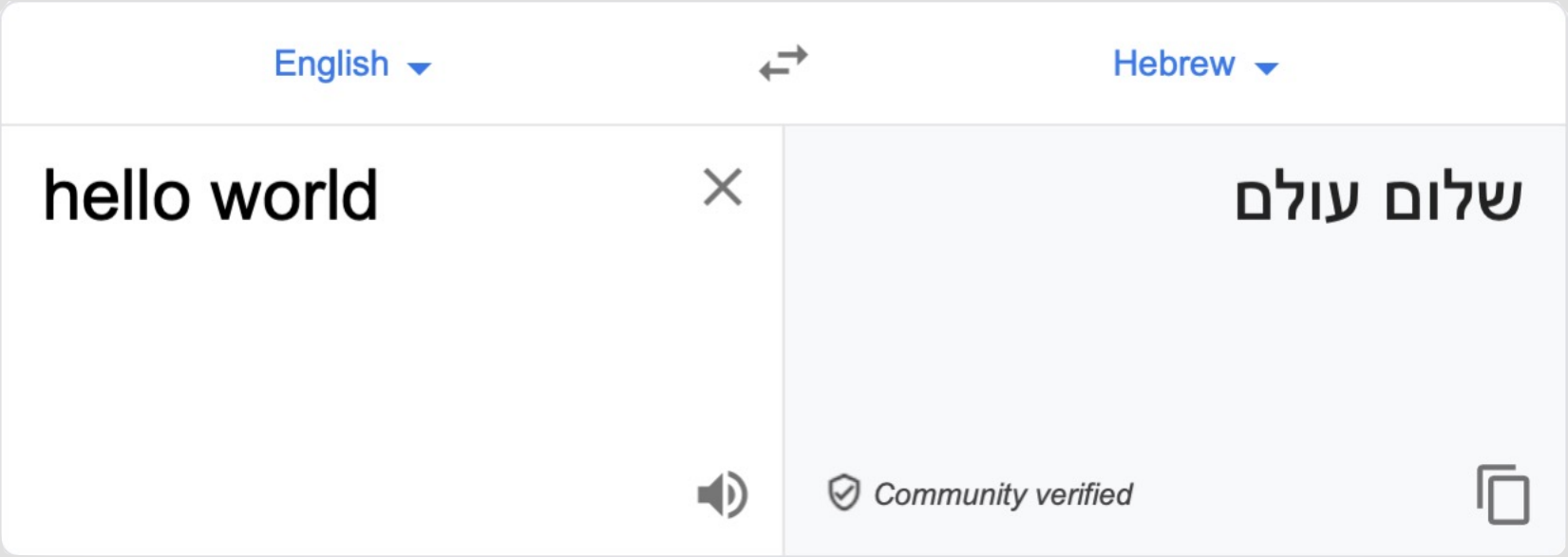
Машинный перевод

- **Модель кодировщик-декодировщик**
- Механизм внимания
- Модель Трансформер
- Метрики качества в машинном переводе
- Улучшение качества машинного перевода
- Машинный перевод без учителя

Модель кодировщик-декодировщик

Модель класса seq2seq

Так работает машинный перевод



Машинный перевод (machine translation)

- Массовая и востребованная задача

Машинный перевод (machine translation)

- Массовая и востребованная задача
- Одна из первых языковых технологий
- Джорджтаунский эксперимент был проведен в 1954 г.

Машинный перевод (machine translation)

- Массовая и востребованная задача
- Одна из первых языковых технологий
- Джорджтаунский эксперимент был проведен в 1954 г.
- Статистические модели (IBM Model) превалировали до середины 2010-х

Машинный перевод (machine translation)

- Массовая и востребованная задача
- Одна из первых языковых технологий
- Джорджтаунский эксперимент был проведен в 1954 г.
- Статистические модели (IBM Model) превалировали до середины 2010-х
- С 2016 нейросетевой машинный перевод (NMT) – индустриальный стандарт

Машинный перевод (machine translation)

- Массовая и востребованная задача
- Одна из первых языковых технологий
- Джорджтаунский эксперимент был проведен в 1954 г.
- Статистические модели (IBM Model) превалировали до середины 2010-х
- С 2016 нейросетевой машинный перевод (NMT) – индустриальный стандарт
- Сейчас машинный перевод хорошо работает для простых или формальных текстов
- Перевод художественной литературы – все еще большая сложность

Модель кодировщик-декодировщик

- Задача: перевести текст с одного языка на другой

Модель кодировщик-декодировщик

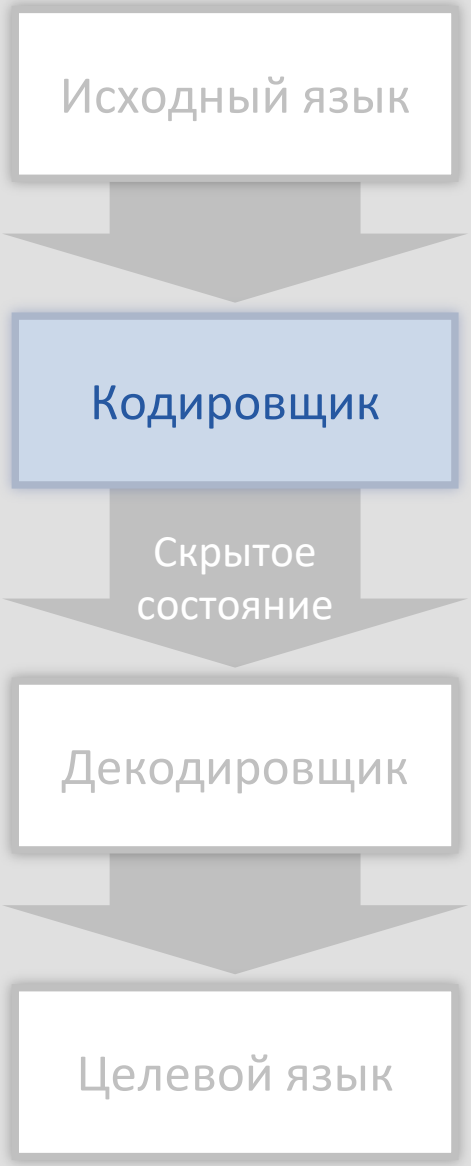
- Задача: перевести текст с одного языка на другой

Модель кодировщик-декодировщик

- Задача: перевести текст с одного языка на другой

Модель кодировщик-декодировщик

- Задача: перевести текст с одного языка на другой

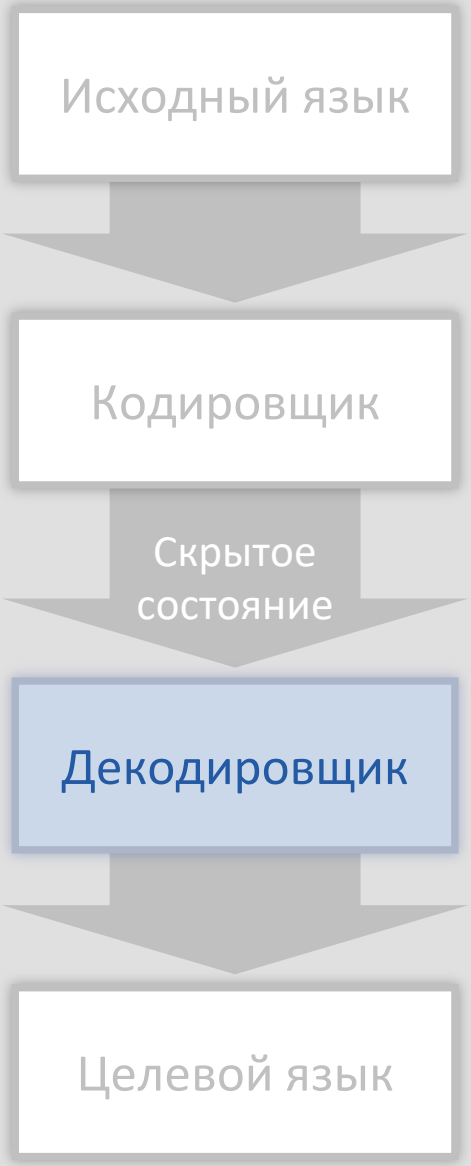


Модель кодировщик-декодировщик

- Задача: перевести текст с одного языка на другой

Модель кодировщик-декодировщик

- Задача: перевести текст с одного языка на другой



Модель кодировщик-декодировщик

- Задача: перевести текст с одного языка на другой

Модель кодировщик-декодировщик

- Задача: перевести текст с одного языка на другой

Постановка задачи

- Предложение на исходном языке:
  $\mathbf{x}_{source} = (x_1, \dots, x_n), x_i \in V_{source}$

Постановка задачи

- Предложение на исходном языке:
  $\mathbf{x}_{source} = (x_1, \dots, x_n), x_i \in V_{source}$
- Предложение на целевом языке:
  $\mathbf{y}_{target} = (y_1, \dots, y_m), y_i \in V_{target}$

Постановка задачи

- Предложение на исходном языке: $\mathbf{x}_{source} = (x_1, \dots, x_n), x_i \in V_{source}$
- Предложение на целевом языке: $\mathbf{y}_{target} = (y_1, \dots, y_m), y_i \in V_{target}$
- $\mathbf{C}$ – параллельный корпус, состоит из пар предложений $(\mathbf{x}, \mathbf{y})$

Постановка задачи

- Предложение на исходном языке: $\mathbf{x}_{source} = (x_1, \dots, x_n), x_i \in V_{source}$
- Предложение на целевом языке: $\mathbf{y}_{target} = (y_1, \dots, y_m), y_i \in V_{target}$
- $\mathbf{C}$ – параллельный корпус, состоит из пар предложений $(\mathbf{x}, \mathbf{y})$
- Максимизируем правдоподобие $p(\mathbf{y}|\mathbf{x})$
- Целевая функция:

$$\mathcal{L}_\theta = \sum_{\mathbf{x},\mathbf{y} \in \mathbf{C}} \log p(\mathbf{y}|\mathbf{x}; \theta)$$

Постановка задачи

- Вероятность целевого предложения при условии исходного предложения:

$$p(\mathbf{y}|\mathbf{x};\theta) = \prod_{j=1}^{m} p(\mathbf{y}_j|\mathbf{y}_{<j},\mathbf{x};\theta)$$

$\mathbf{y}_j$ – текущее слово,

$\mathbf{y}_{<j}$ – предыдущие слова

Постановка задачи

- Вероятность целевого предложения при условии исходного предложения:

$$p(\mathbf{y}|\mathbf{x};\theta) = \prod_{j=1}^{m} p(\mathbf{y}_j|\mathbf{y}_{<j},\mathbf{x};\theta)$$

$\mathbf{y}_j$ – текущее слово,

$\mathbf{y}_{<j}$ – предыдущие слова

Языковая модель, обусловленная на $\mathbf{x}$

Нейросетевой машинный перевод

- Подготовить векторные представления

$$f_i^s = E_{source} x_i$$

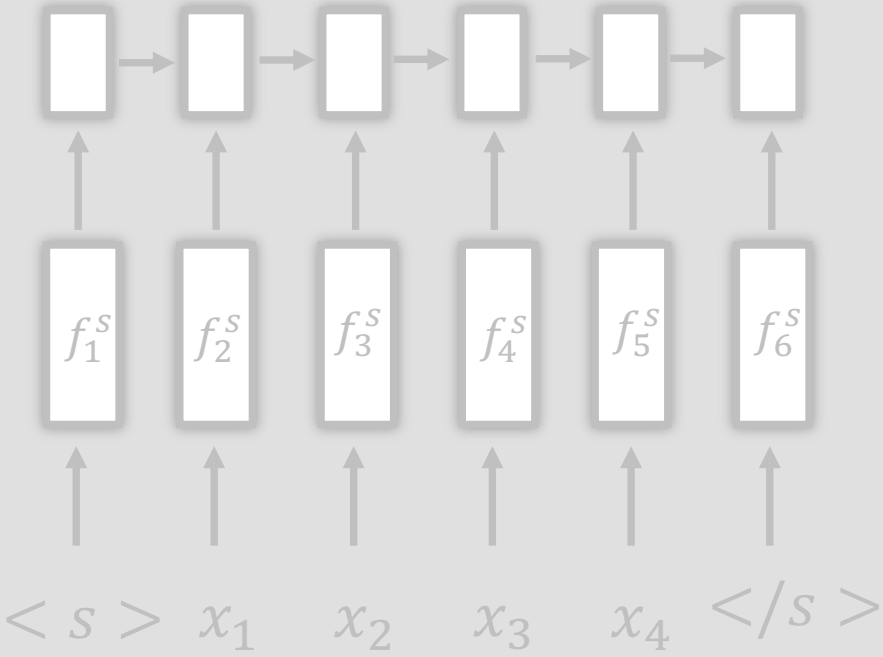
Нейросетевой машинный перевод

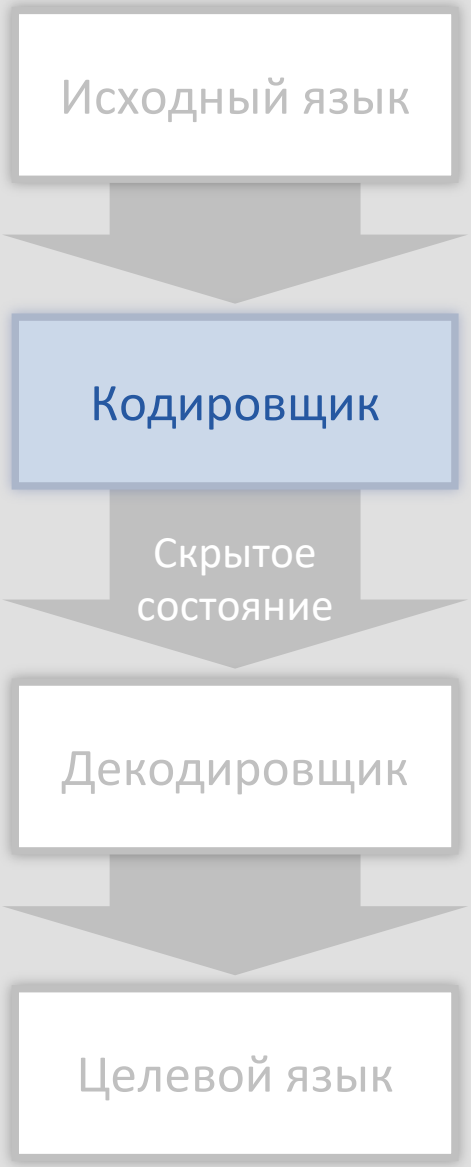
- Подготовить векторные представления

$$f_i^s = E_{source} x_i$$

Нейросетевой машинный перевод

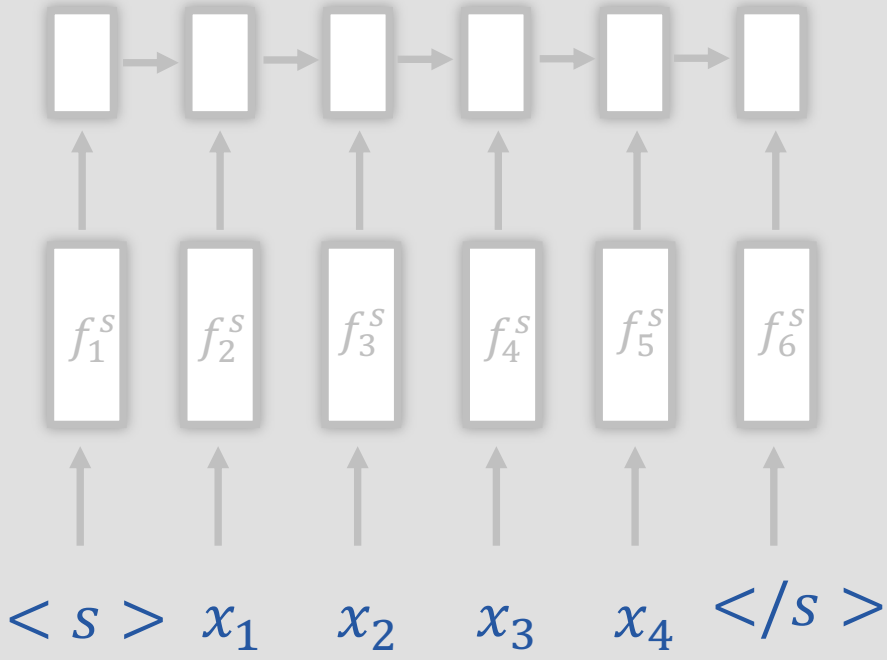
- Архитектура кодировщика

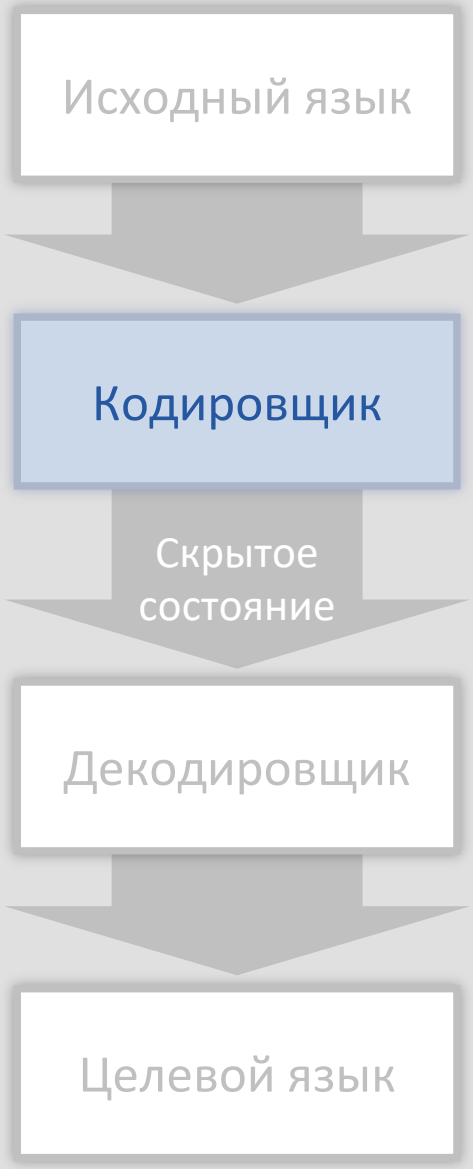




Нейросетевой машинный перевод

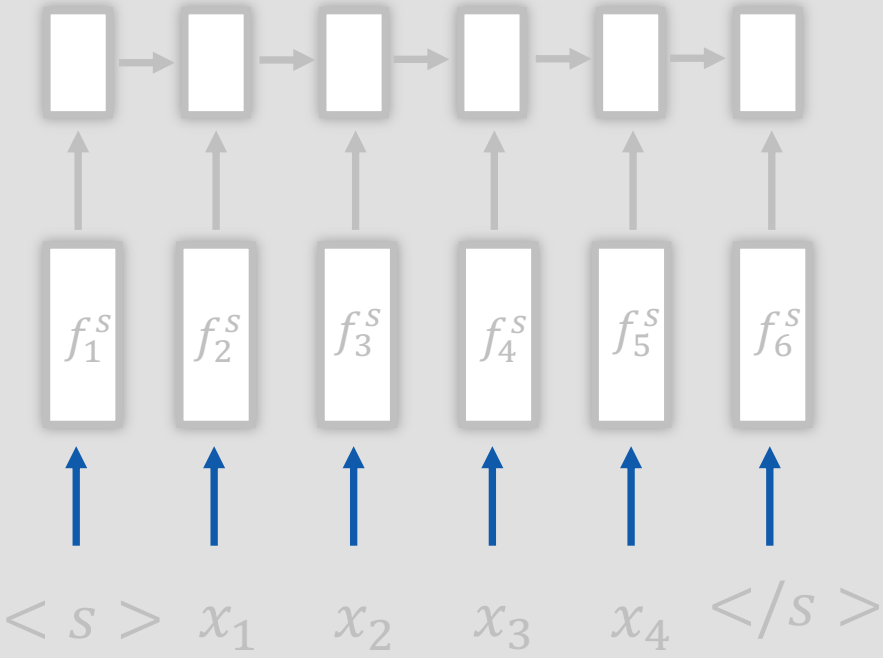
- Архитектура кодировщика

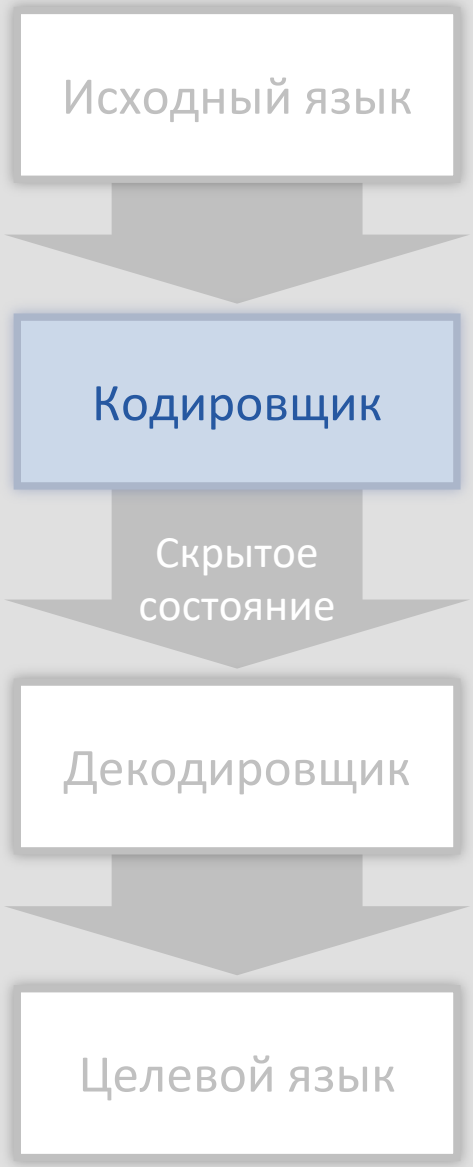




Нейросетевой машинный перевод

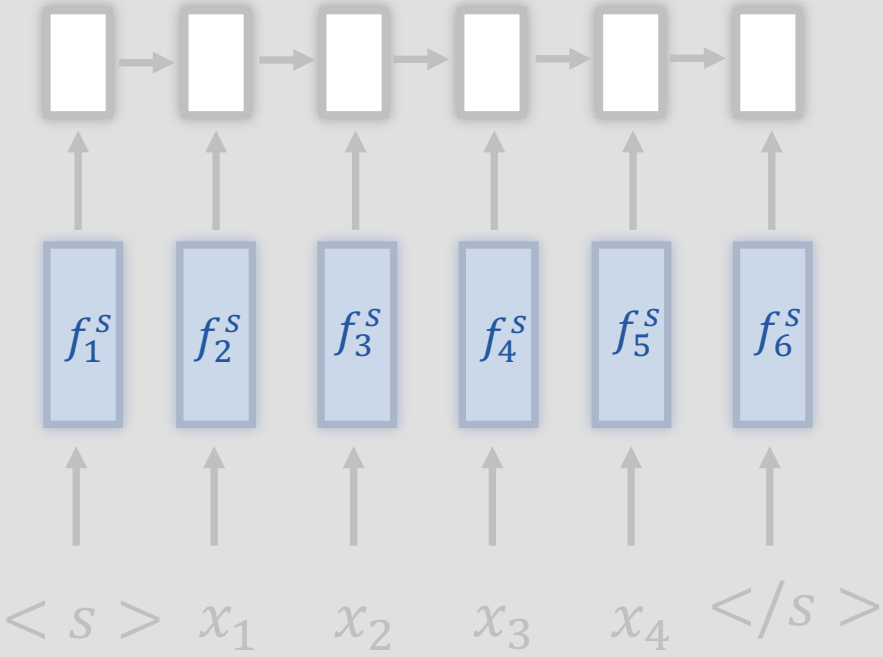
- Архитектура кодировщика

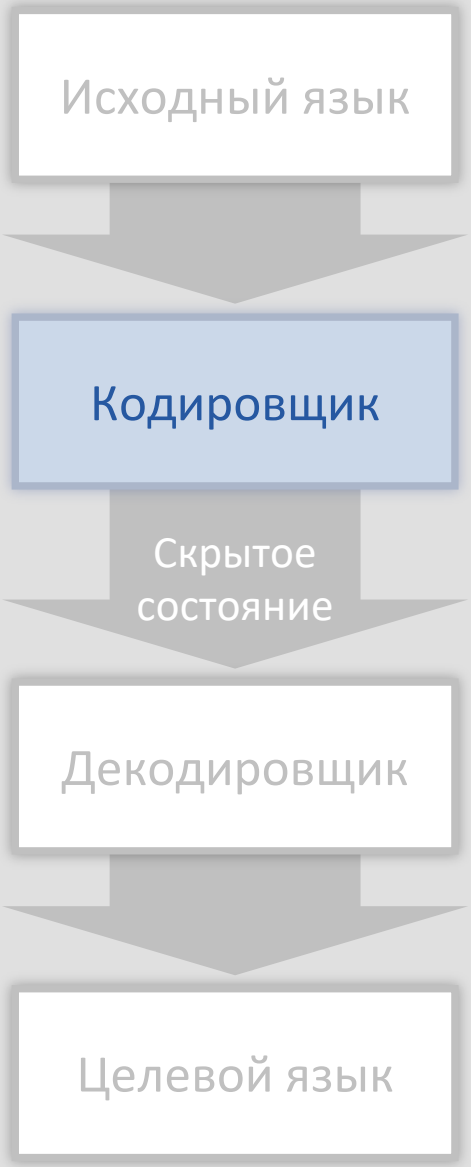




Нейросетевой машинный перевод

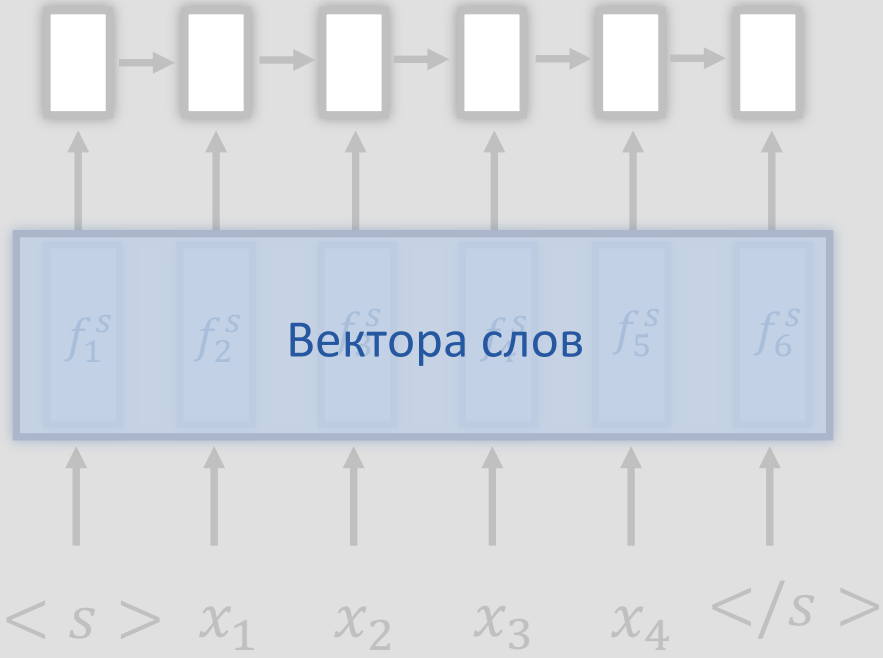
- Архитектура кодировщика

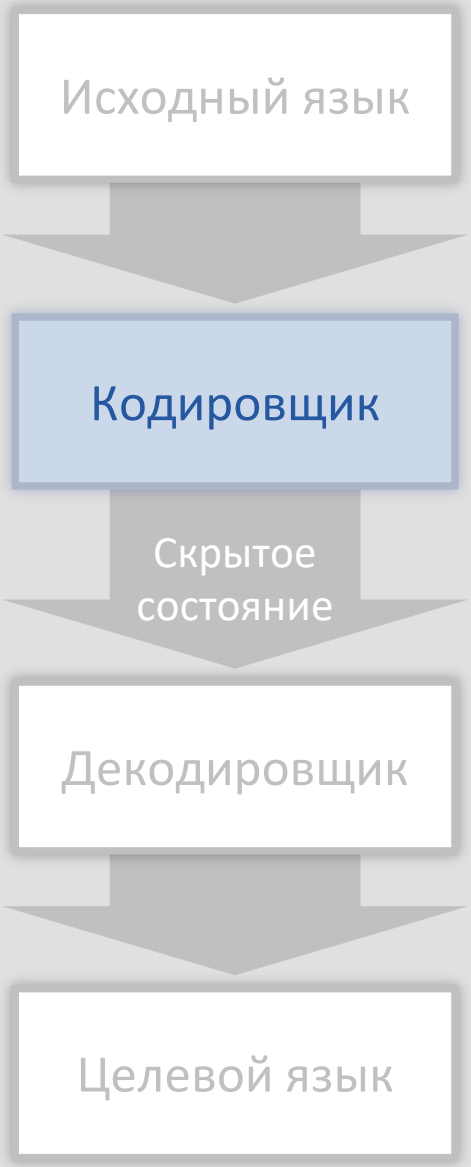




Нейросетевой машинный перевод

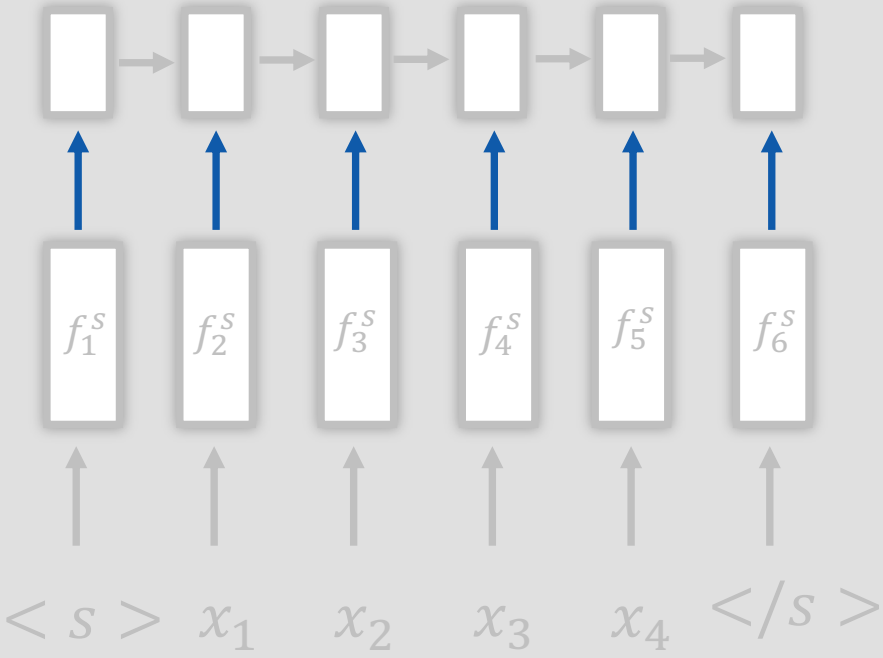
- Архитектура кодировщика

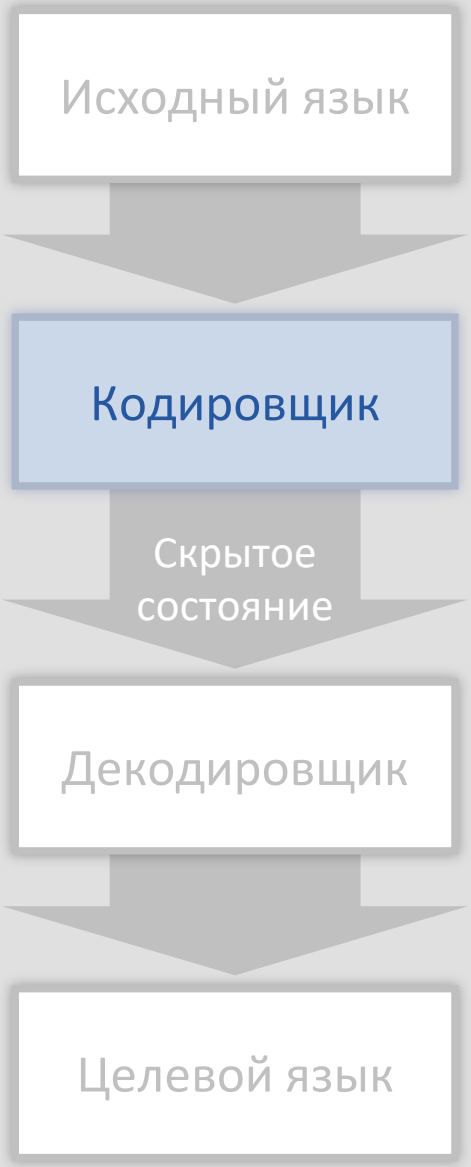




Нейросетевой машинный перевод

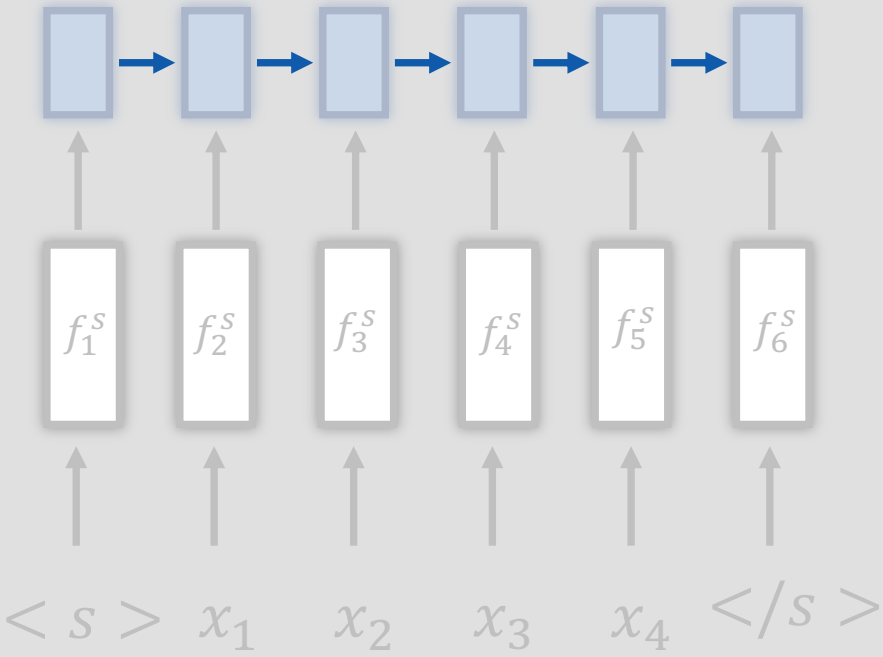
- Архитектура кодировщика

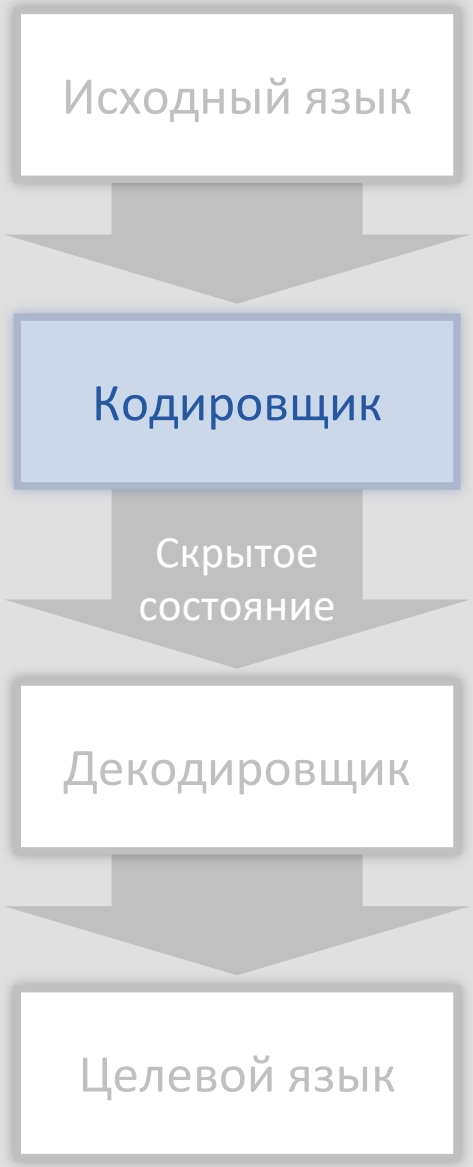




Нейросетевой машинный перевод

- Архитектура кодировщика

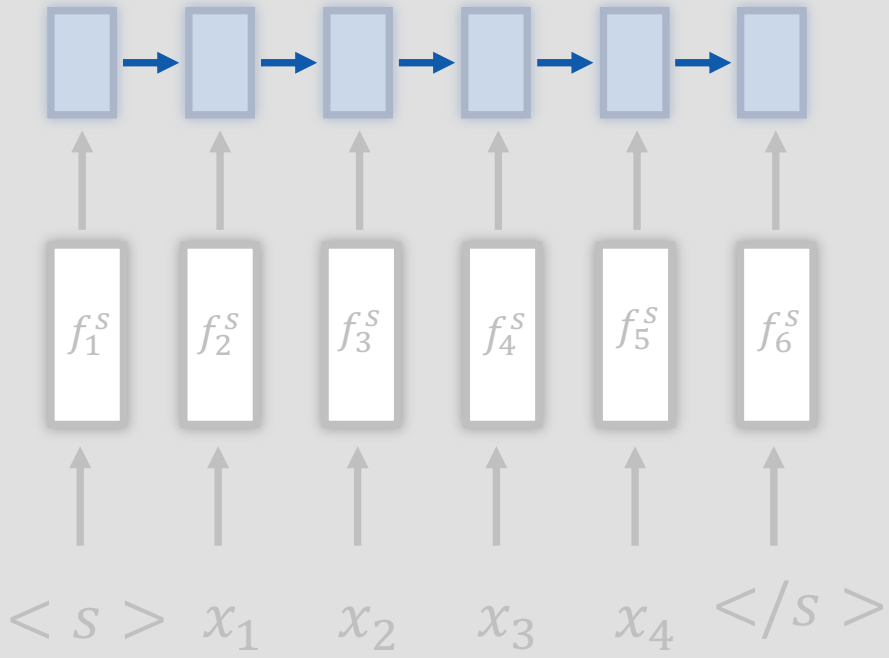




Нейросетевой машинный перевод

- Закодировать исходное предложение

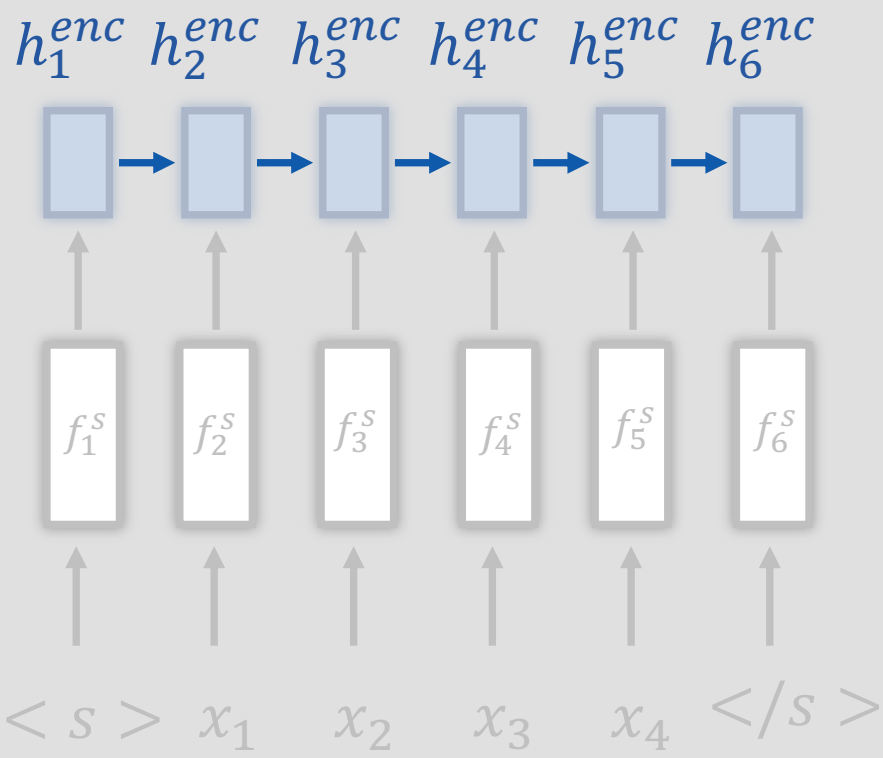
$$h^{encoder} = RNN(\mathbf{x})$$



Нейросетевой машинный перевод

- Закодировать исходное предложение

$$h^{encoder} = RNN(\mathbf{x})$$



Нейросетевой машинный перевод

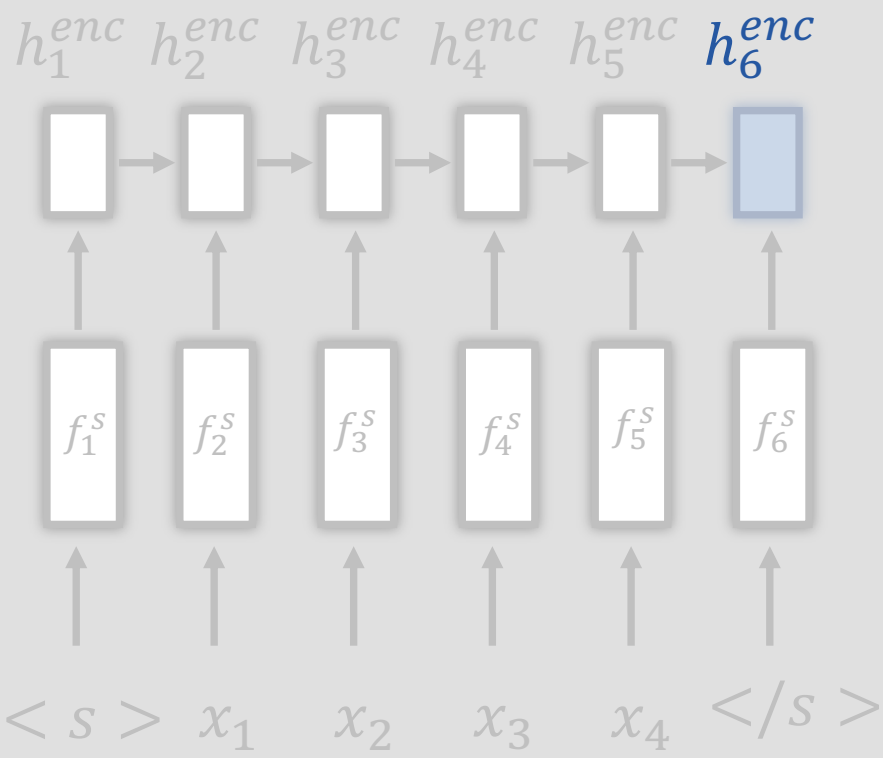
- Закодировать исходное предложение

$$h^{encoder} = RNN(\mathbf{x})$$

Нейросетевой машинный перевод

- Закодировать исходное предложение

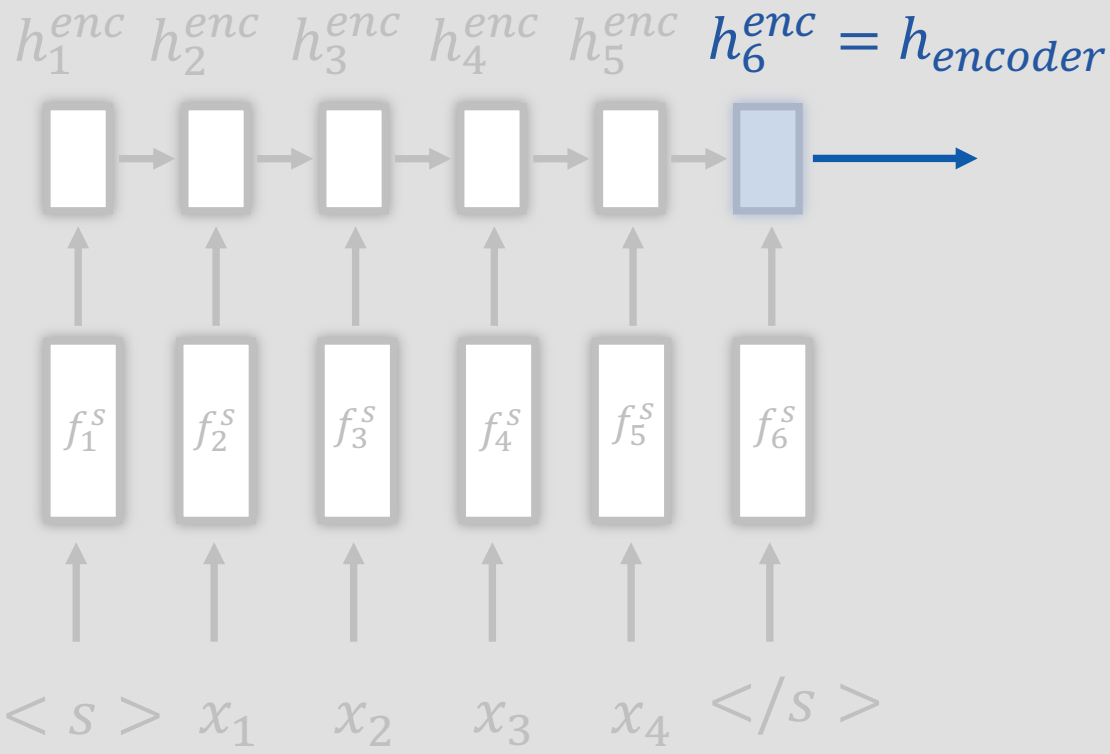
$$h^{encoder} = RNN(\mathbf{x})$$



Нейросетевой машинный перевод

- Закодировать исходное предложение

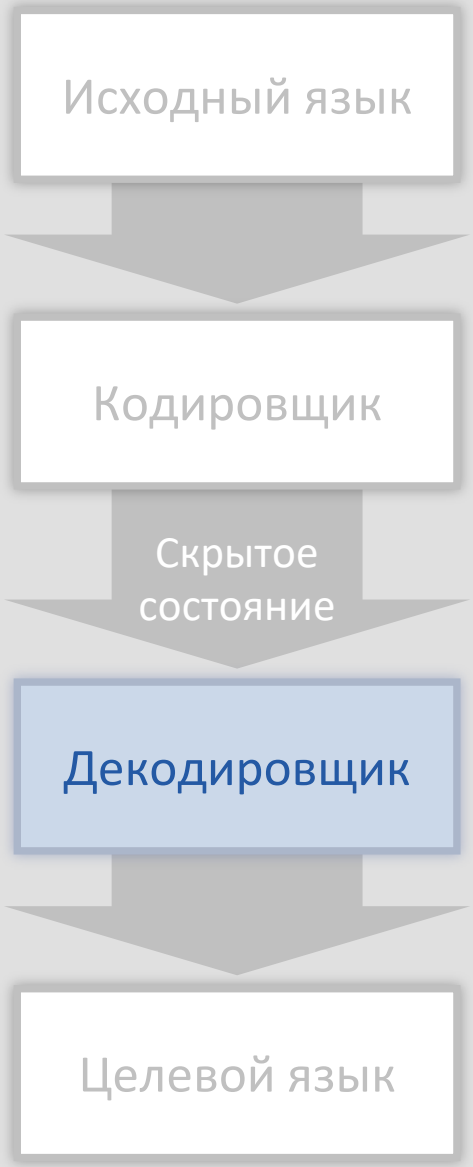
$$h^{encoder} = RNN(\mathbf{x})$$



Архитектура декодировщика

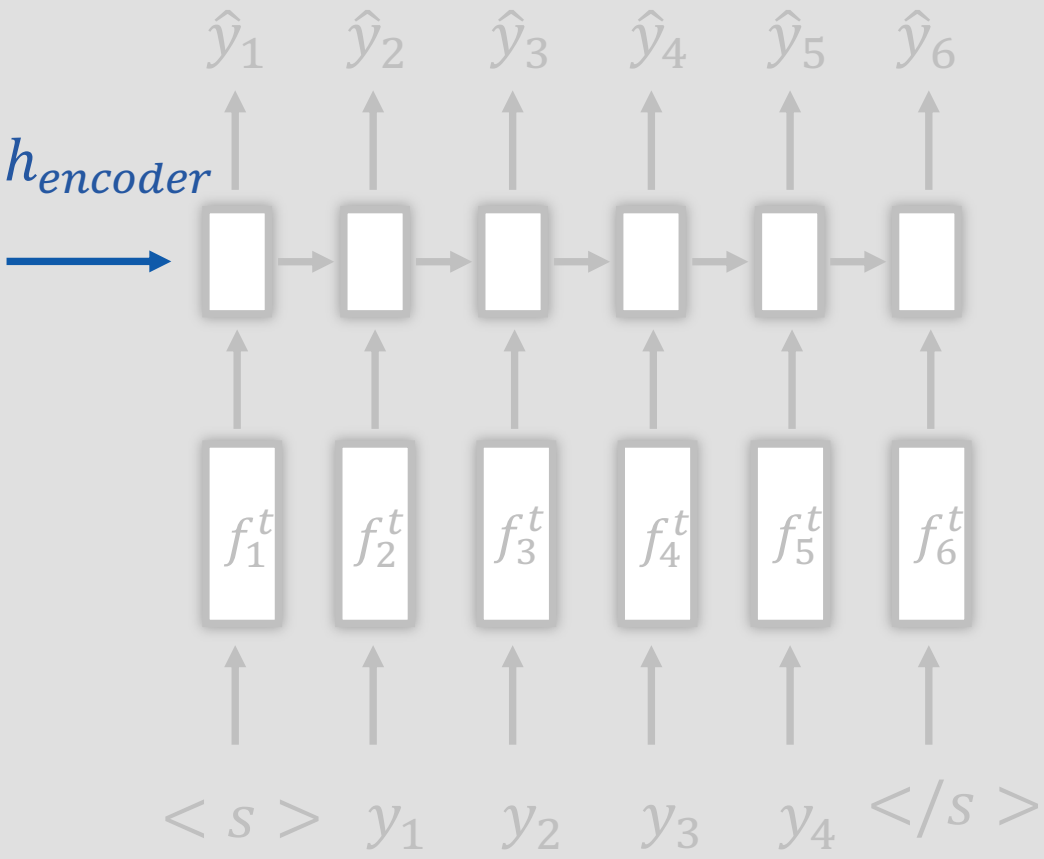
- Раскодировать целевое предложение

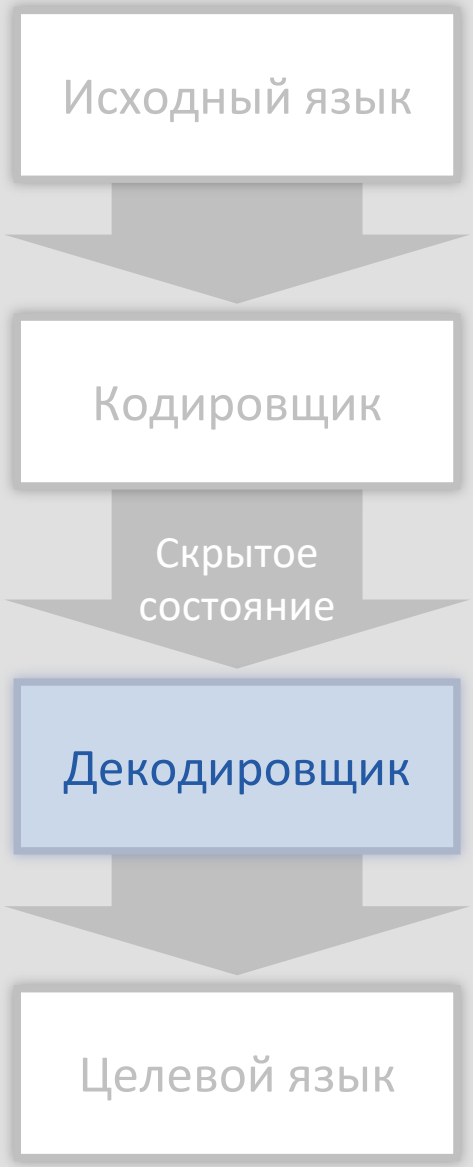
$h_{encoder}$



Архитектура декодировщика

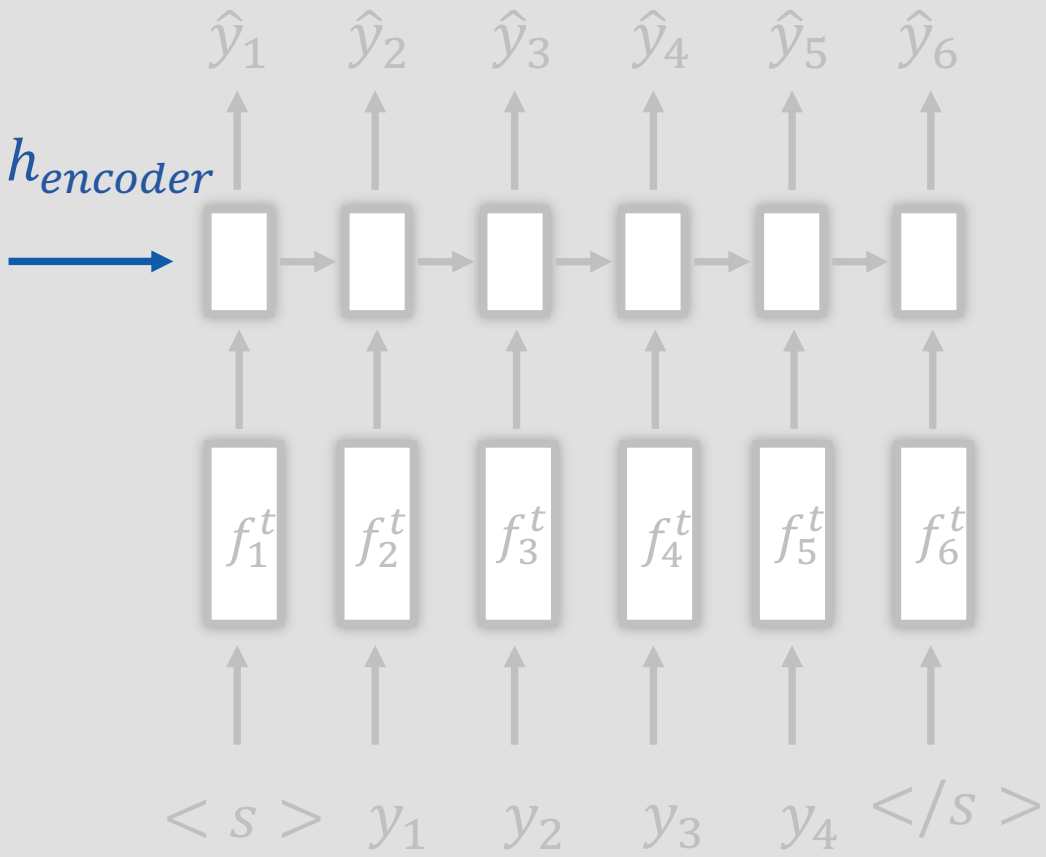
- Раскодировать целевое предложение

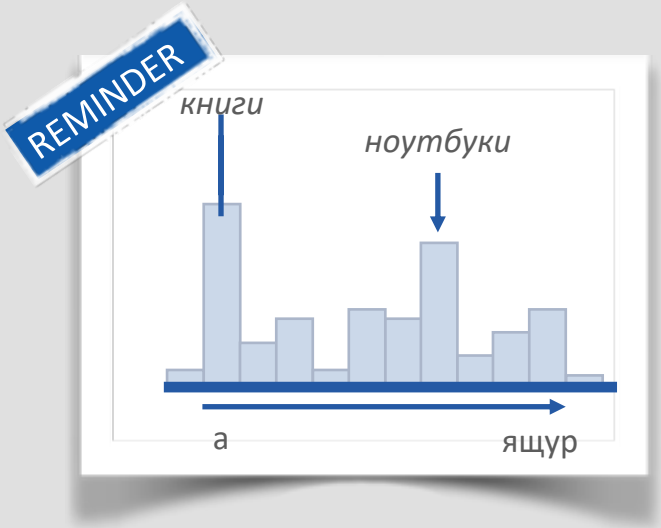




Архитектура декодировщика

- Раскодировать целевое предложение





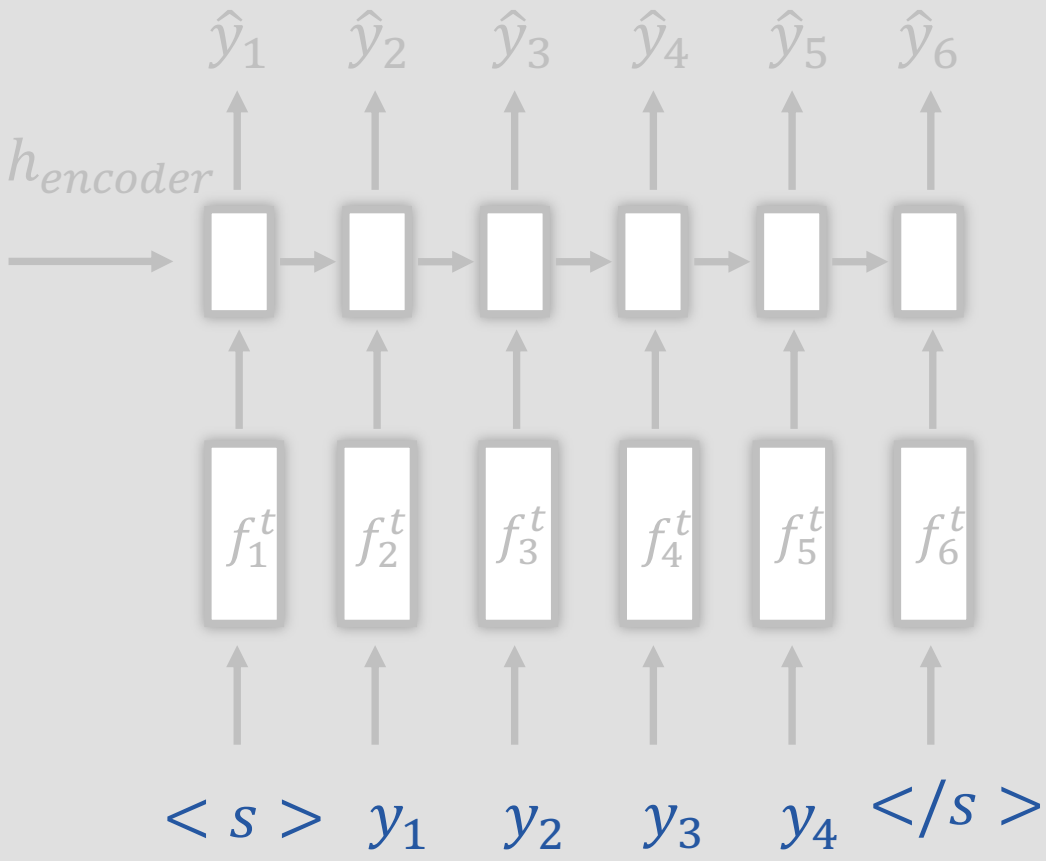
Обучение кодировщика - декодировщика

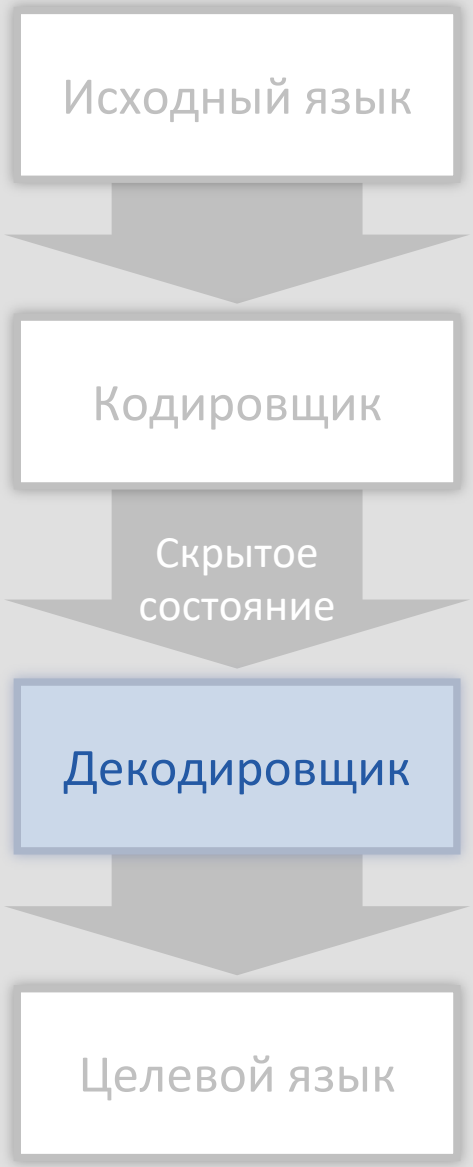
- Функция потерь: отрицательный логарифм правдоподобия

$$L(\mathbf{y}, \hat{\mathbf{y}}) = \sum_{j=1}^{V_{target}} y_j \log \hat{y}_j$$

Архитектура декодировщика

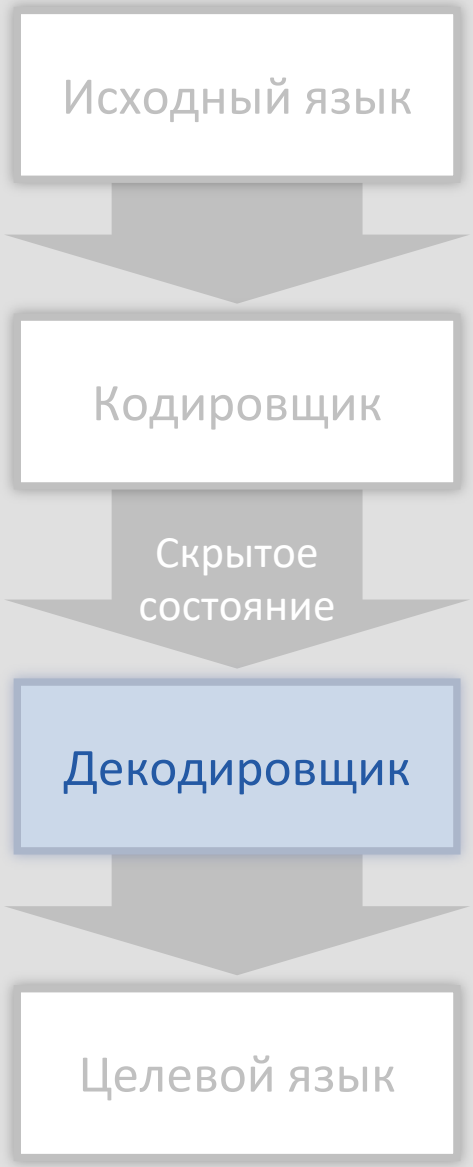
- Раскодировать целевое предложение





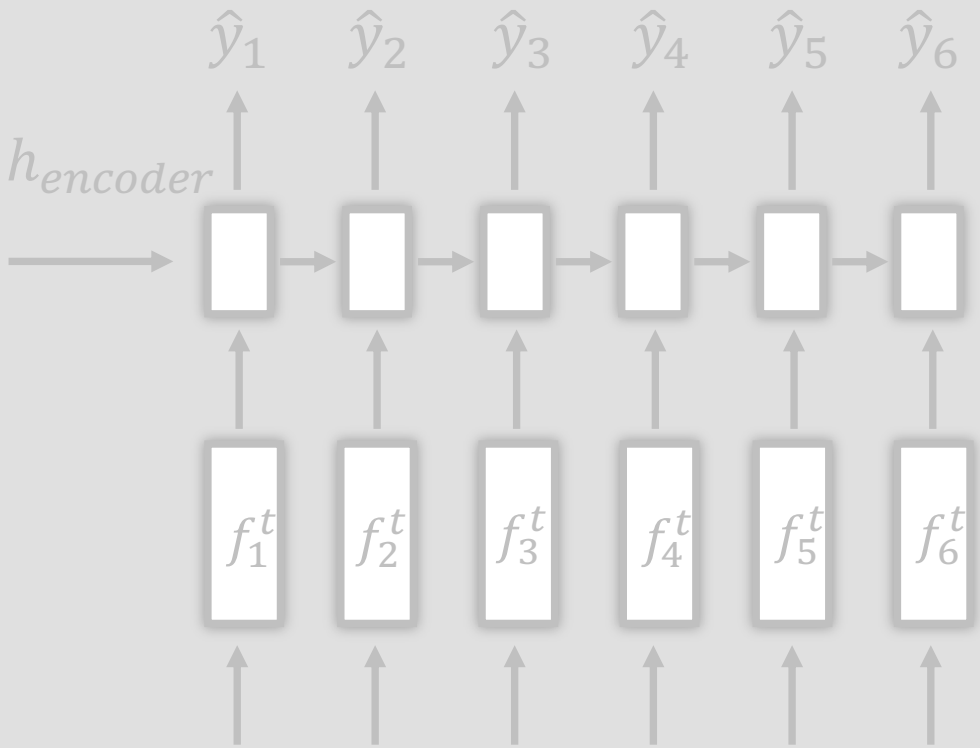
Архитектура декодировщика

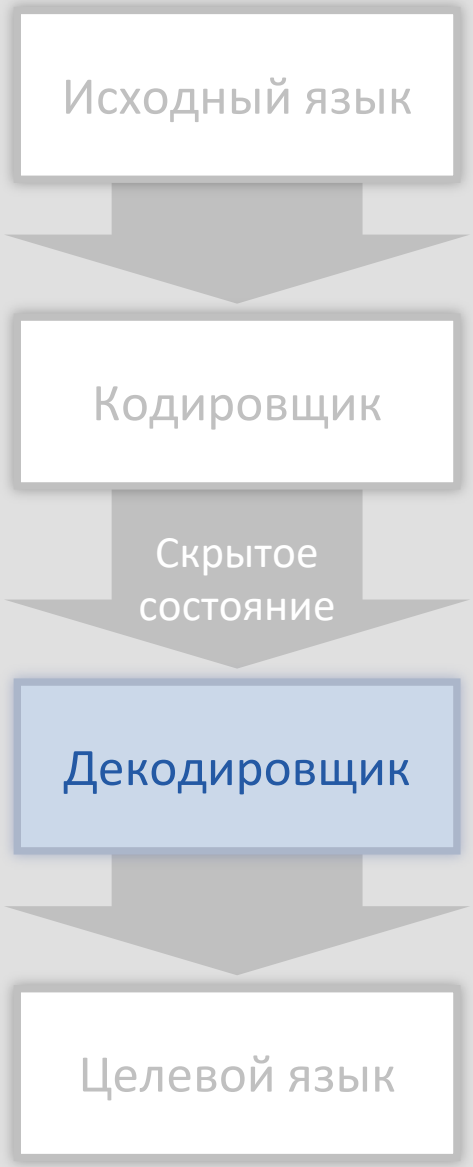
- Раскодировать целевое предложение



Архитектура декодировщика

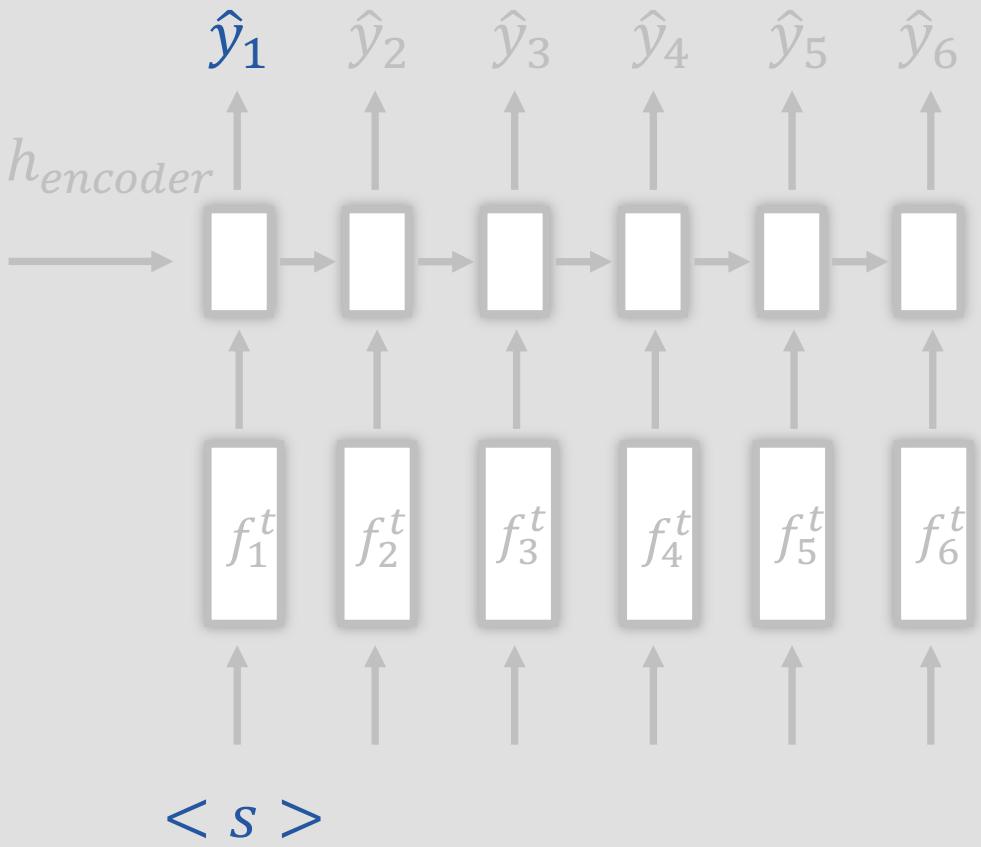
- Раскодировать целевое предложение

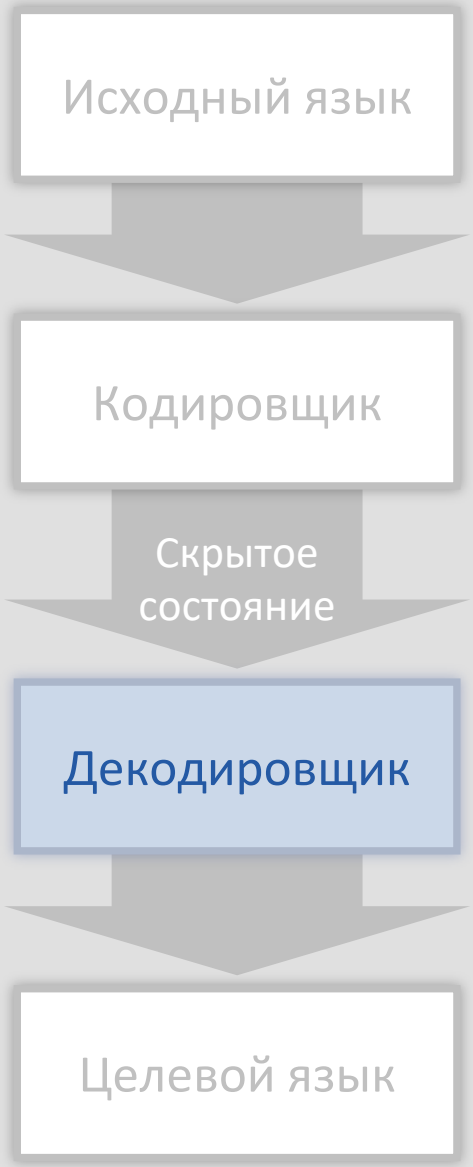




Архитектура декодировщика

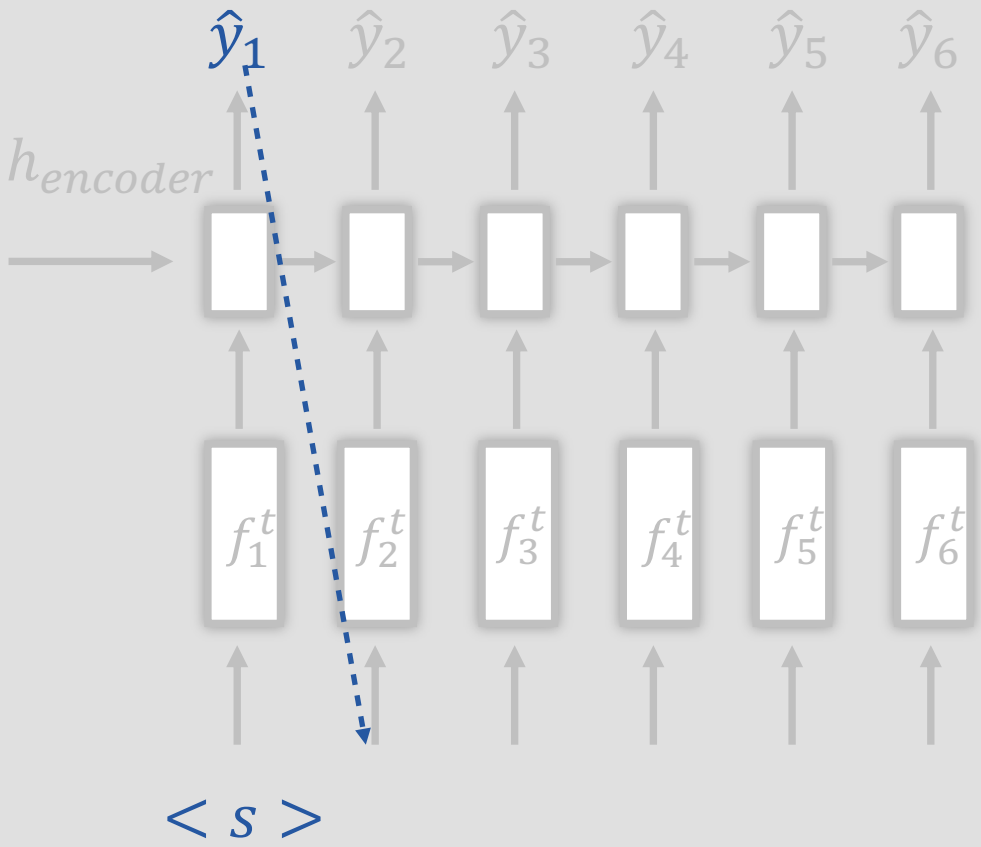
- Раскодировать целевое предложение

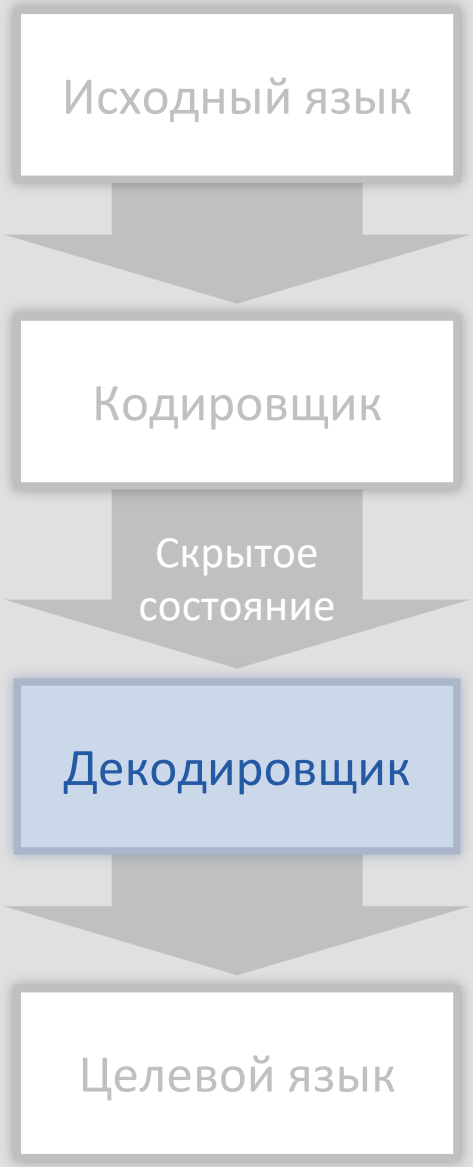




Архитектура декодировщика

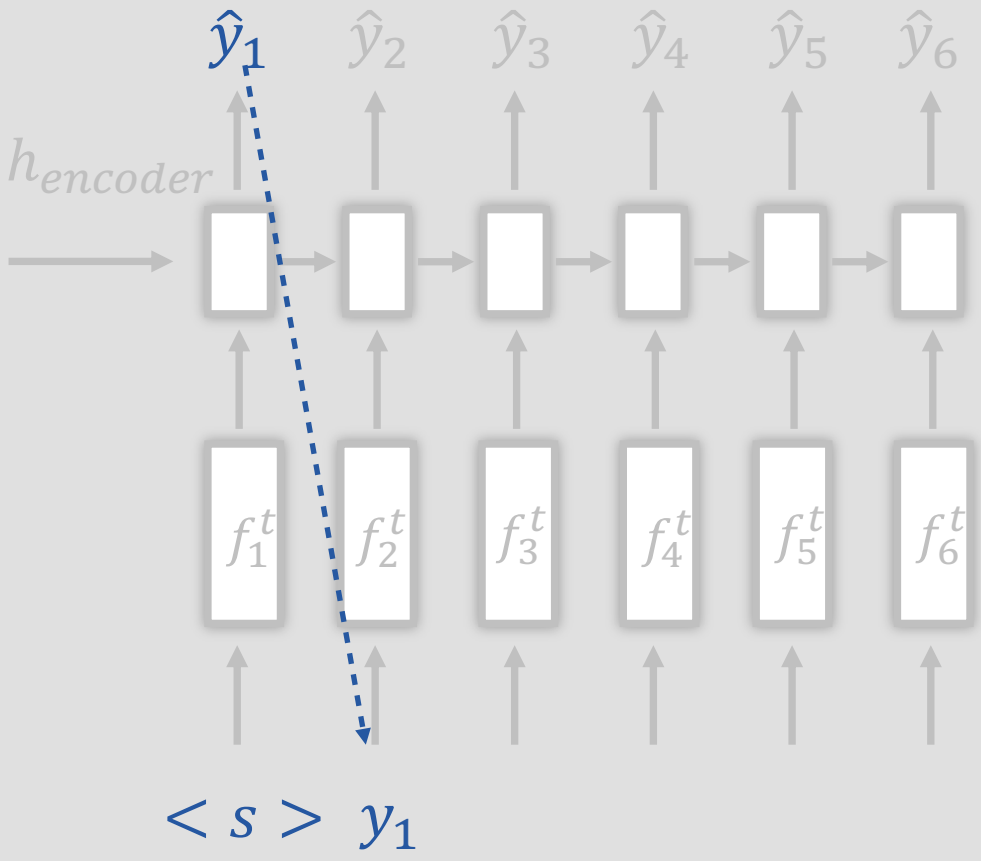
- Раскодировать целевое предложение

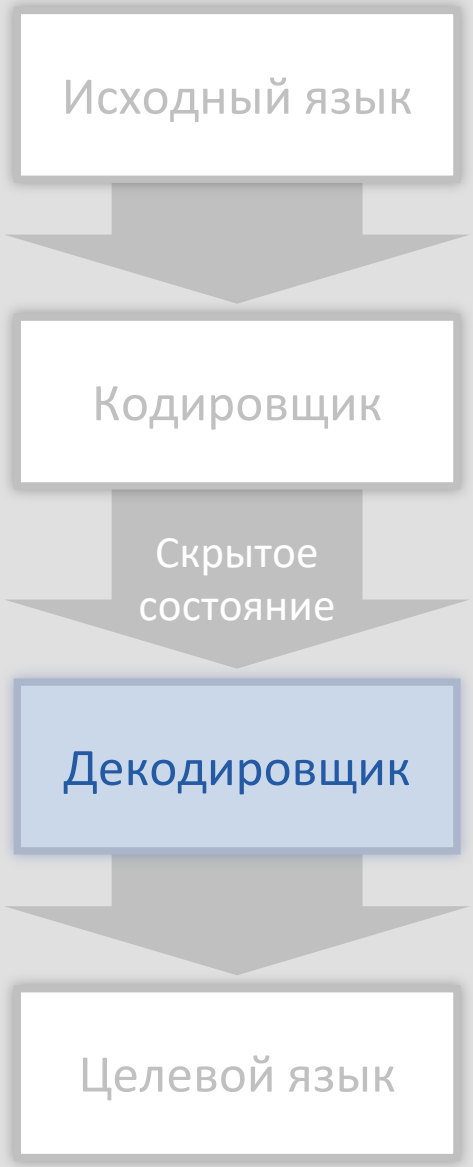




Архитектура декодировщика

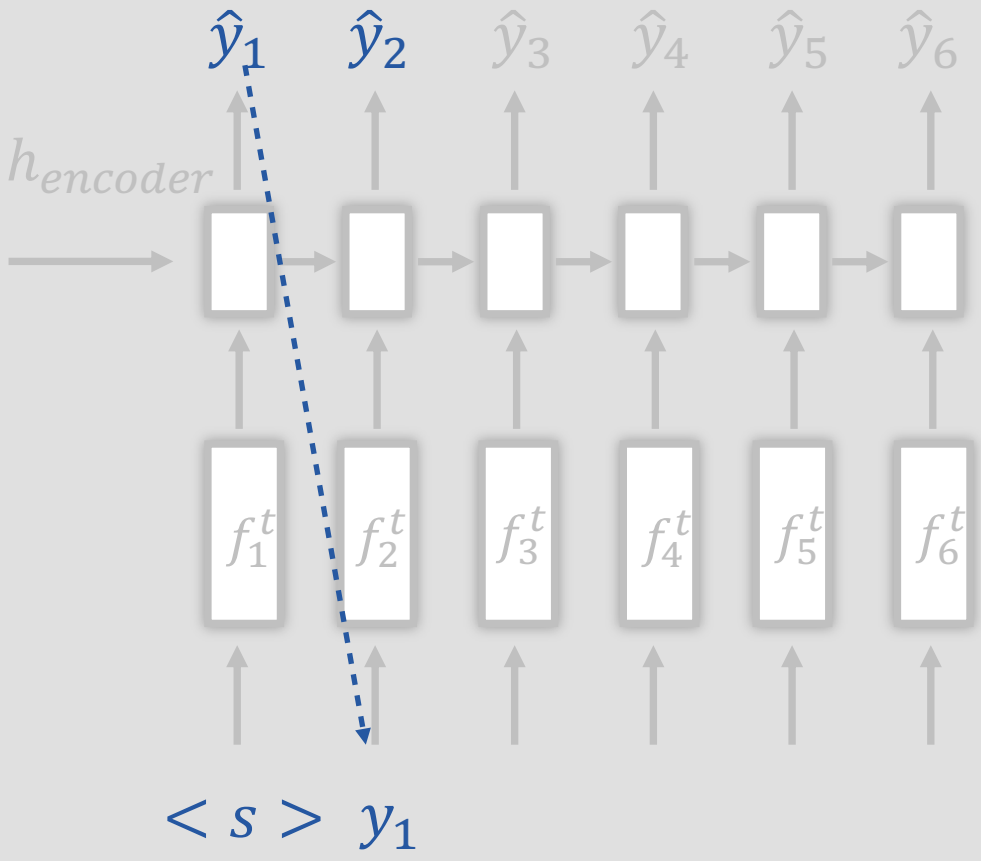
- Раскодировать целевое предложение

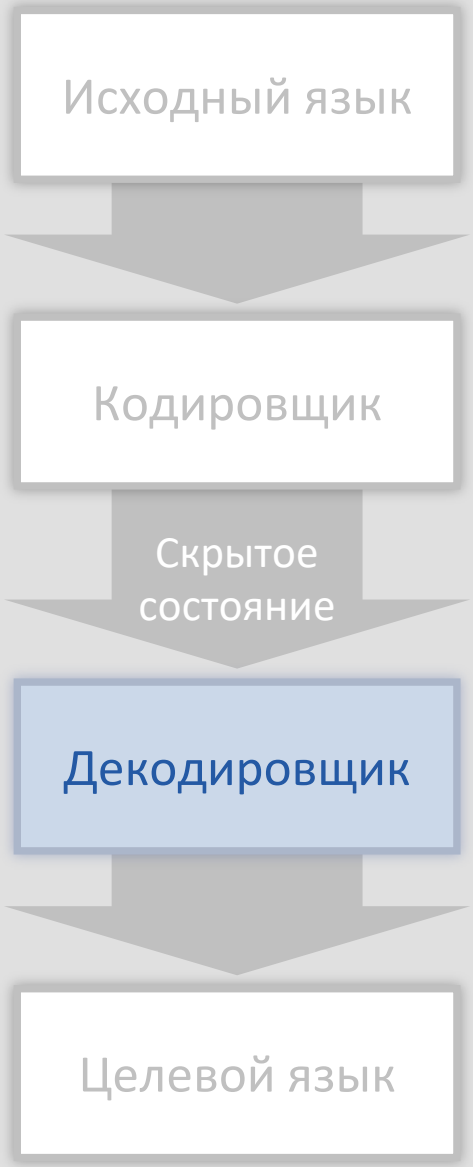




Архитектура декодировщика

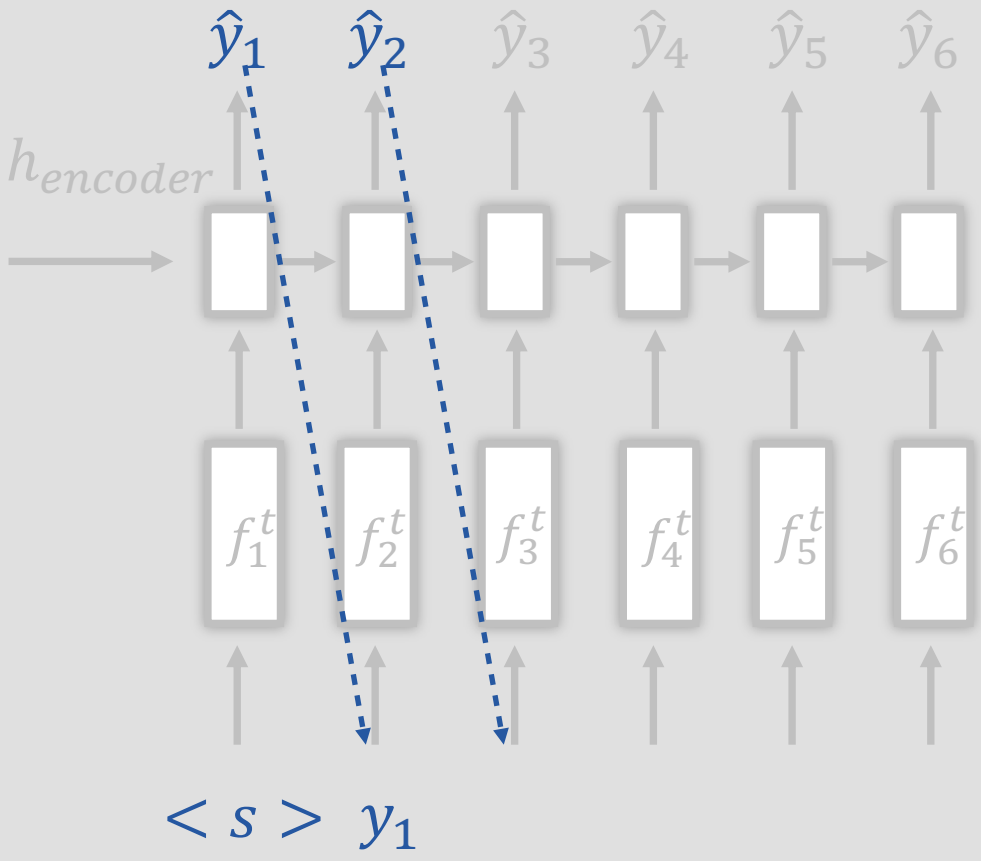
- Раскодировать целевое предложение

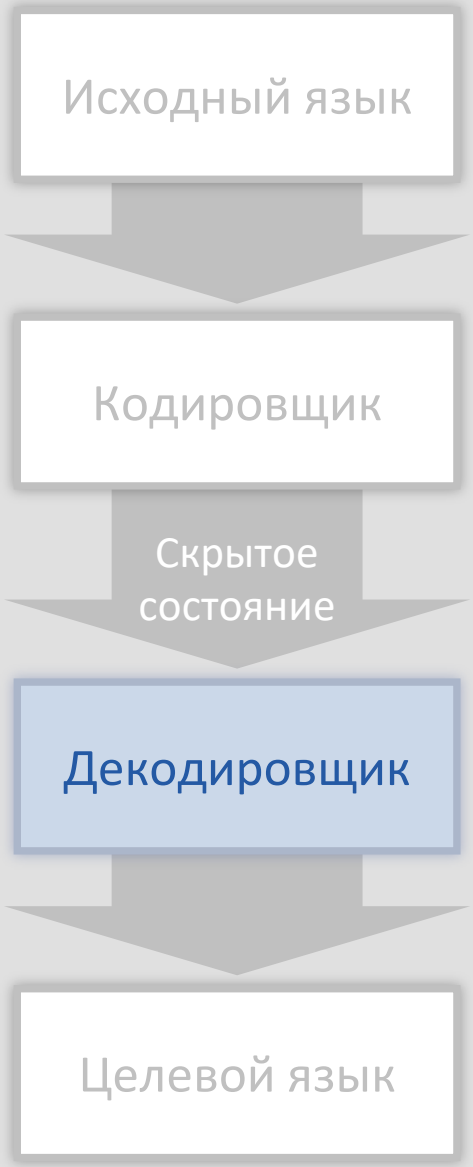




Архитектура декодировщика

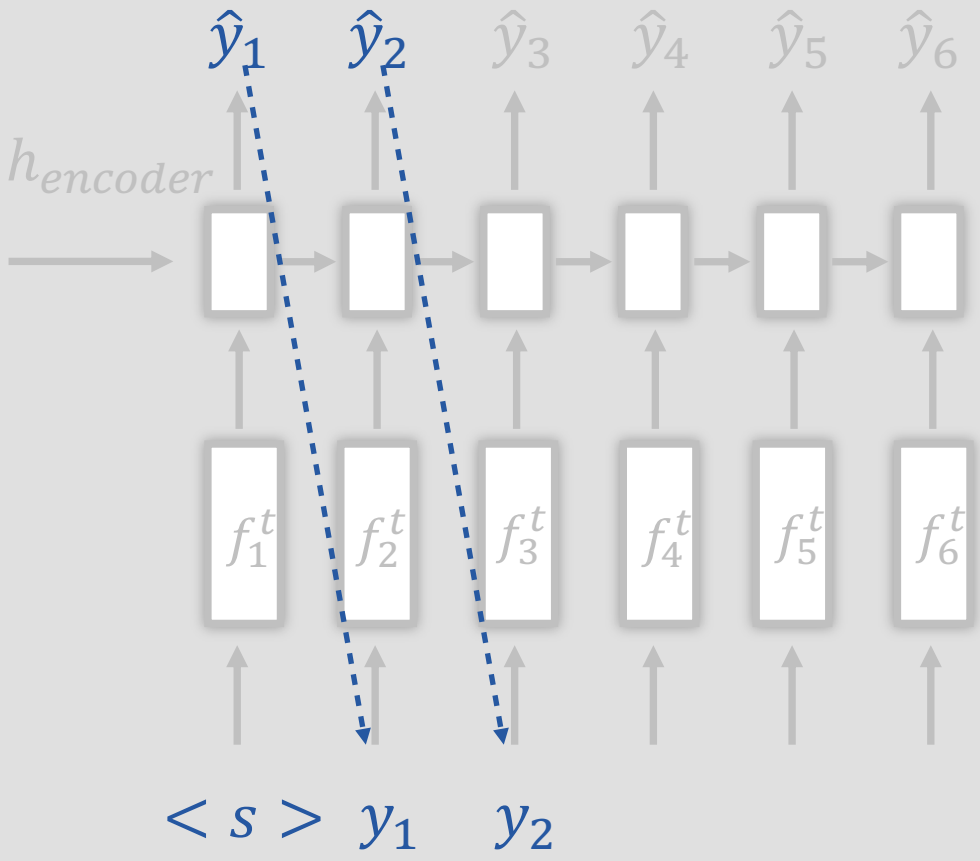
- Раскодировать целевое предложение

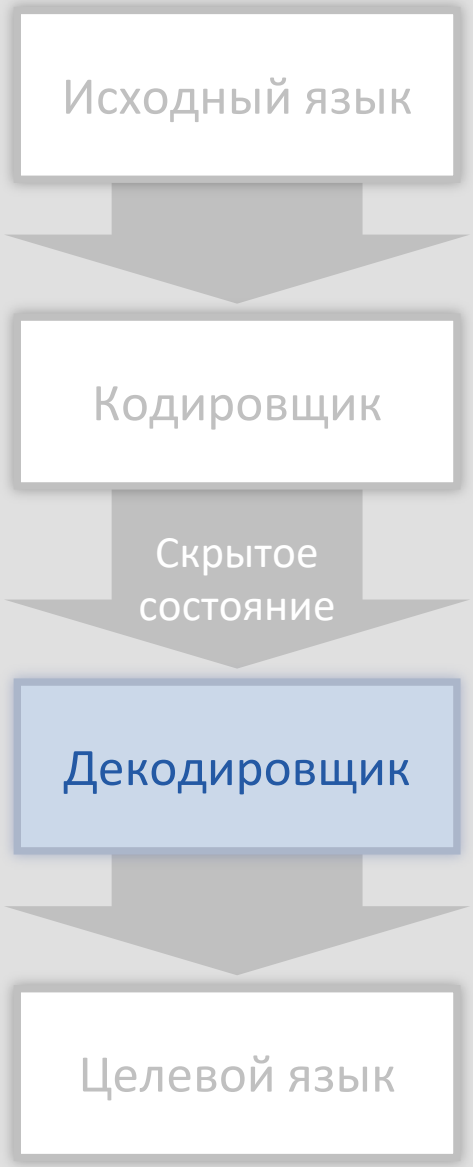




Архитектура декодировщика

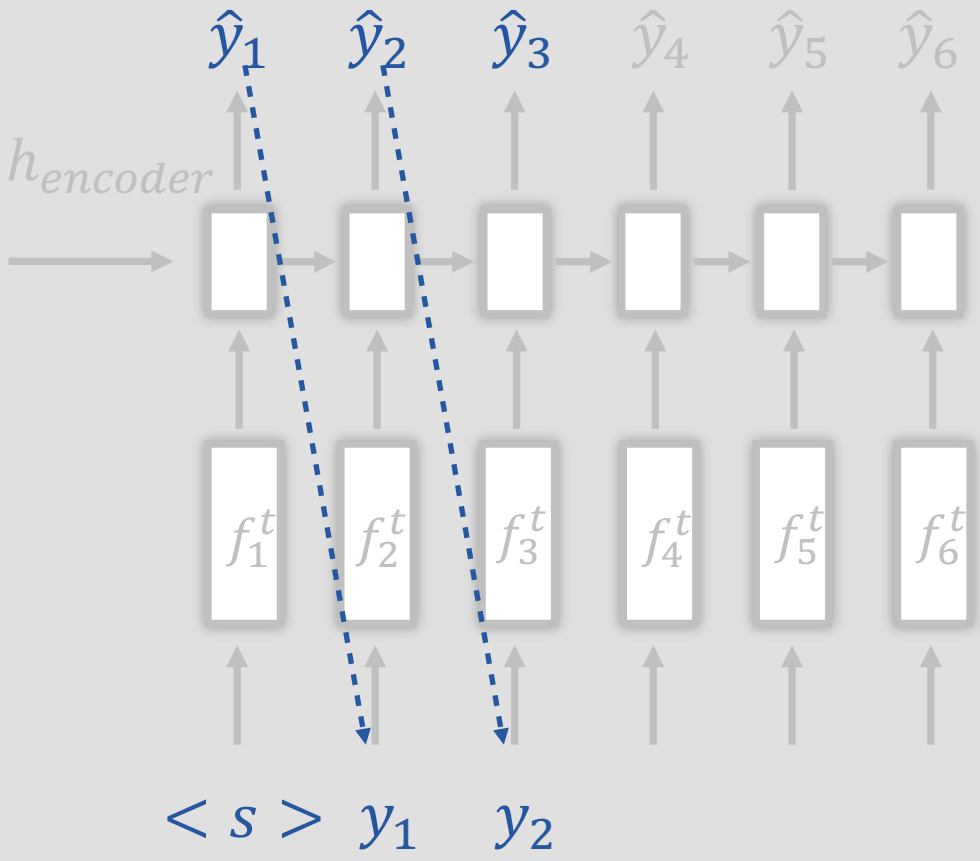
- Раскодировать целевое предложение

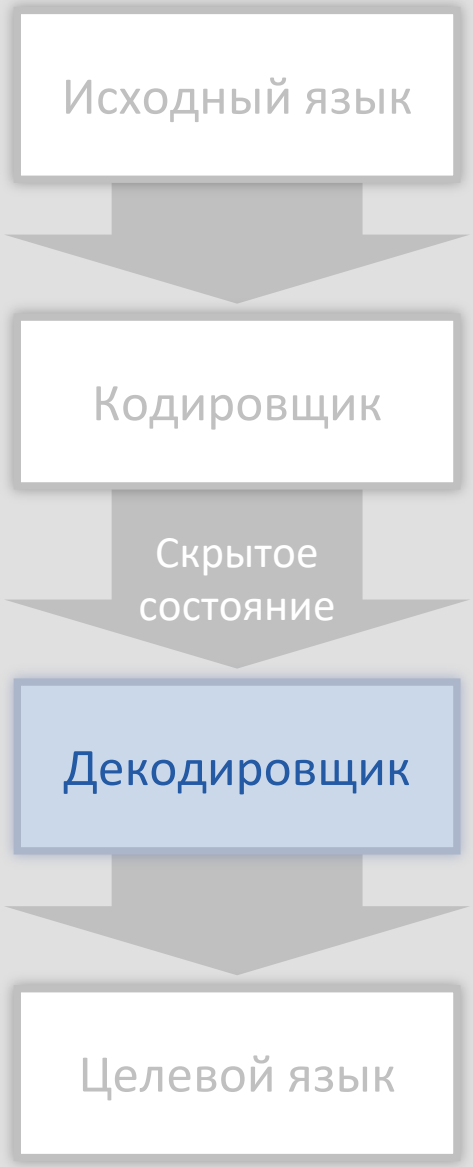




Архитектура декодировщика

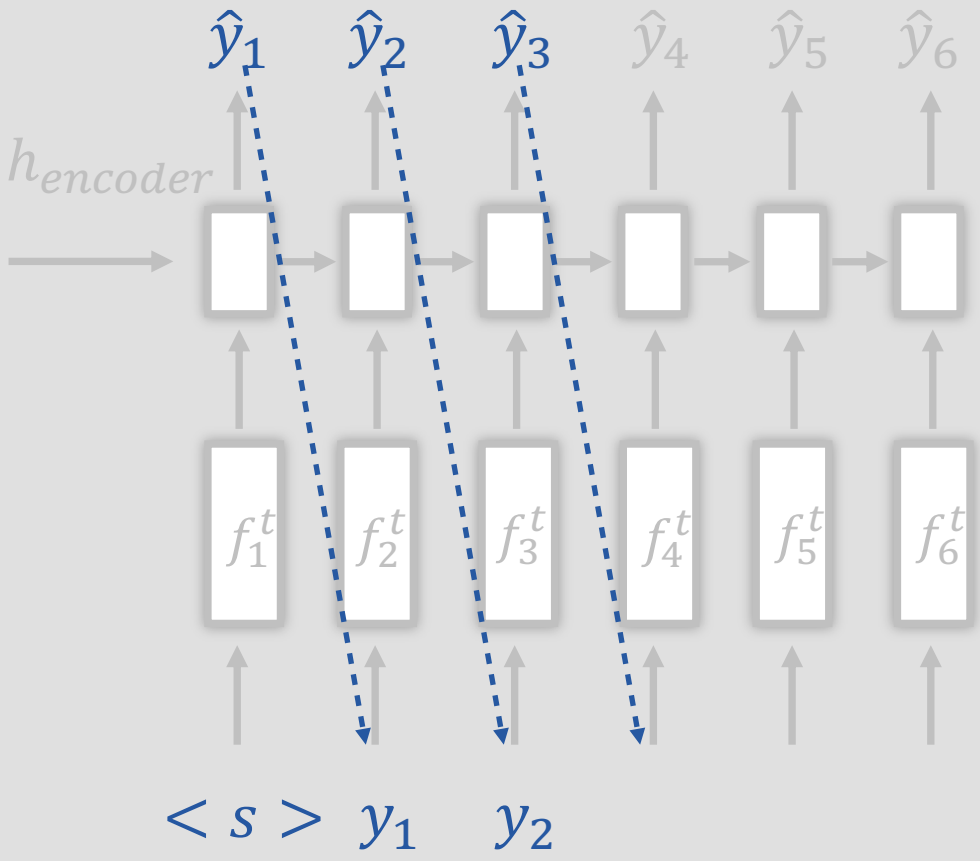
- Раскодировать целевое предложение

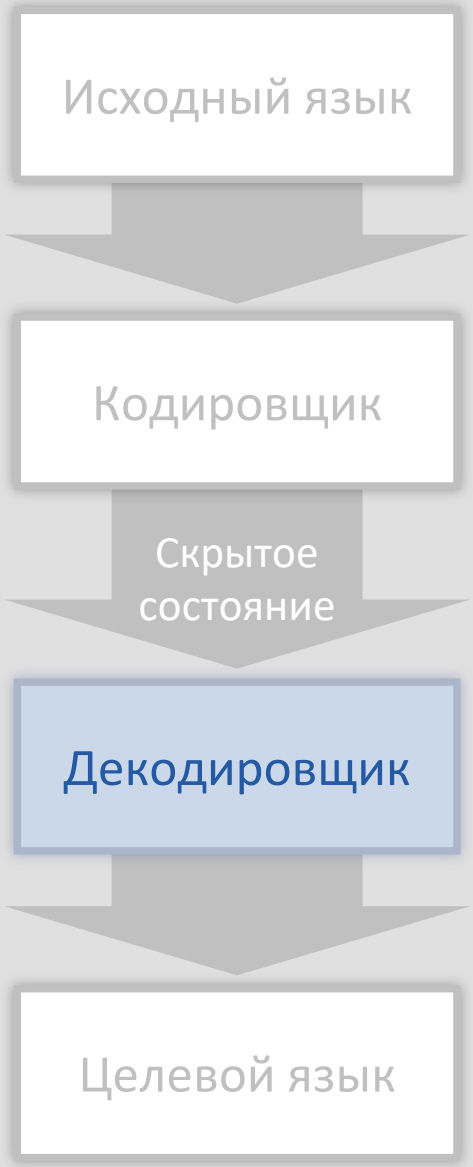




Архитектура декодировщика

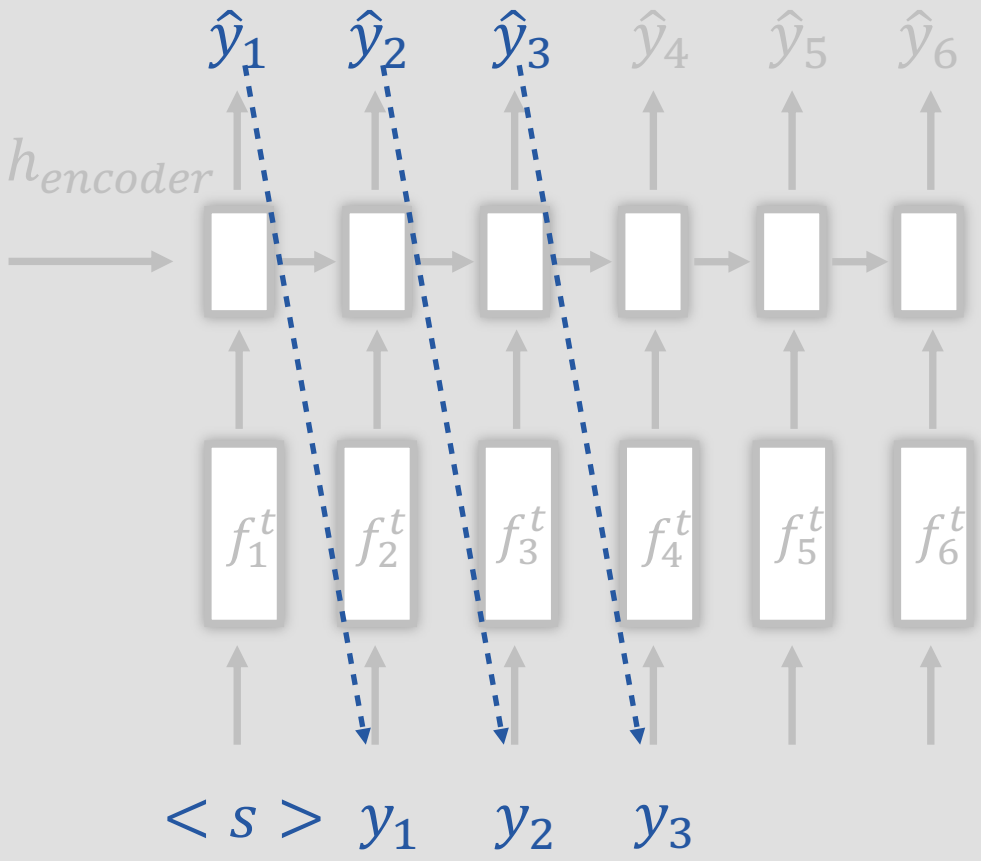
- Раскодировать целевое предложение

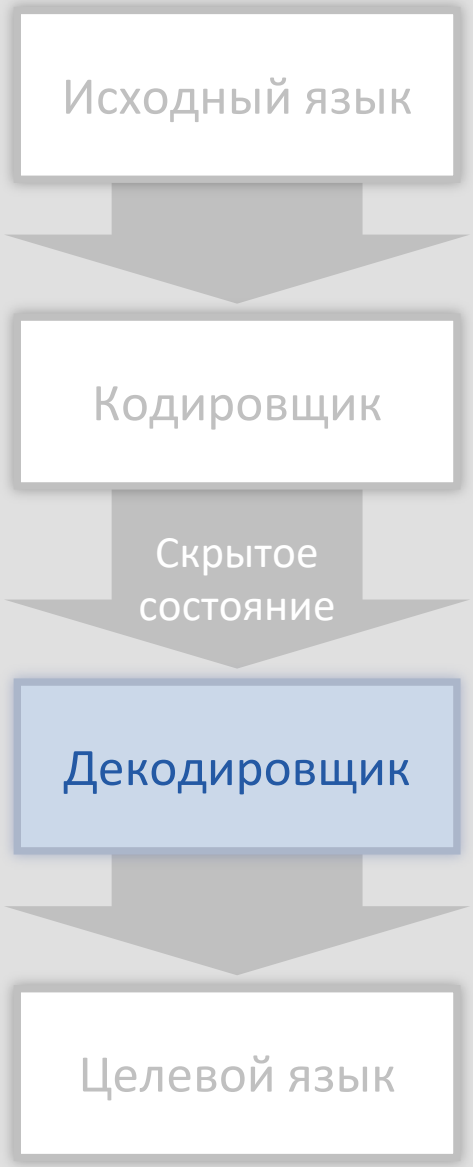




Архитектура декодировщика

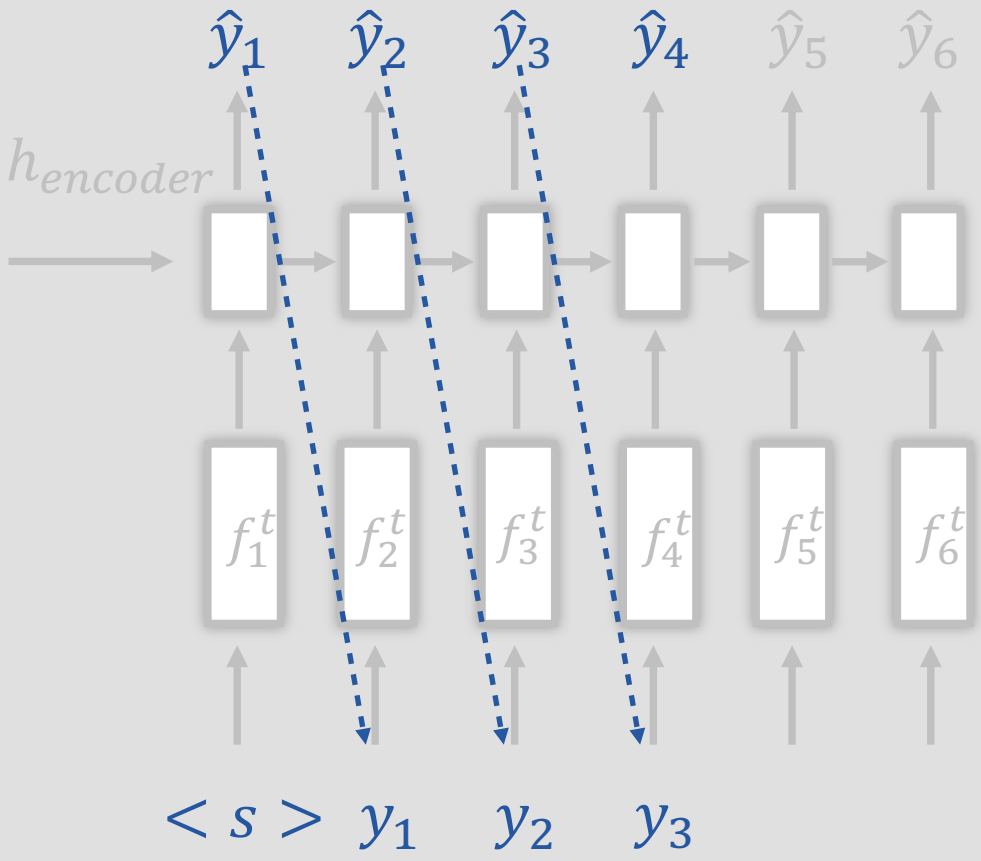
- Раскодировать целевое предложение

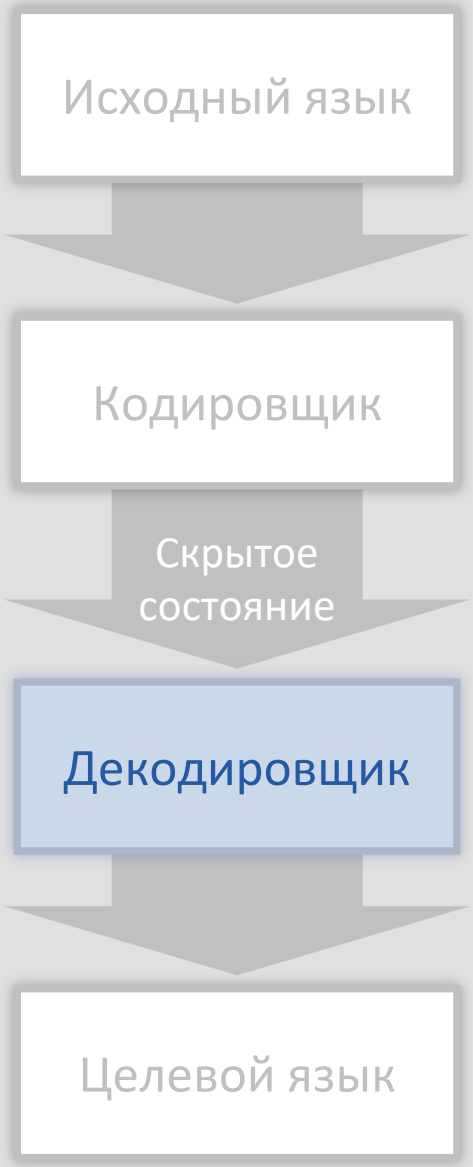




Архитектура декодировщика

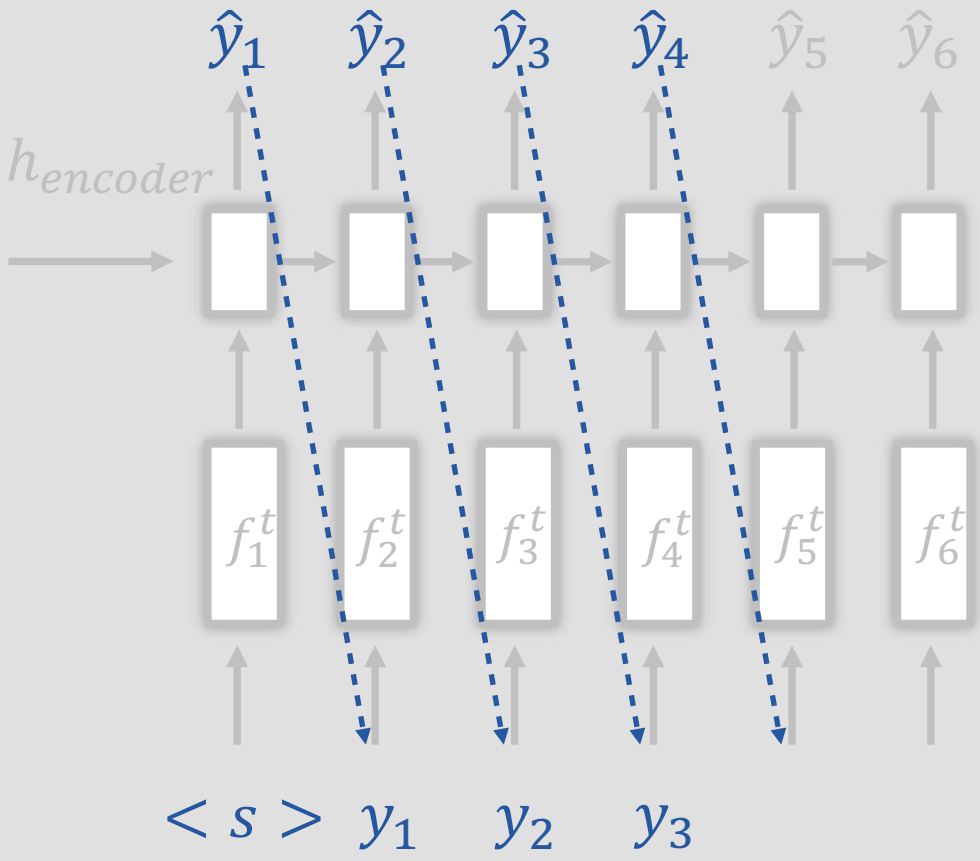
- Раскодировать целевое предложение

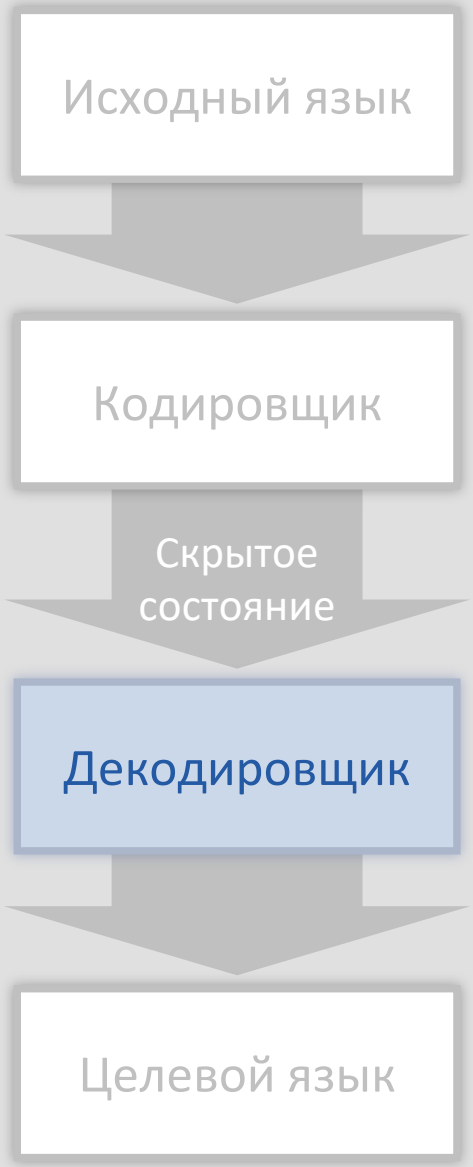




Архитектура декодировщика

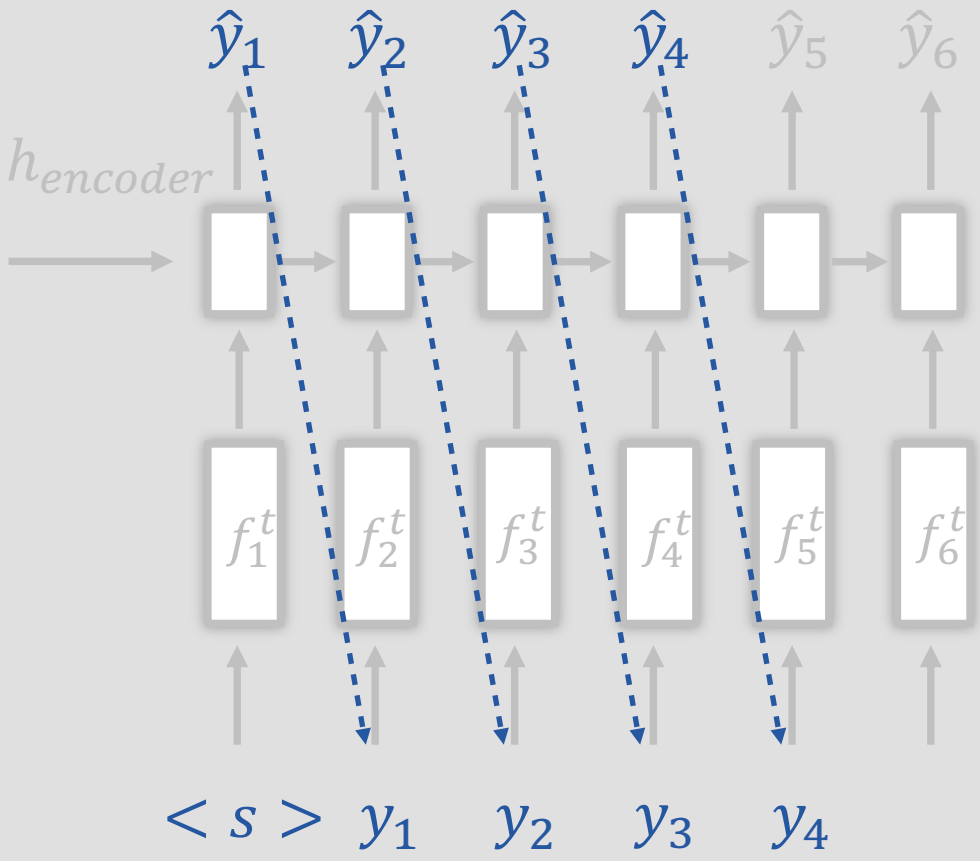
- Раскодировать целевое предложение

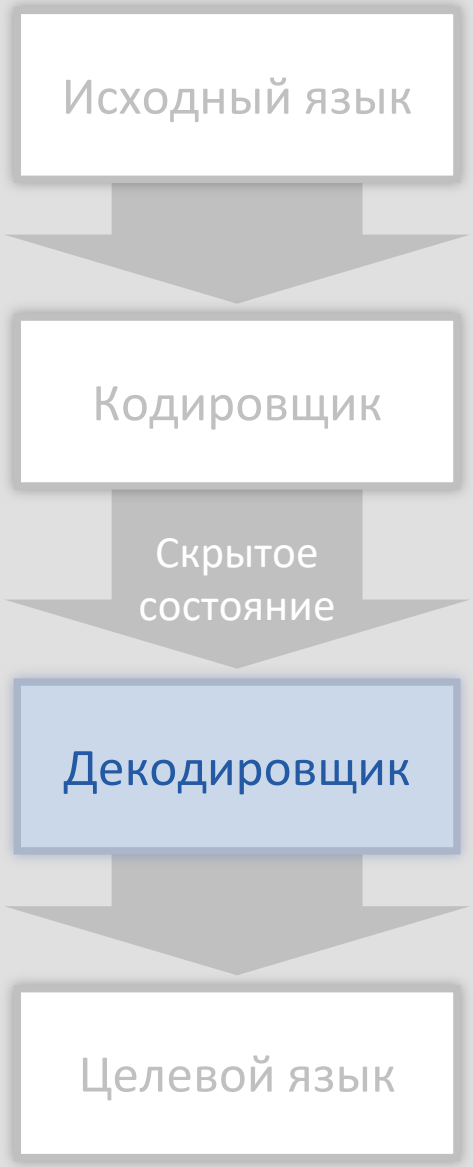




Архитектура декодировщика

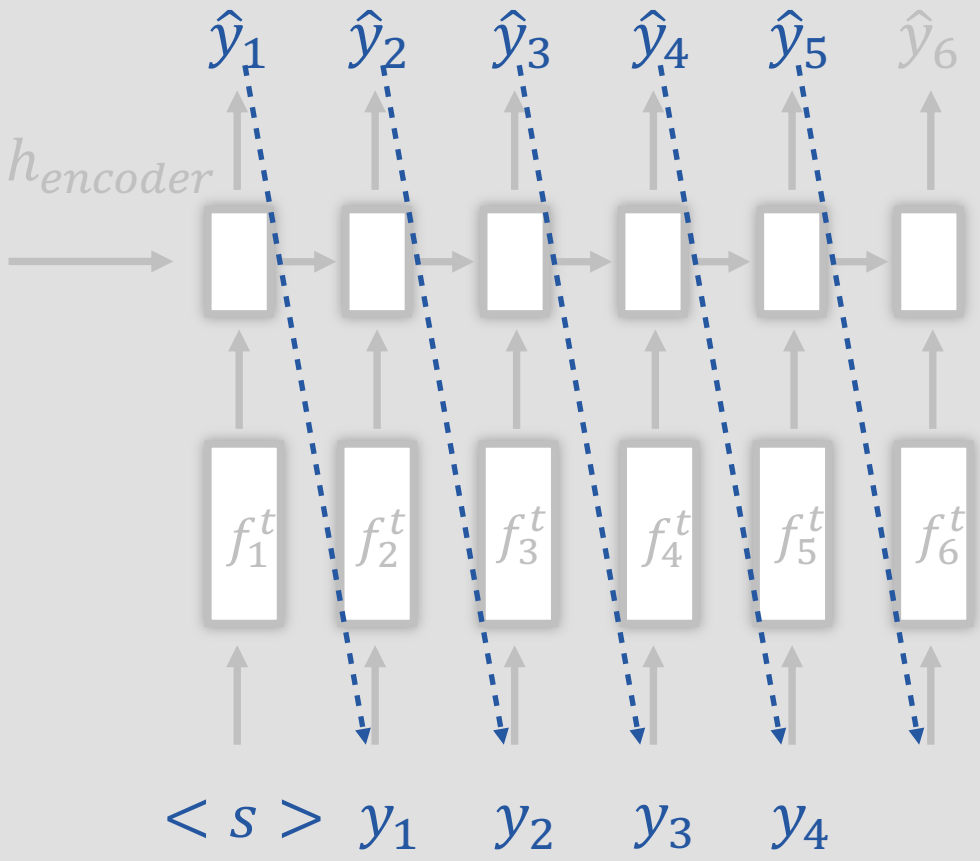
- Раскодировать целевое предложение

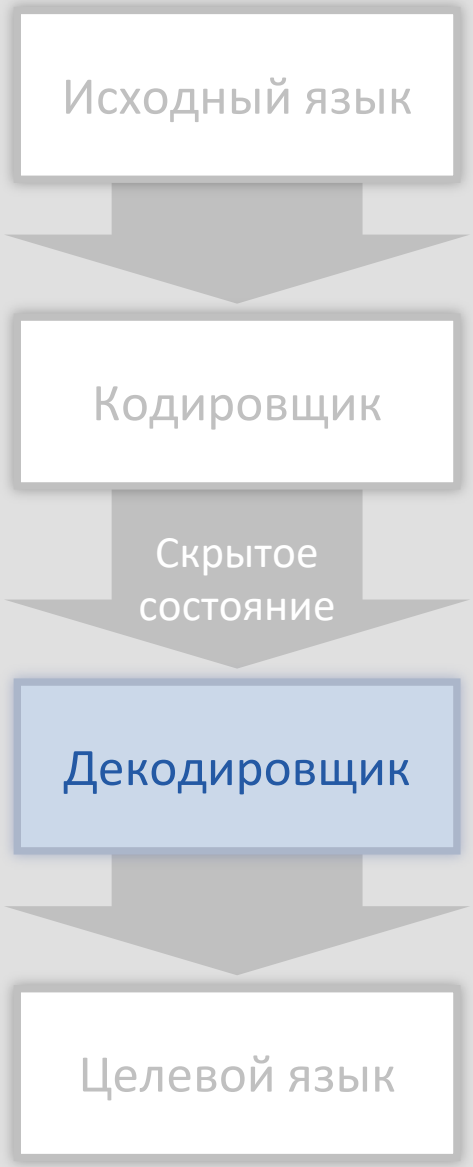




Архитектура декодировщика

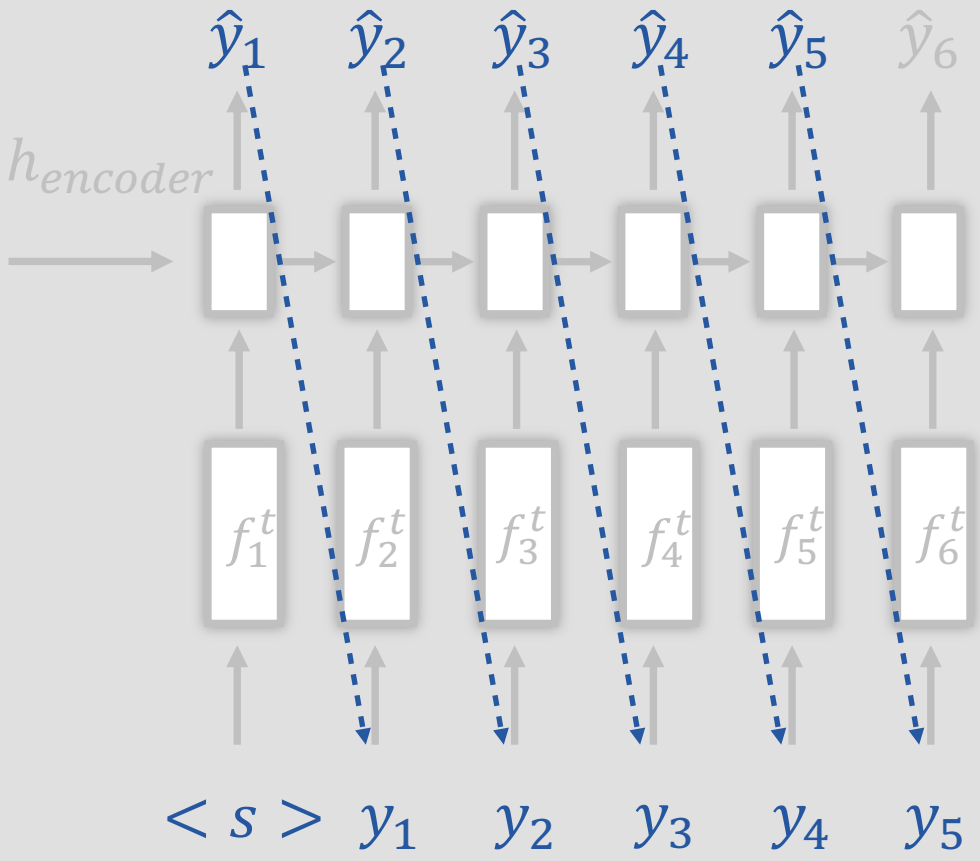
- Раскодировать целевое предложение

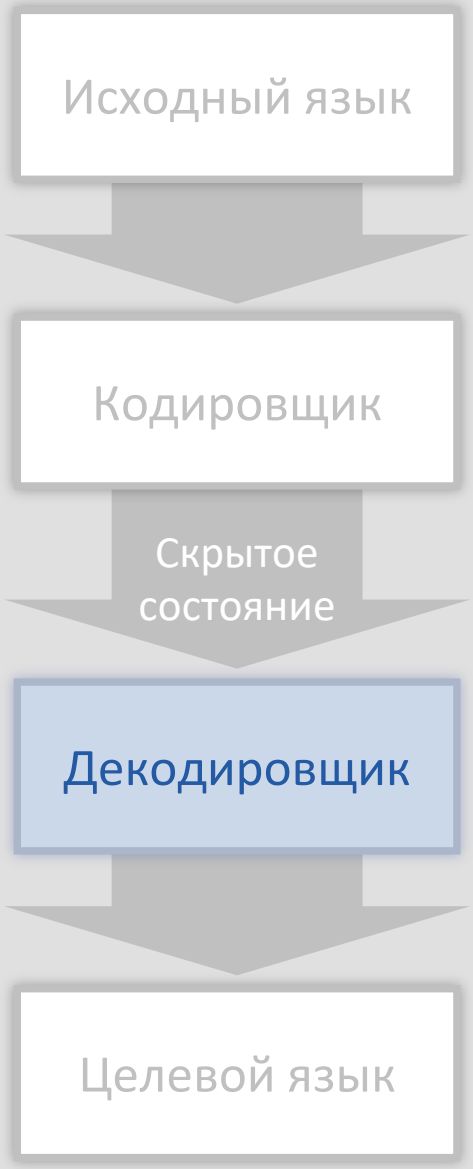




Архитектура декодировщика

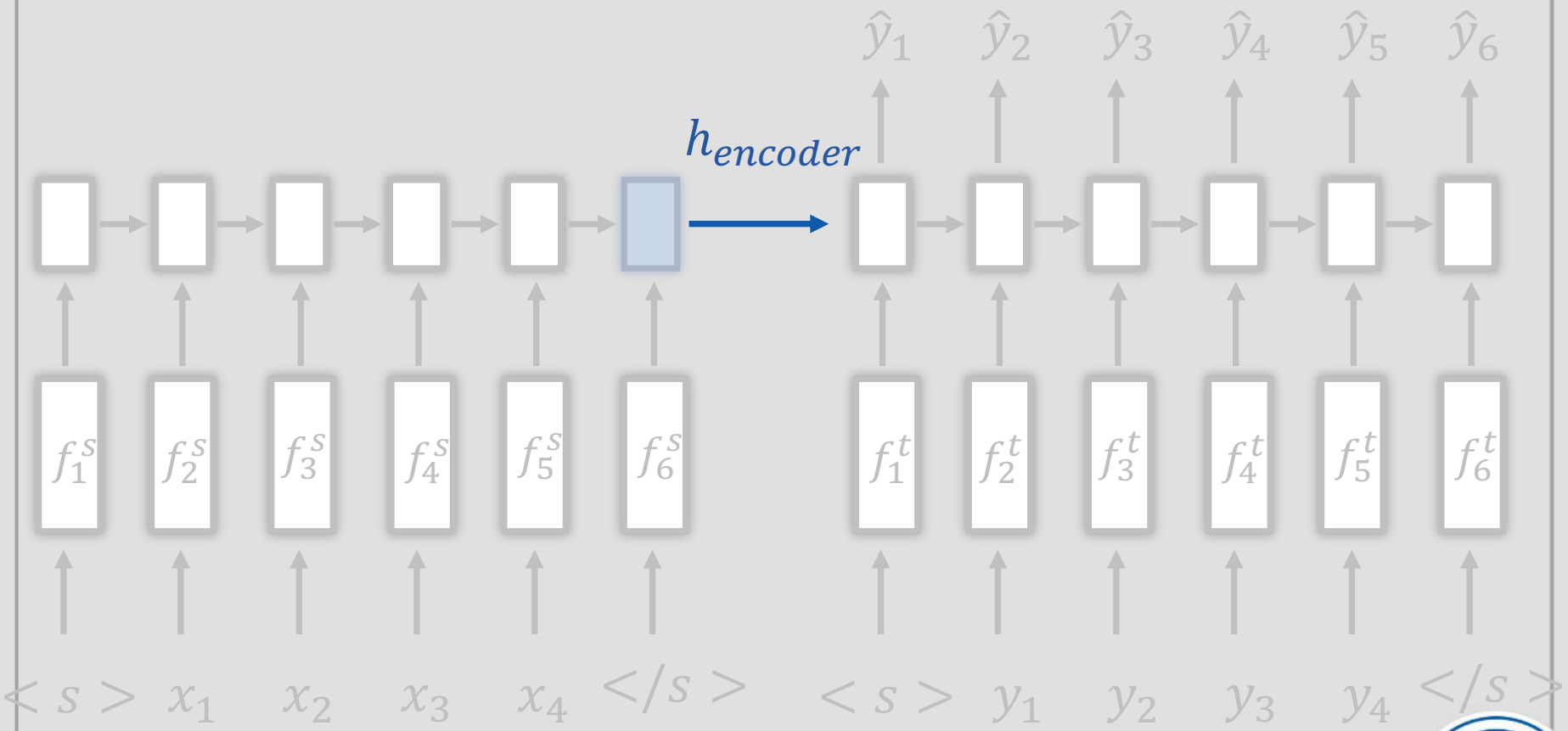
- Раскодировать целевое предложение





Нейросетевой машинный перевод

- Архитектура кодировщика



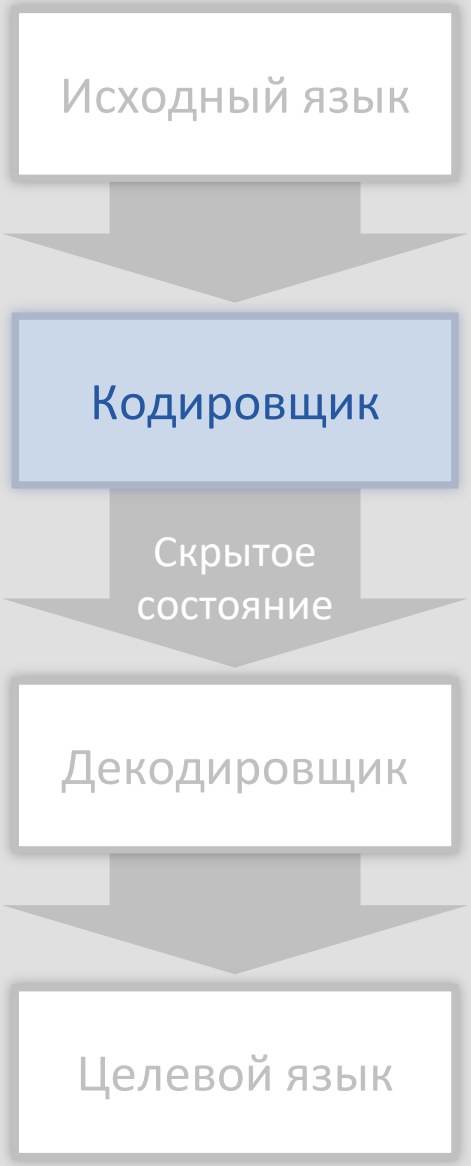
Модель кодировщик - декодировщик

Состоит из двух частей:

Модель кодировщик - декодировщик

Состоит из двух частей:

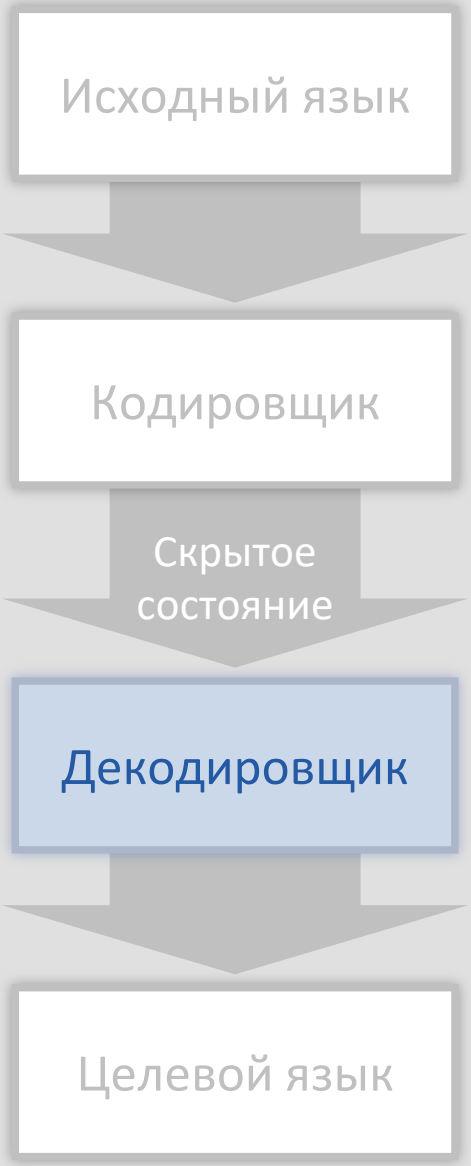
- Кодировщик кодирует входную последовательность слов



Модель кодировщик - декодировщик

Состоит из двух частей:

- Кодировщик кодирует входную последовательность слов
- Декодировщик генерирует выходную последовательность слов

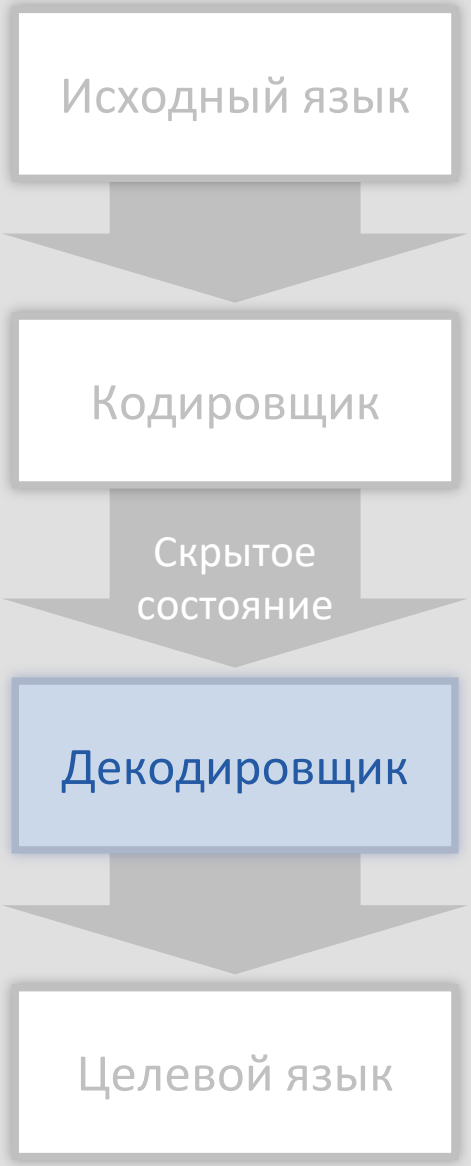


Модель кодировщик - декодировщик

Состоит из двух частей:

- Кодировщик кодирует входную последовательность слов
- Декодировщик генерирует выходную последовательность слов

Декодировщик – это языковая модель!



Модель кодировщик - декодировщик

Модели класса кодировщик-декодировщик могут быть использованы в других приложениях:

Модель кодировщик - декодировщик

Модели класса кодировщик-декодировщик могут быть использованы в других приложениях:

- Вопросно-ответные системы

Модель кодировщик - декодировщик

Модели класса кодировщик-декодировщик могут быть использованы в других приложениях:

- Вопросно-ответные системы
- Генерация подписи к изображению (кодировщик должен обрабатывать изображение)

Модель кодировщик - декодировщик

Модели класса кодировщик-декодировщик могут быть использованы в других приложениях:

- Вопросно-ответные системы
- Генерация подписи к изображению (кодировщик должен обрабатывать изображение)
- Распознавание речи (кодировщик обрабатывает аудиозапись)

Модель кодировщик - декодировщик

- Проблема: используется только одно, последнее скрытое состояние кодировщика ⇒ мало данных для декодирования

Модель кодировщик - декодировщик

- Проблема: используется только одно, последнее скрытое состояние кодировщика ⇒ мало данных для декодирования
- Последние входные слова оказывают большее влияние, чем первые

Модель кодировщик - декодировщик

- Проблема: используется только одно, последнее скрытое состояние кодировщика ⇒ мало данных для декодирования
- Последние входные слова оказывают большее влияние, чем первые
- Лучше переводятся короткие предложения

Модель кодировщик - декодировщик

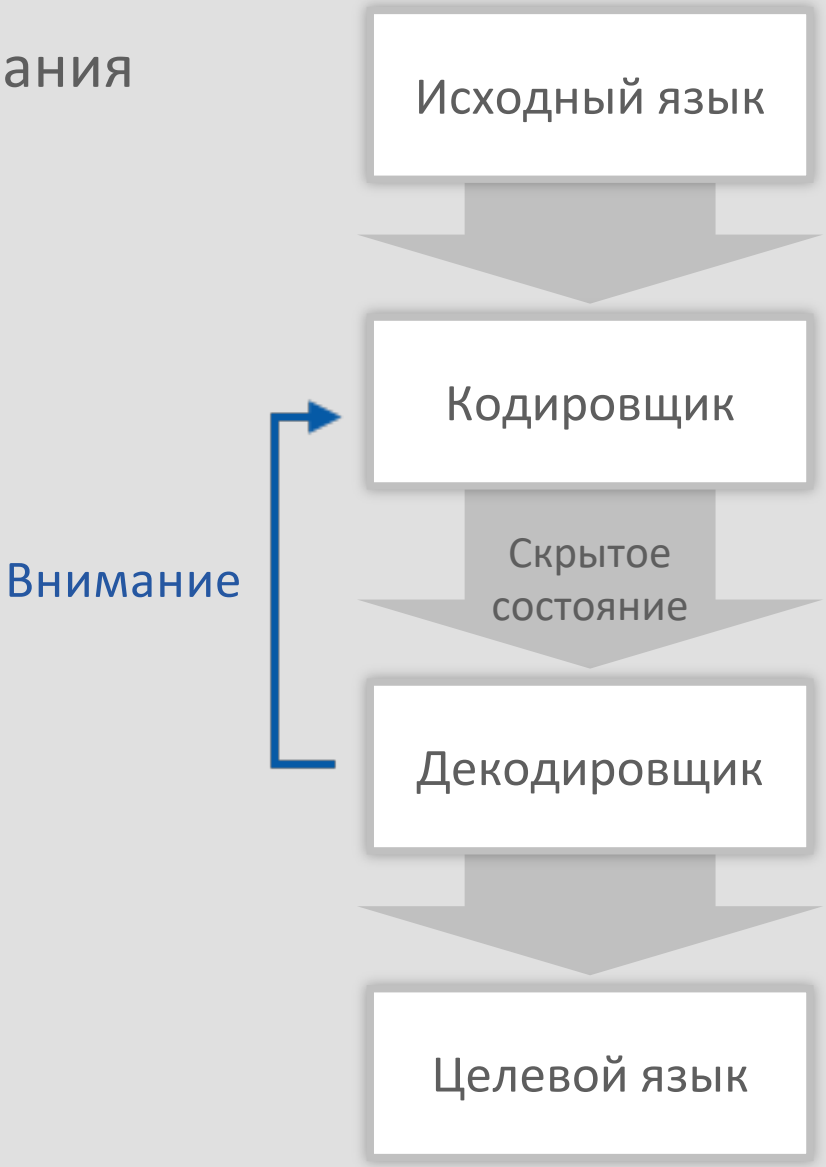
- Проблема: используется только одно, последнее скрытое состояние кодировщика ⇒ мало данных для декодирования
- Последние входные слова оказывают большее влияние, чем первые
- Лучше переводятся короткие предложения
- Архитектуру можно существенно улучшить, добавив механизм внимания

Механизм внимания

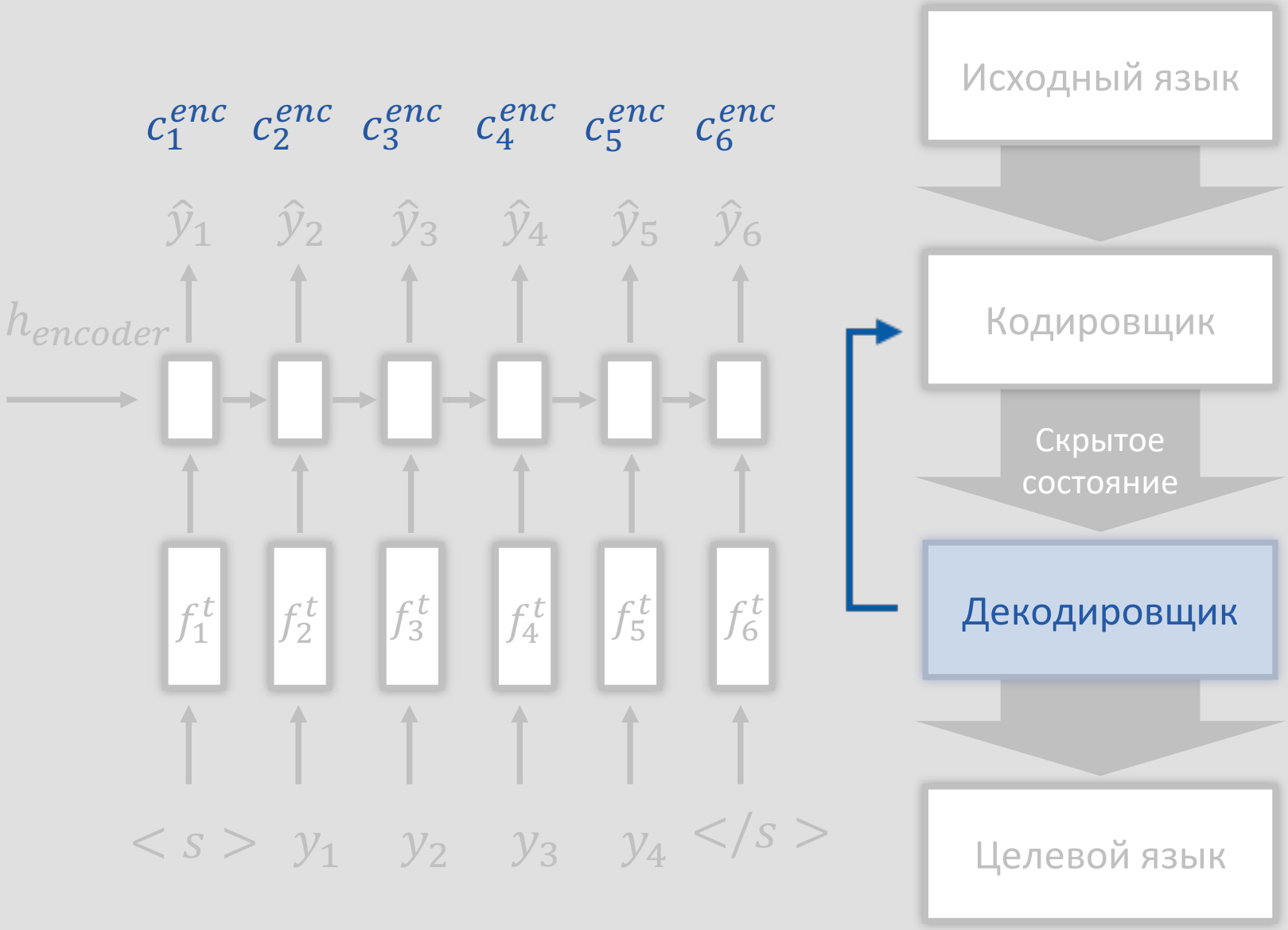
Механизм внимания

Механизм внимания

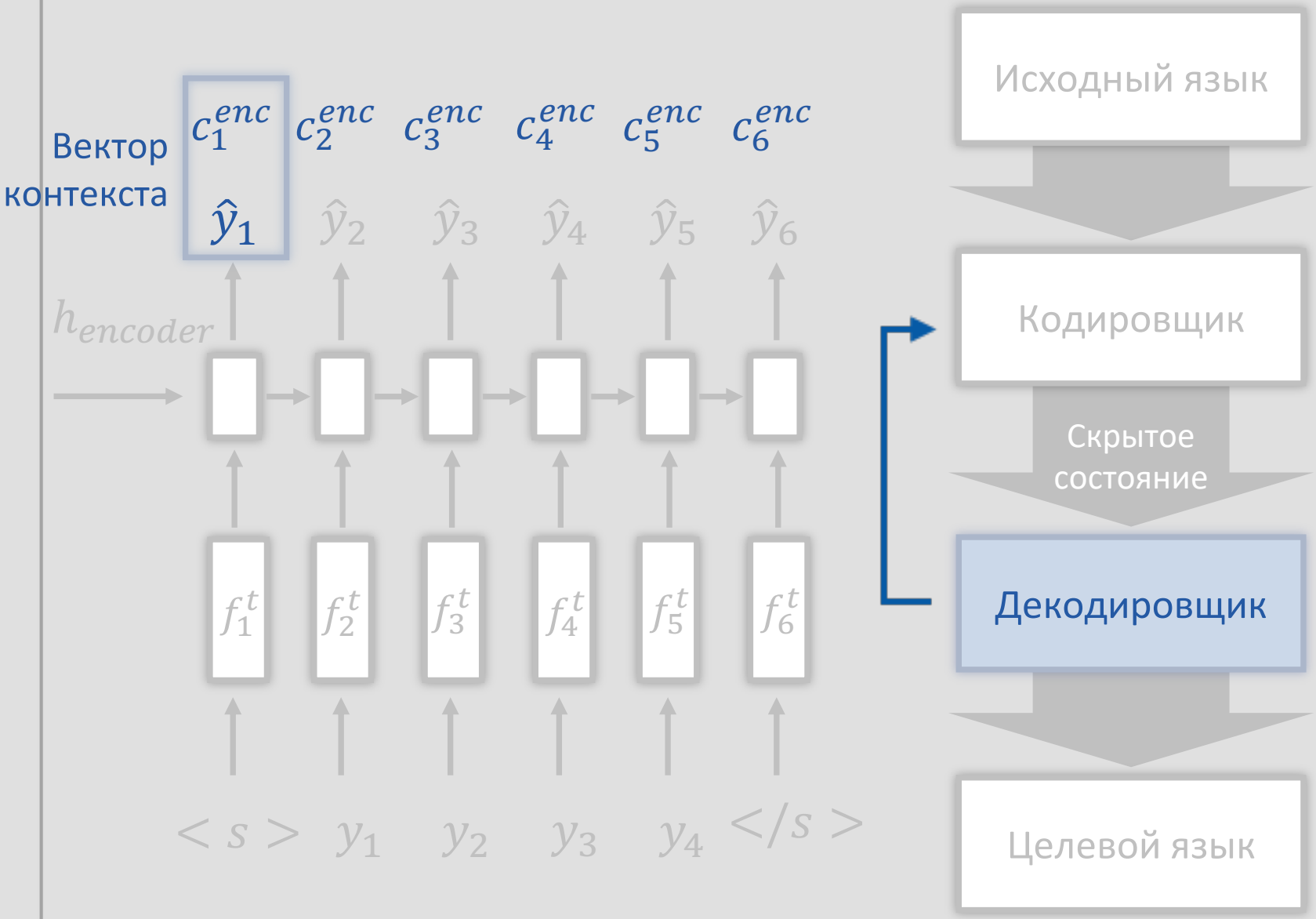
- Включить механизм внимания



Механизм внимания



Механизм внимания



Механизм внимания

- Веса выравнивания:

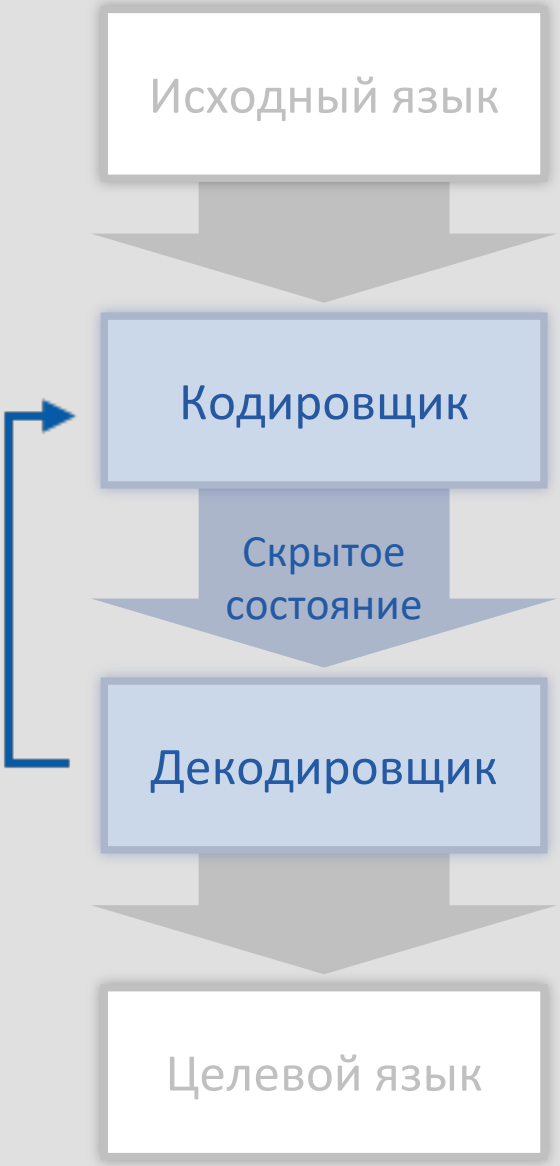
$$a_{ij} = sim(h_i^{enc}, h_j^{dec})$$

- Нормировка весов внимания:

$$\alpha_{ij} = \frac{e^{a_{ij}}}{\sum_k e^{a_{kj}}}$$

- Вектор контекста:

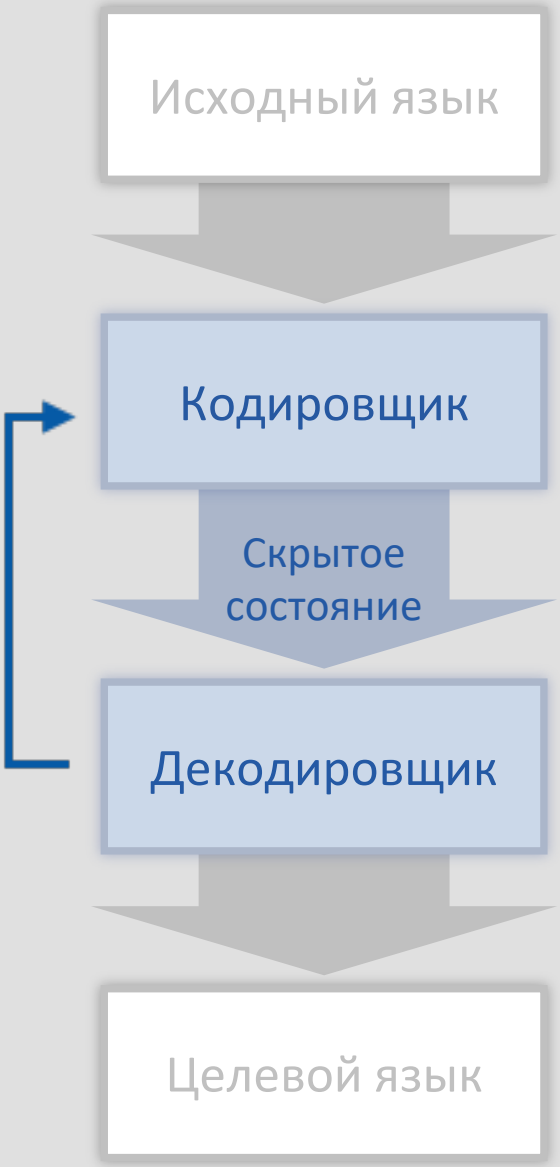
$$c_j = \sum_i \alpha_{ij} h_i^{enc}$$



Механизм внимания

- Скалярное произведение:

$$sim(h, s) = h^{\top} s$$



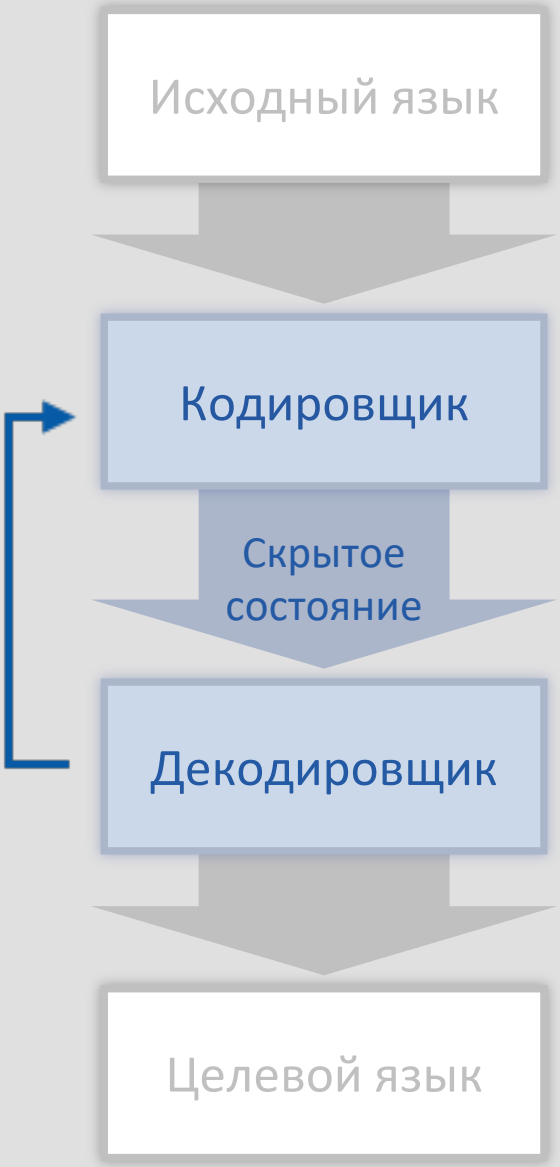
Механизм внимания

- Скалярное произведение:

$$sim(h, s) = h^{\top} s$$

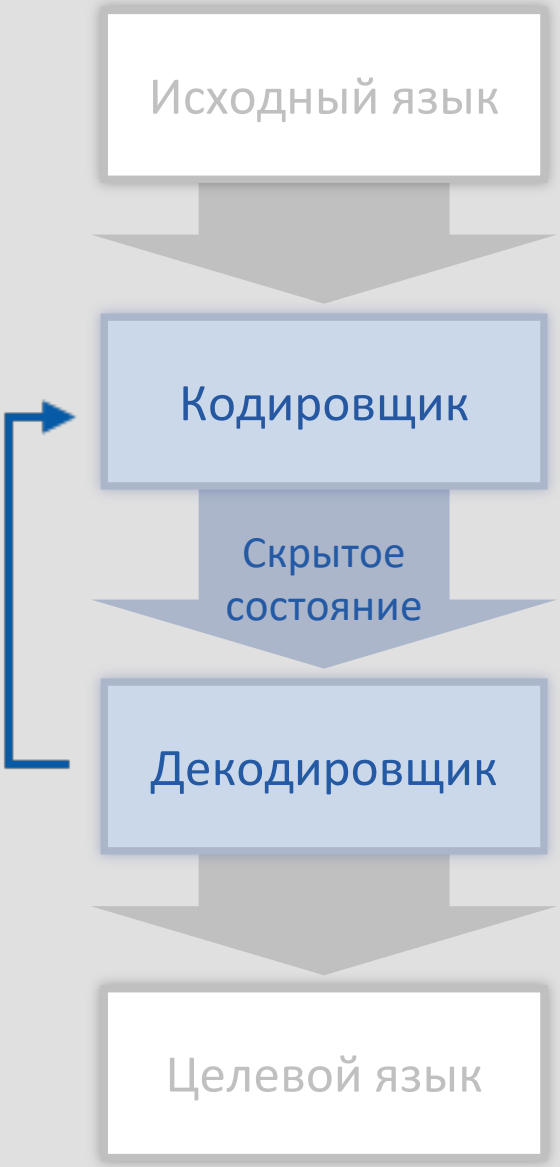
- Аддитивный механизм внимания:

$$sim(h, s) = w^{\top} \tanh(W_h h + W_s s)$$



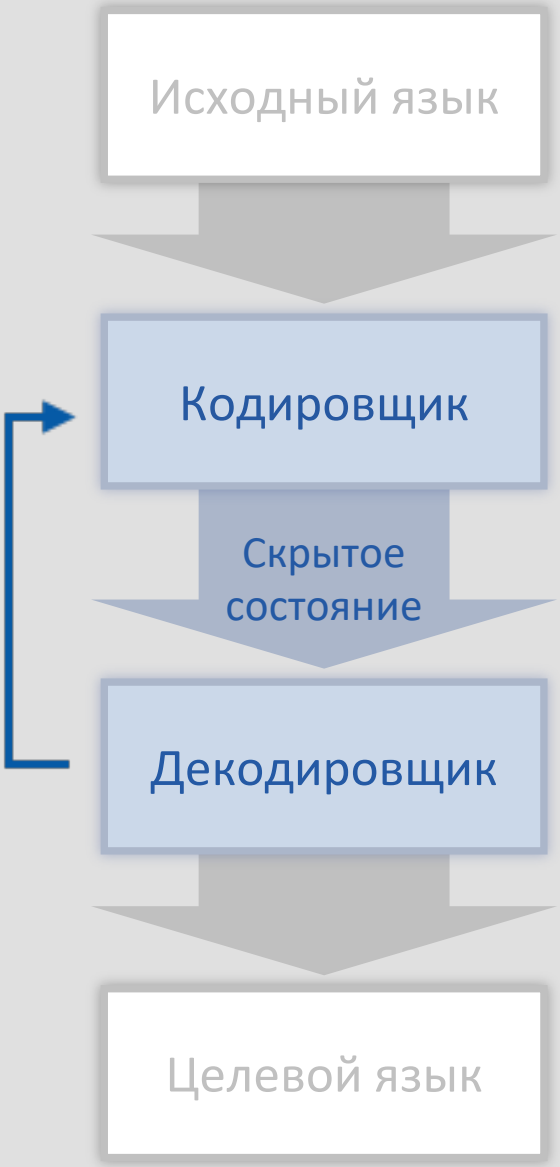
Механизм внимания

- Скалярное произведение:
$$sim(h, s) = h^\top s$$
- Аддитивный механизм внимания:
$$sim(h, s) = w^\top \tanh(W_h h + W_s s)$$
- Мультипликативный механизм внимания:
$$sim(h, s) = h^\top W s$$

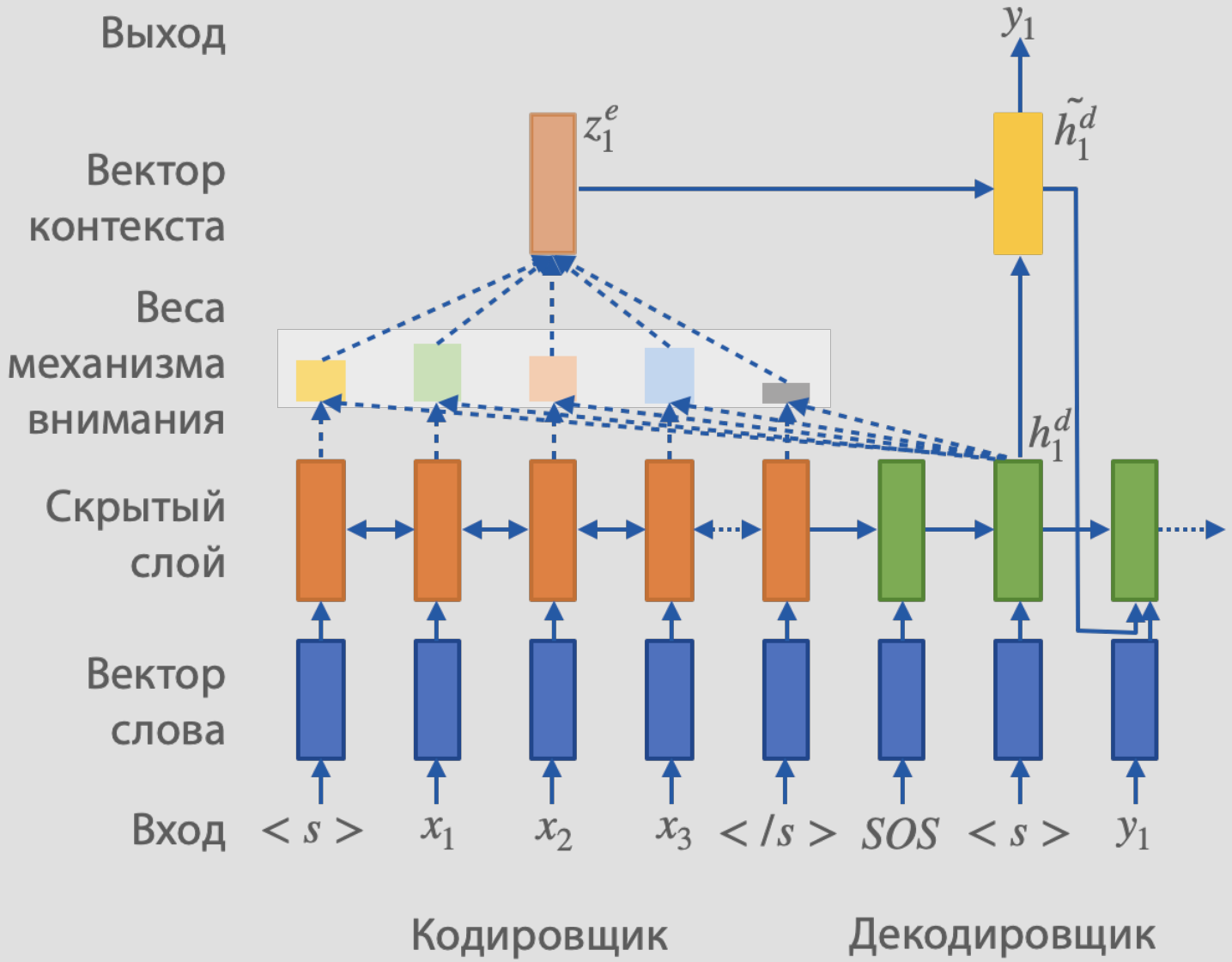


Механизм внимания

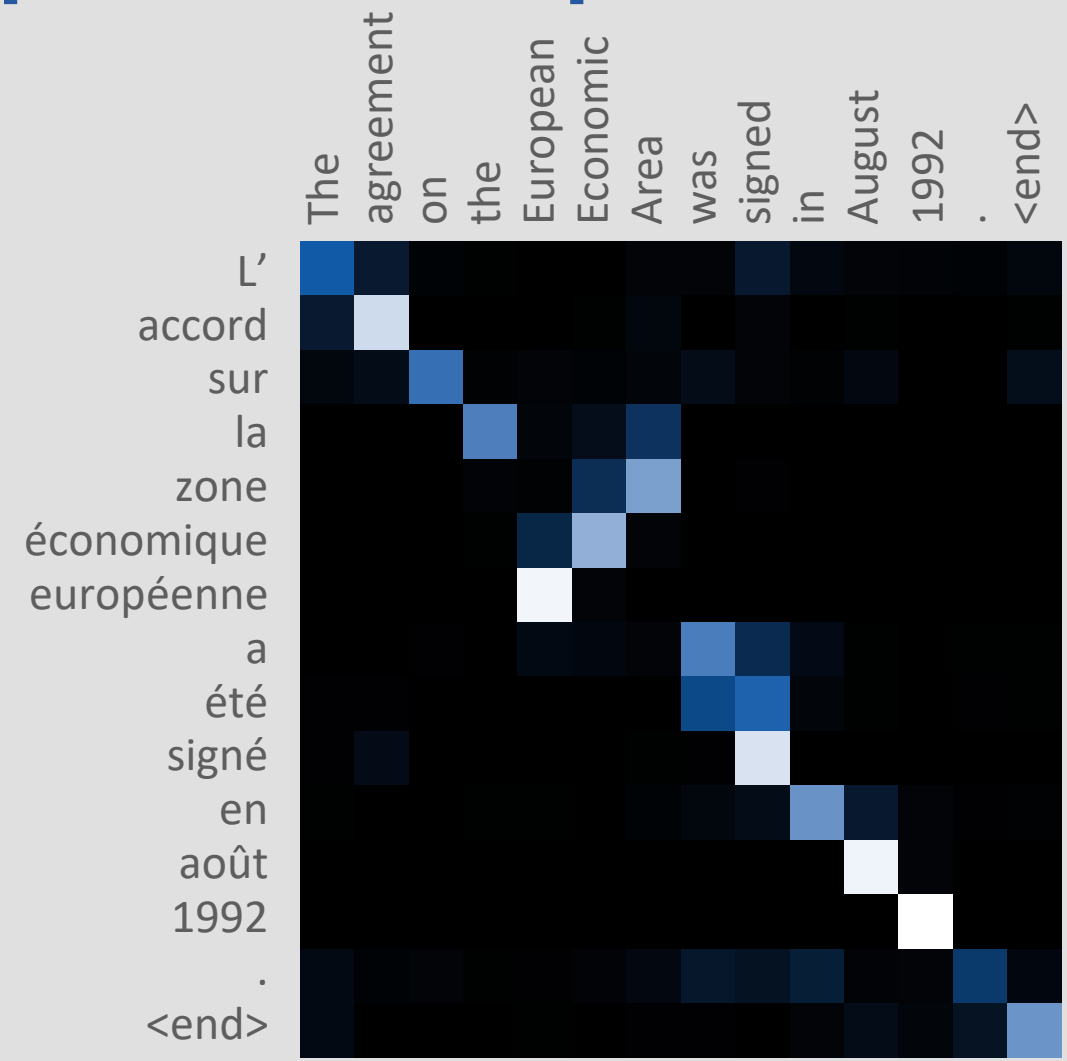
- Скалярное произведение:
$$sim(h, s) = h^\top s$$
- Аддитивный механизм внимания:
$$sim(h, s) = w^\top \tanh(W_h h + W_s s)$$
- Мультипликативный механизм внимания:
$$sim(h, s) = h^\top W s$$
- Параметры весовых функций (при наличии) обучаются вместе со всей моделью



Механизм внимания



Карты весов выравнивания



Вместо заключения

Вместо заключения

知 识 就 是 力 量 <end>

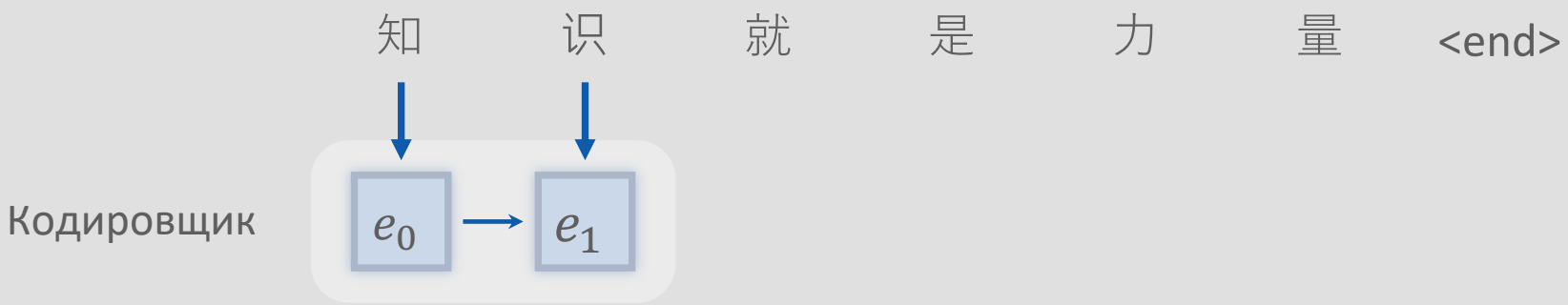
Вместо заключения

知 识 就 是 力 量 <end>

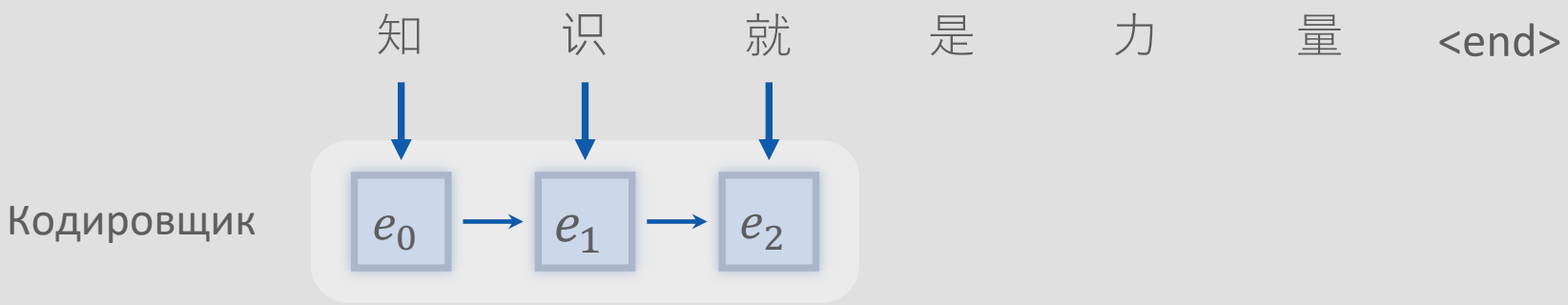
Кодировщик

e_0

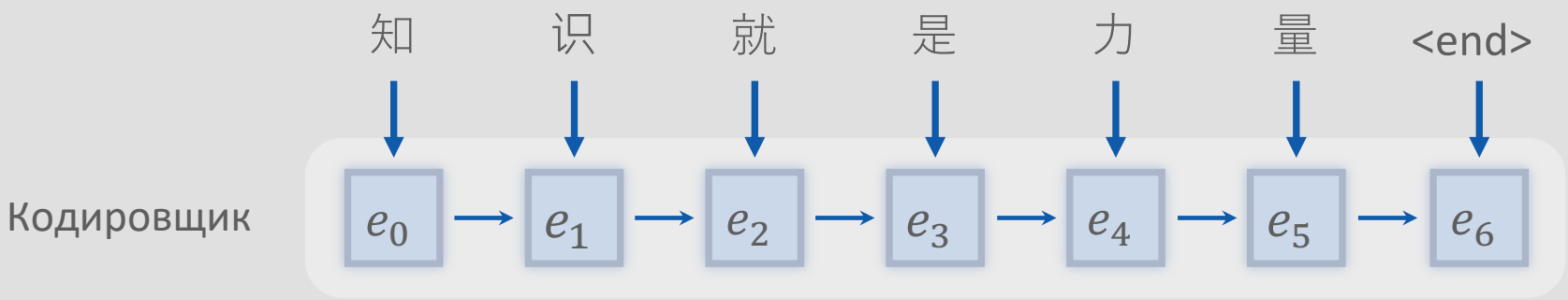
Вместо заключения



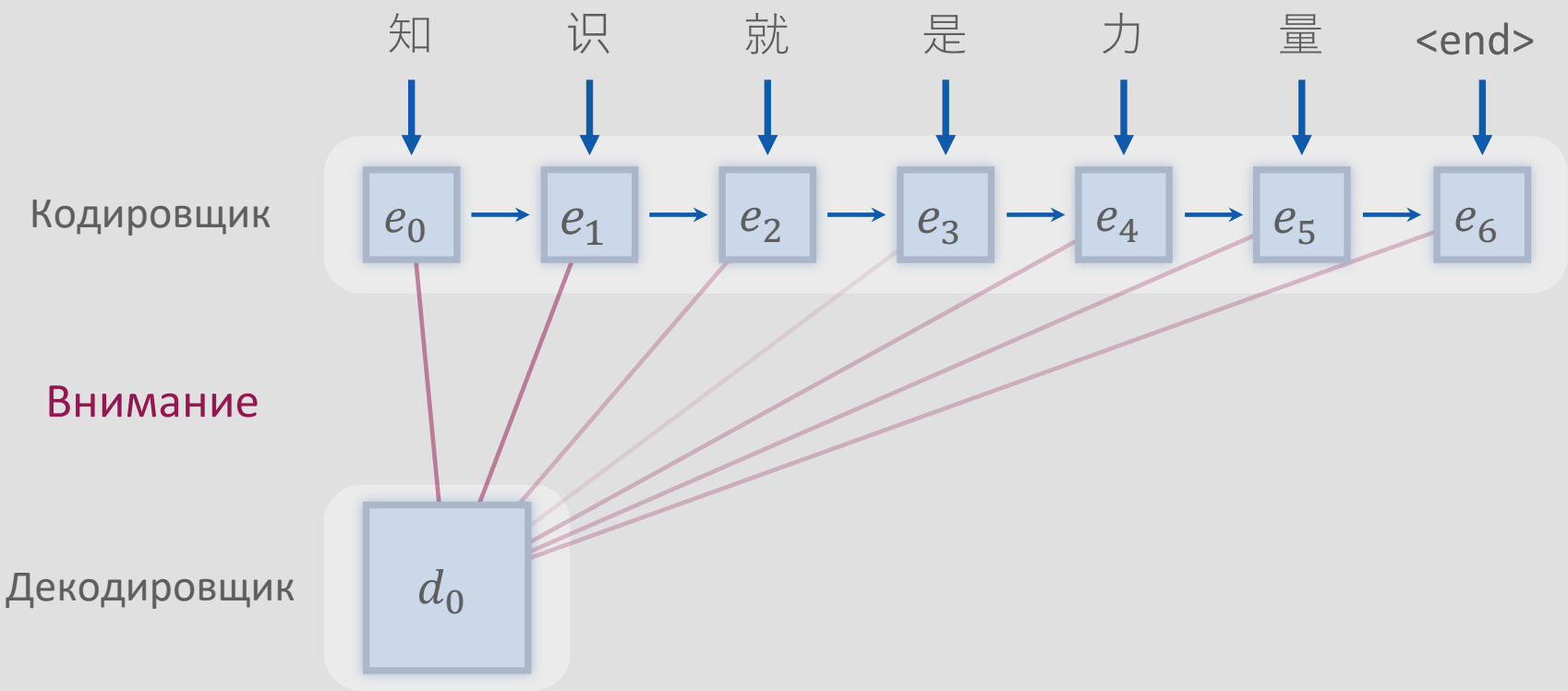
Вместо заключения



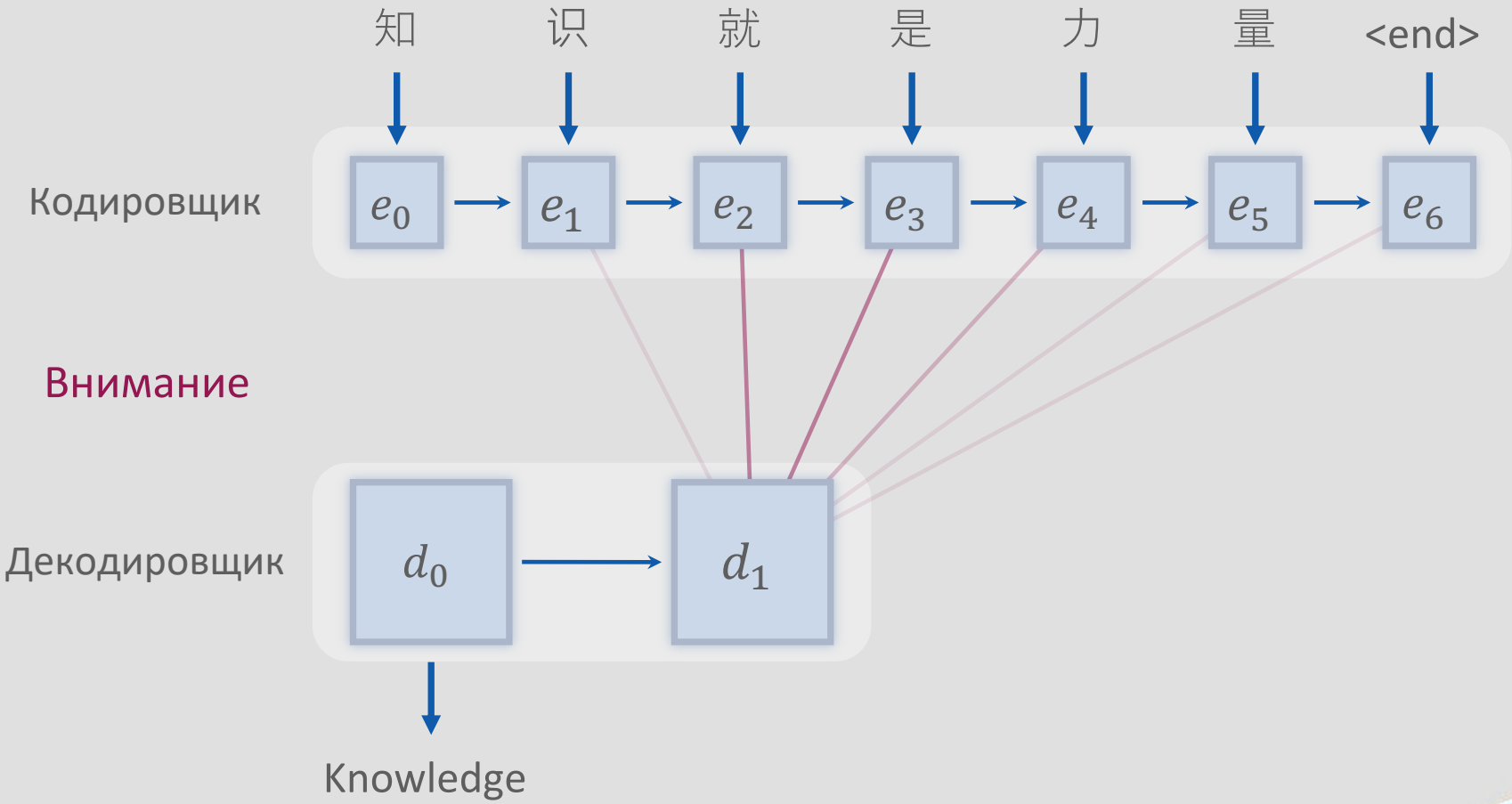
Вместо заключения



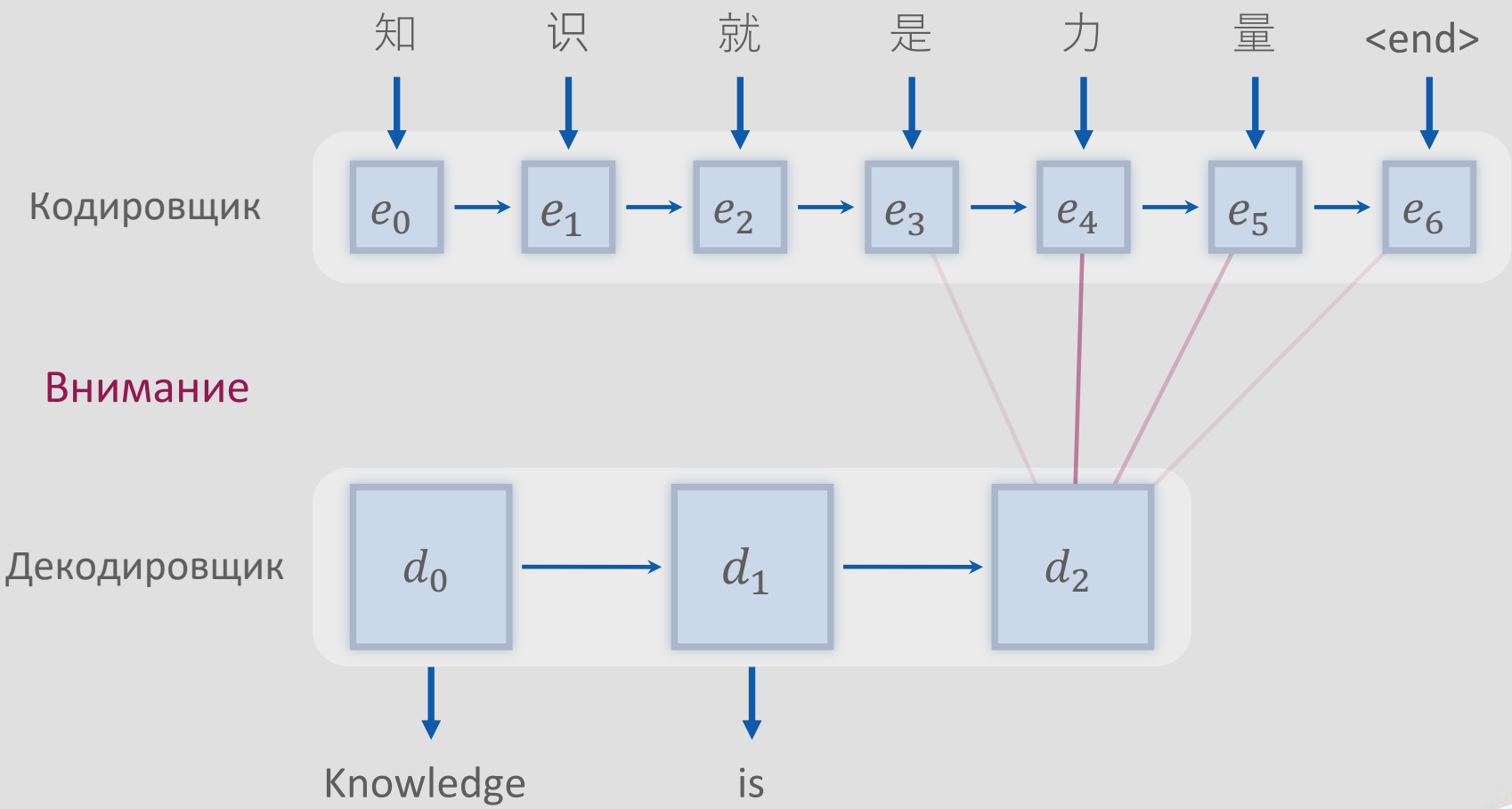
Вместо заключения



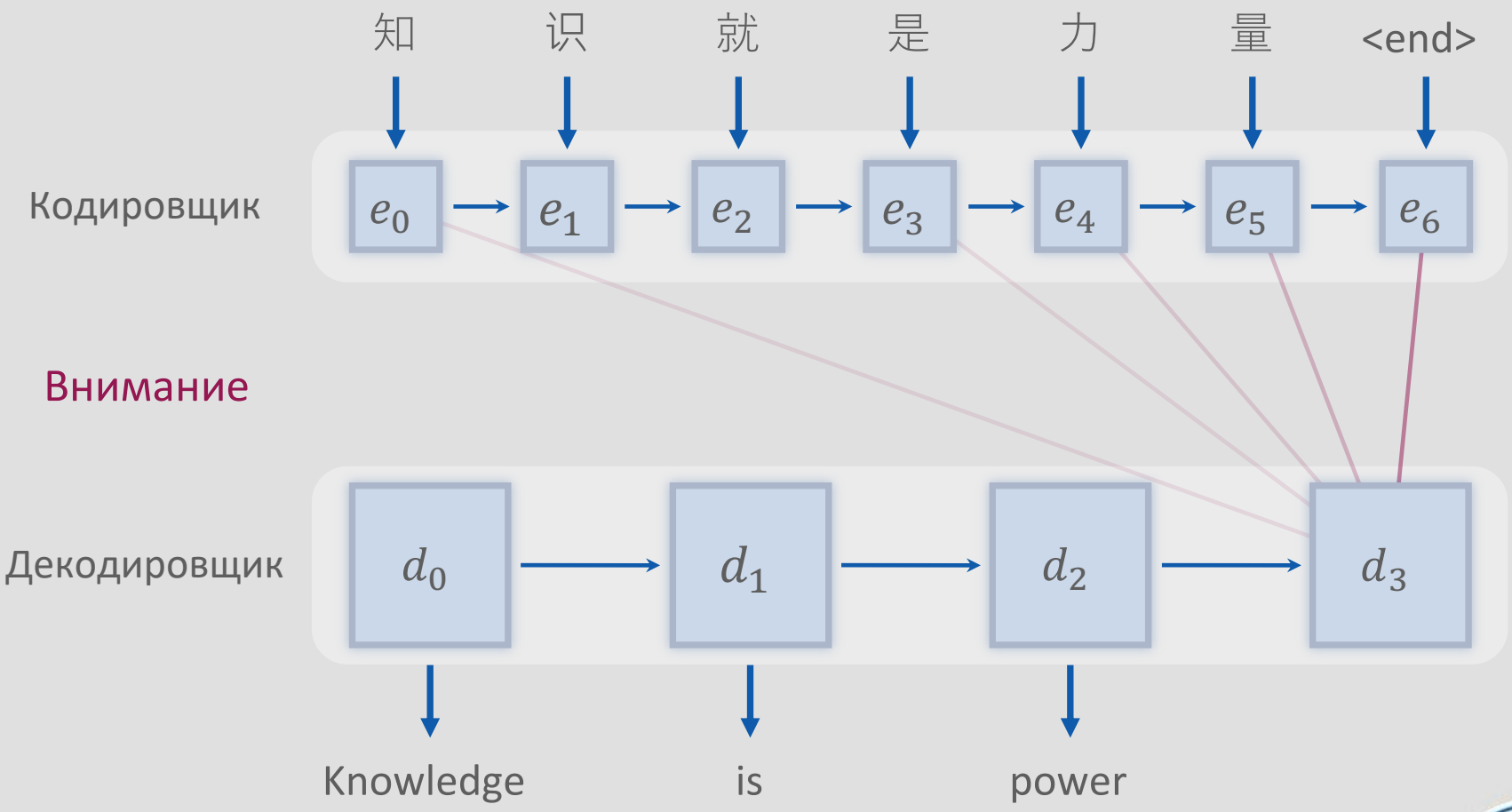
Вместо заключения



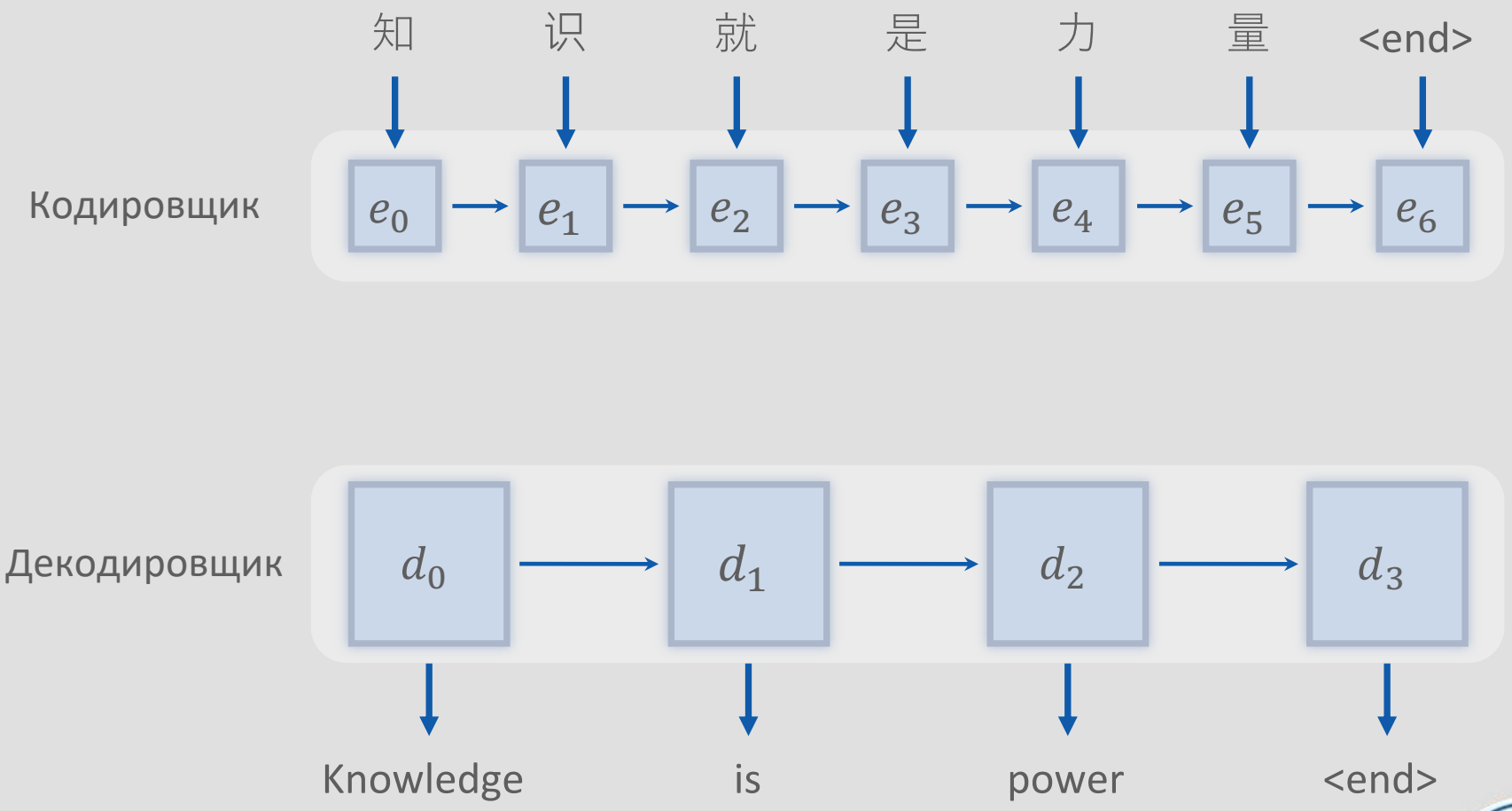
Вместо заключения



Вместо заключения

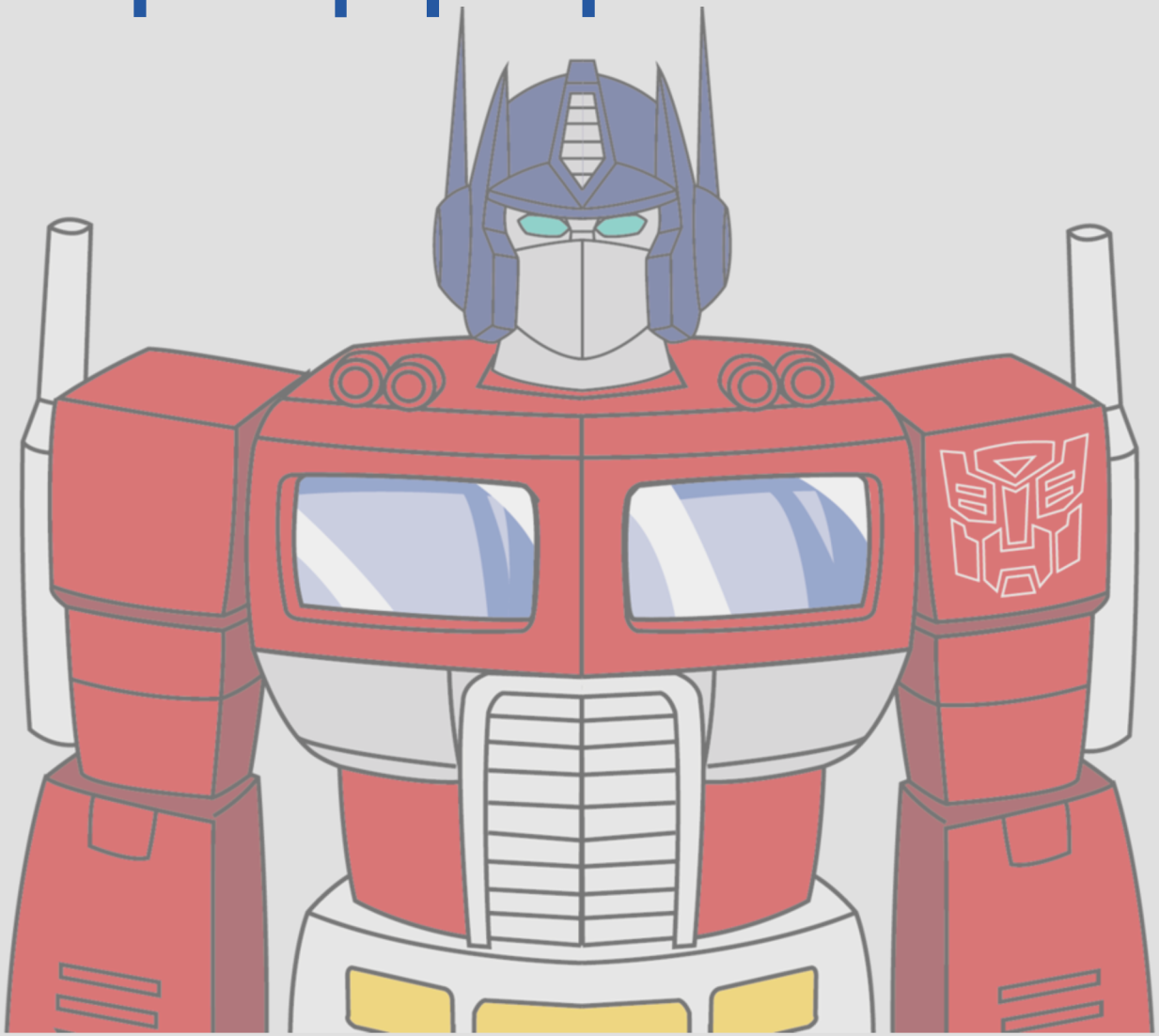


Вместо заключения



Модель Трансформер

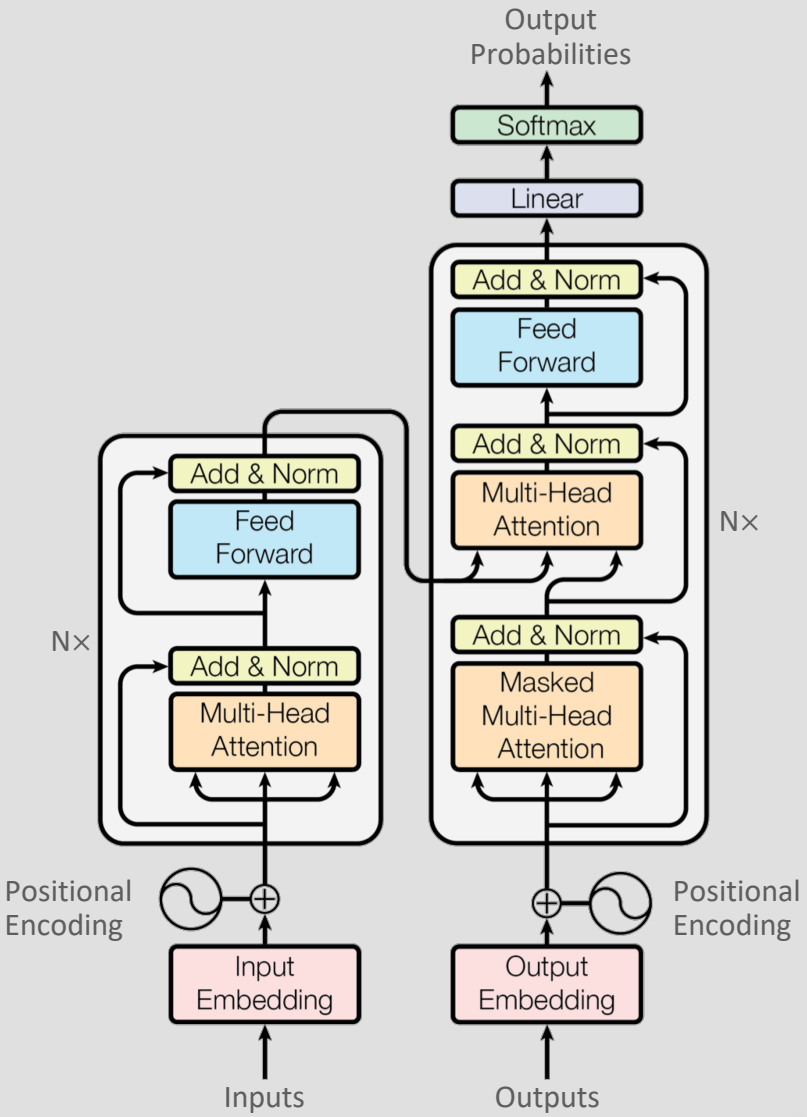
Модель Трансформер



- *Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." In Advances in neural information processing systems, pp. 5998-6008. 2017.*

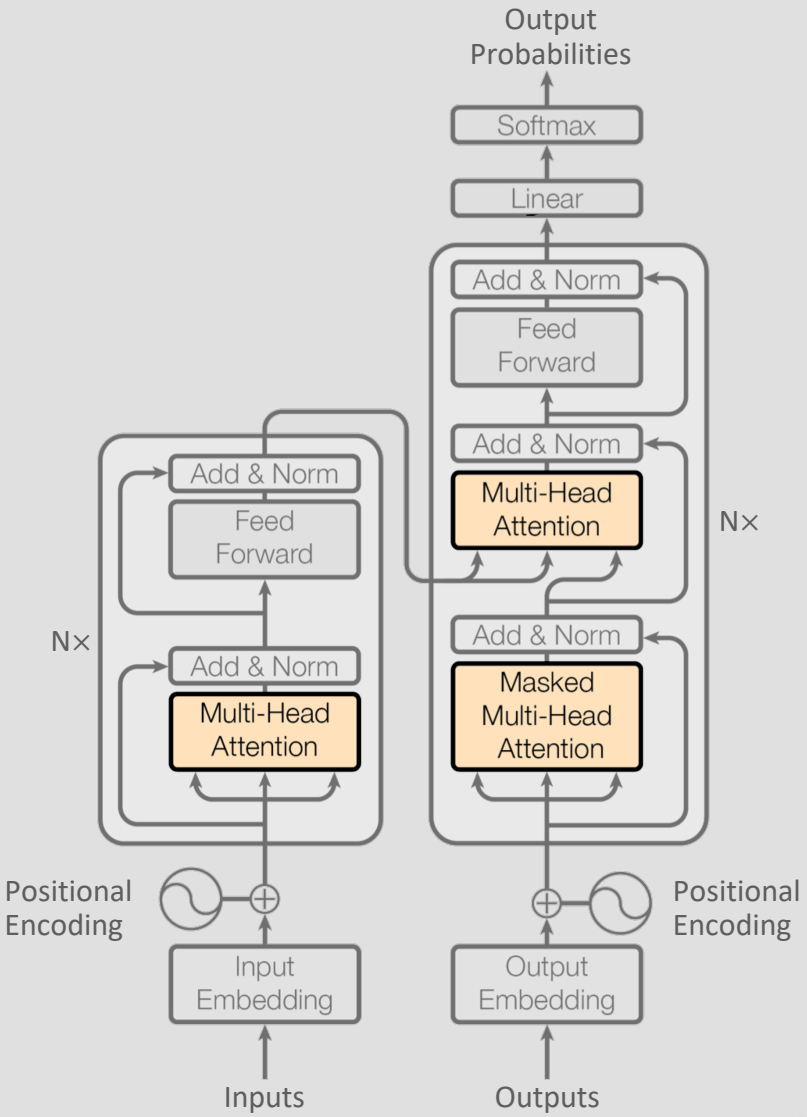
Модель Трансформер

- Кодировщик-декодировщик архитектура без рекуррентных сетей



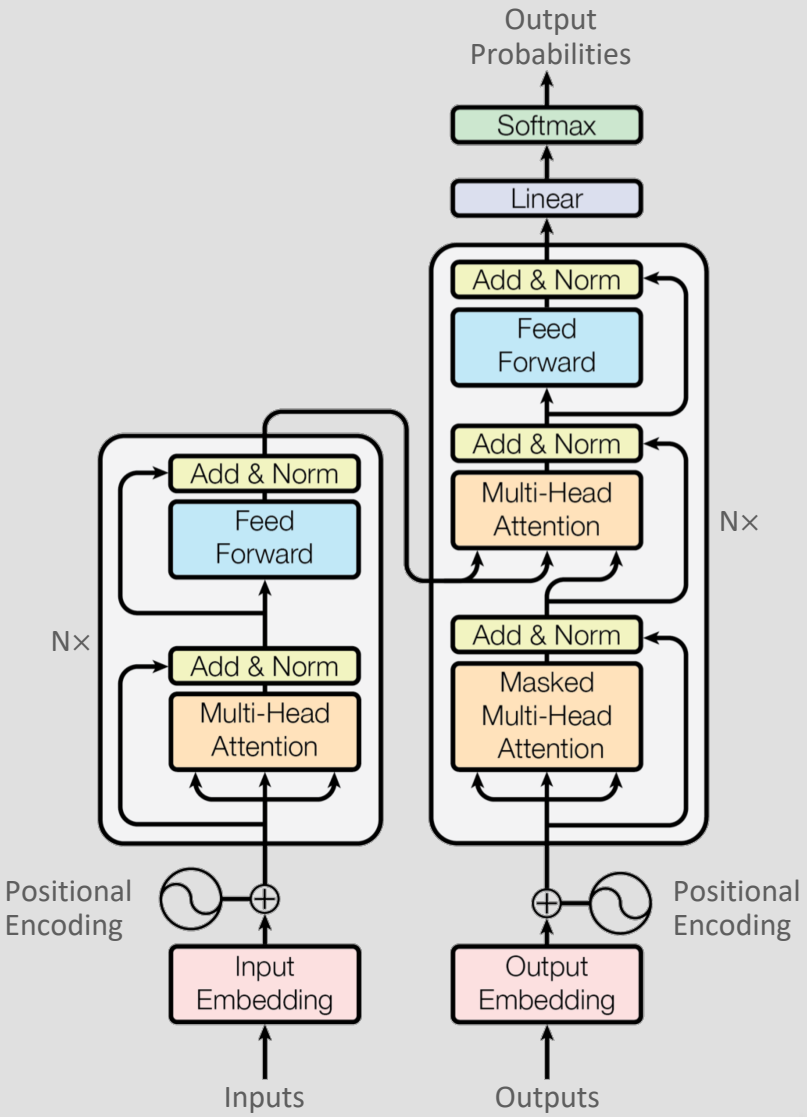
Модель Трансформер

- Кодировщик-декодировщик архитектура без рекуррентных сетей
- Новый слой: multi-head self-attention
- Превосходящие результаты как по скорости обучения, так и по качеству перевода



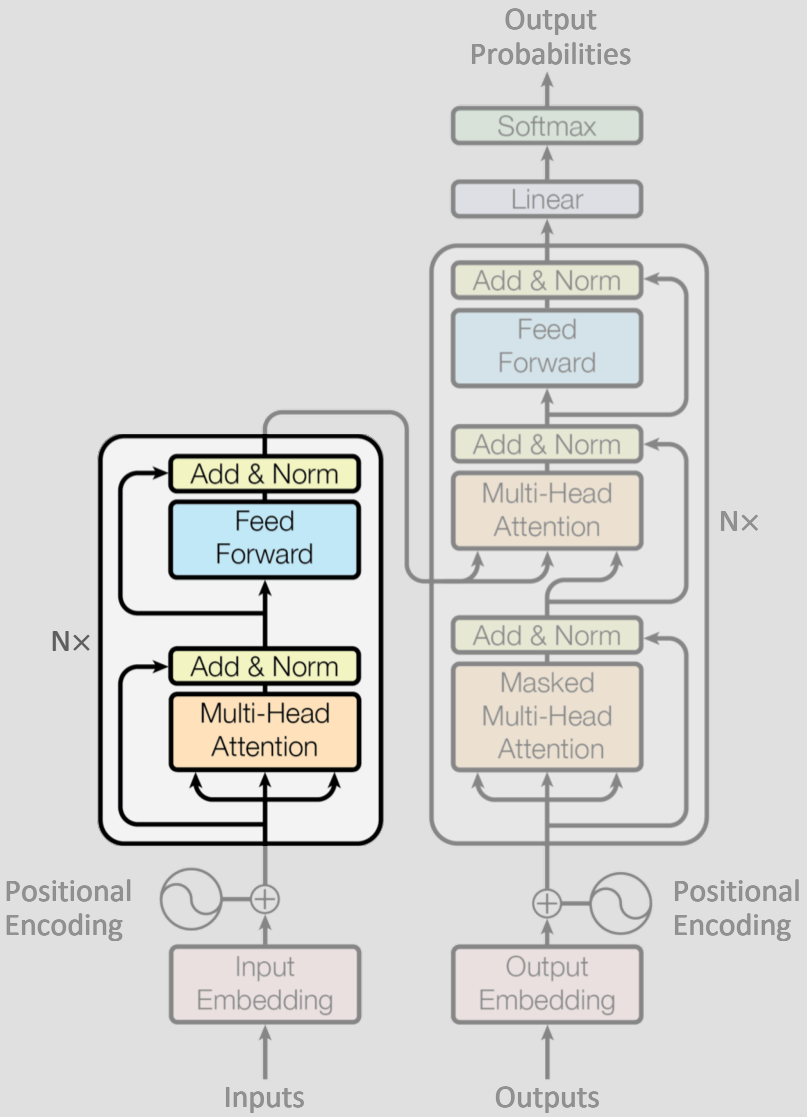
Модель Трансформер

- Кодировщик и декодировщик состоят из шести одинаковых блоков
- Блоки организованы последовательно
- В каждом блоке собственные веса



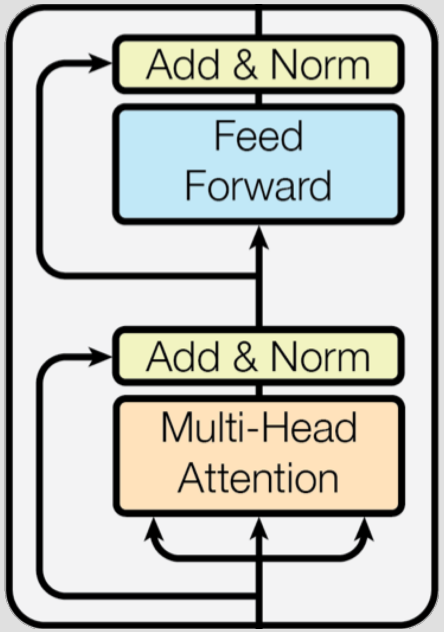
Кодировщик

- На вход первого блока подаются вектора слов, на вход остальных блоков – выход предшествующих блоков
- Каждый блок кодировщика состоит из двух частей: механизма self-attention и полносвязного слоя



Кодировщик

- Рассмотрим механизм self-attention: каждый входной вектор преобразуем три раза:
  $q_i = W^Q x_i$ – запрос, query
  $k_i = W^K x_i$ – ключ, key
  $v_i = W^V x_i$ – значения, value
- Матрицы весов W^Q, W^K, W^V обучаются



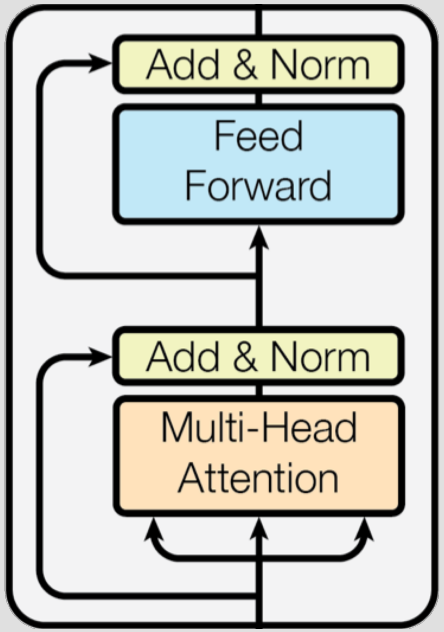
Кодировщик

- Механизм self-attention в матричном виде:

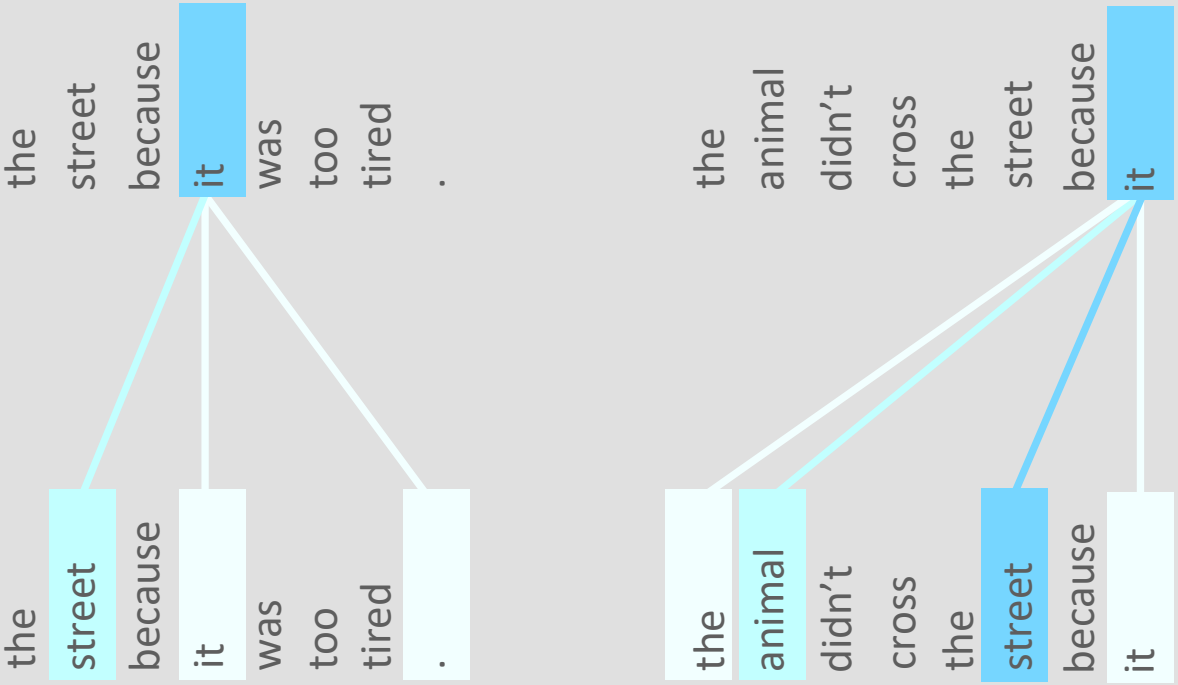
$$Z = softmax\frac{(QK^{\top})}{\sqrt{d_k}}V$$

- Будем параллельно вычислять несколько голов self-attention: $Z_1, \dots, Z_8$
- Конкатенируем вектора self-attention и получим промежуточное представление входной последовательности:

$$Z = W^O[Z_1, \dots, Z_8]$$



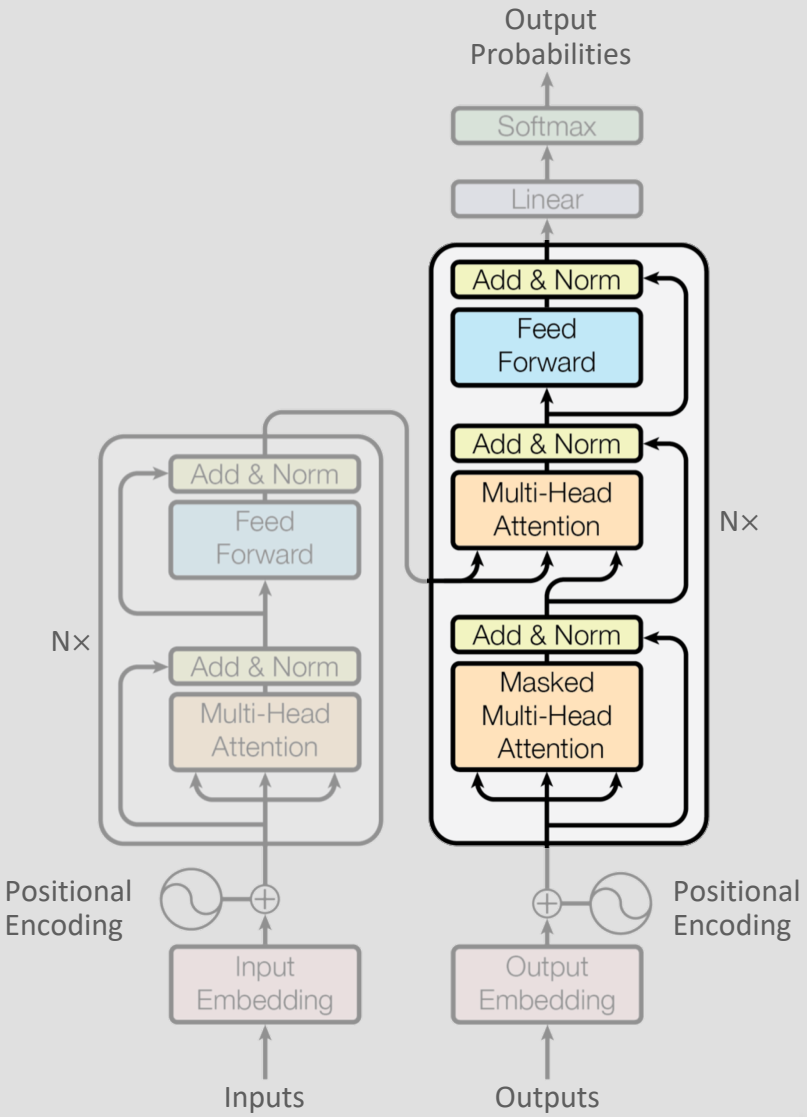
Веса в механизме внимания



ai.googleblog.com/2017/08/transformer-novel-neural-network.html

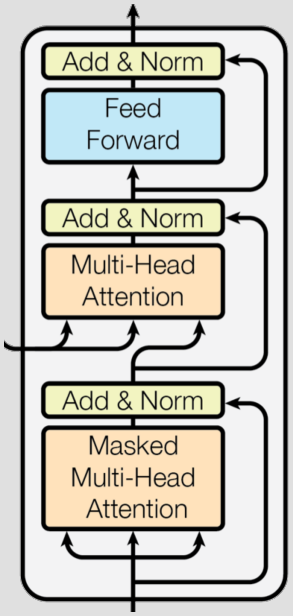
Декодировщик

- Кодировщик преобразует входные слова x в матрицы $Q^{enc}, K^{enc}, V^{enc}$
- Кодировщик передает матрицы K^{enc}, V^{enc} декодировщику
- Декодировщик преобразует входные слова y в матрицу Q^{dec}

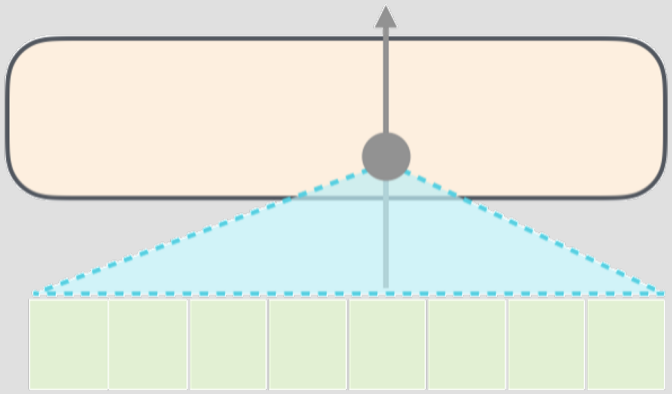


Декодировщик

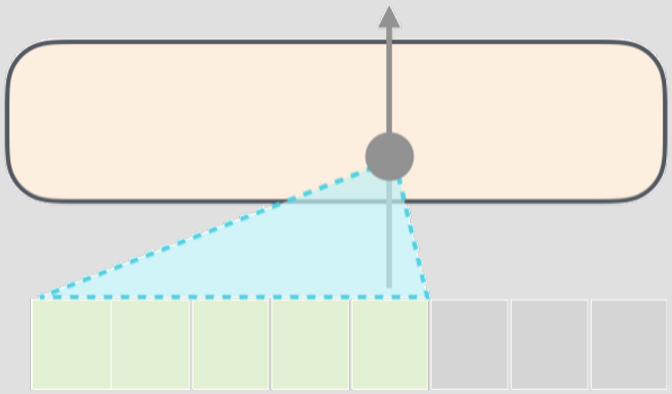
- Декодировщик последовательно переводит слова
- Механизм masked attention в декодировщике позволяет ограничить доступ к словам, идущим после текущего, считая их веса равными $-\infty$
- Механизм masked attention используется только на этапе обучения



Механизм внимания в модели Трансформер



Механизм внимания
в кодировщике



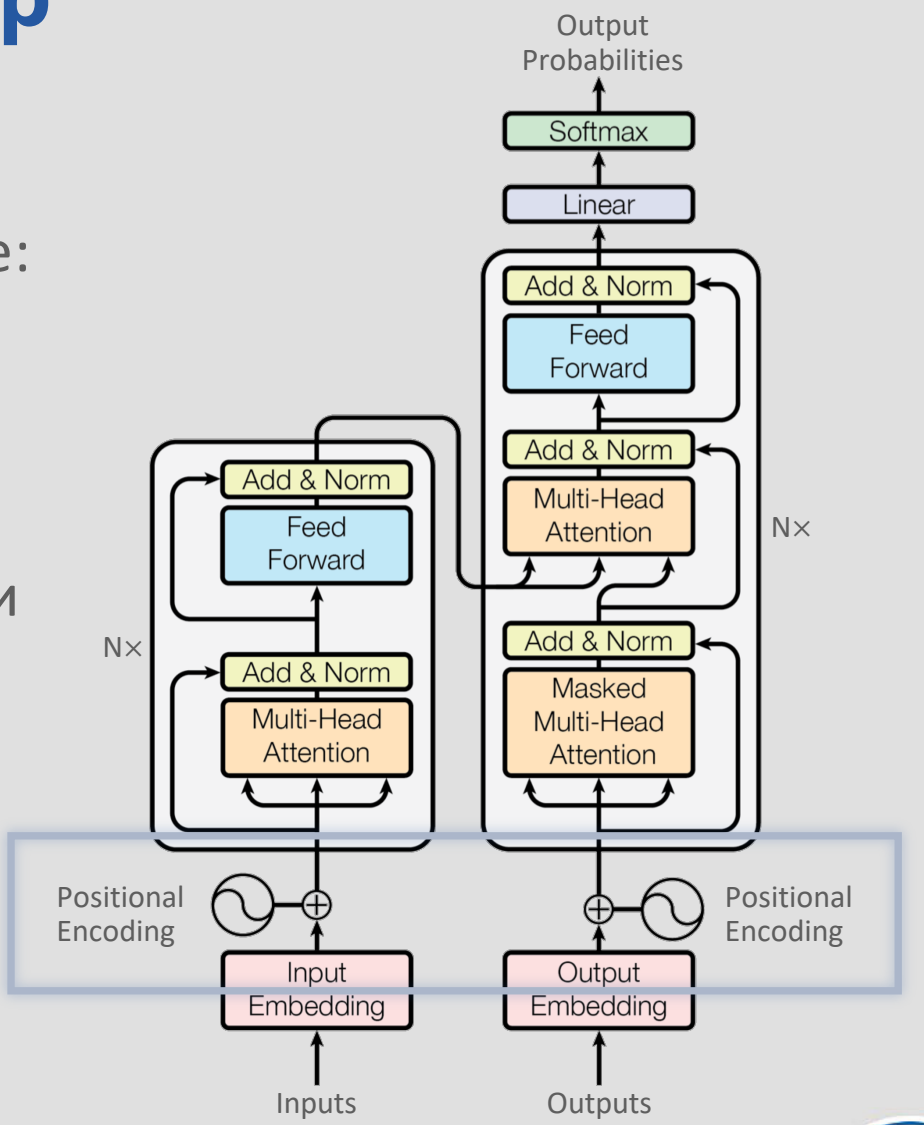
Механизм внимания
в декодировщике

Модель Трансформер

- Как учесть порядок слов?
- Позиционное кодирование:

$$x_i = x_i + pe_i$$

- Передает в вектора слов информацию о взаимном расстоянии между словами

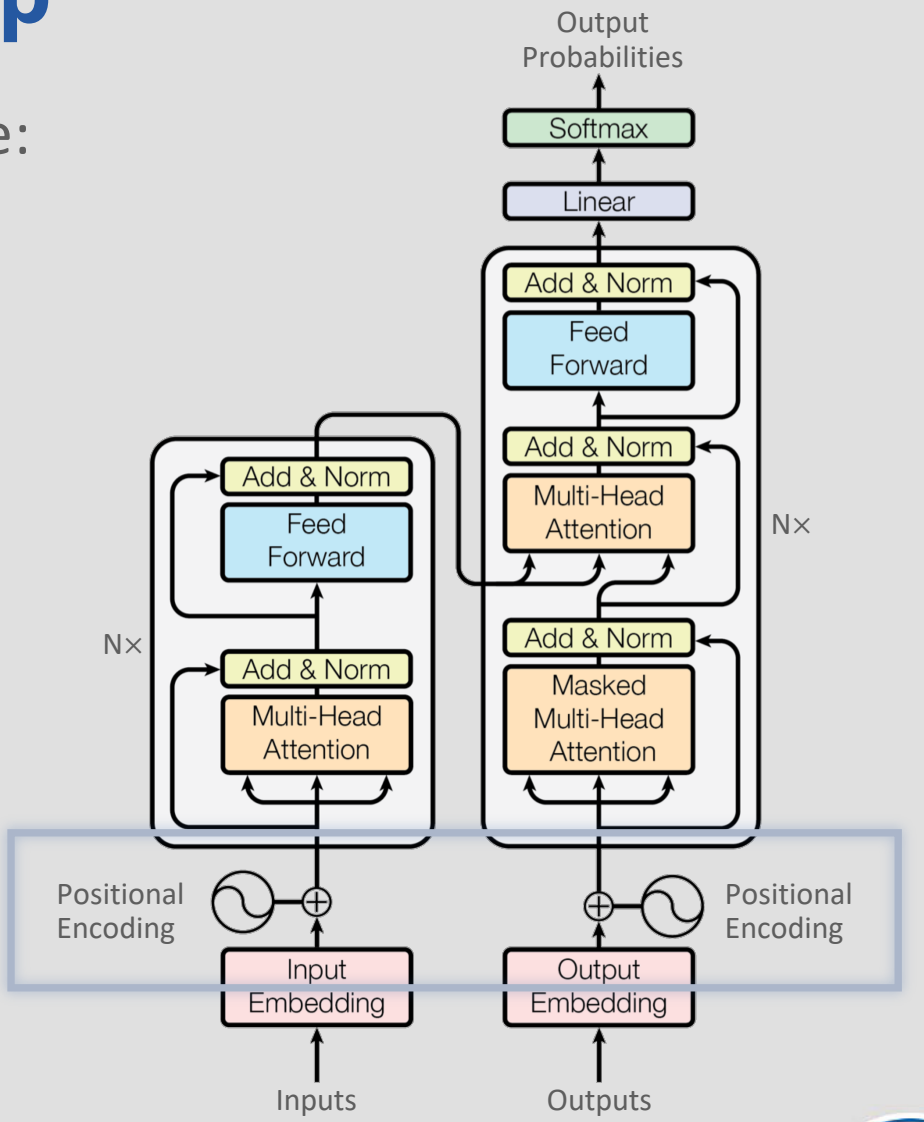


Модель Трансформер

- Позиционное кодирование:

$$x_i = x_i + pe_i$$

- **Обучаемое** позиционное кодирование обучается одновременно с основной архитектурой

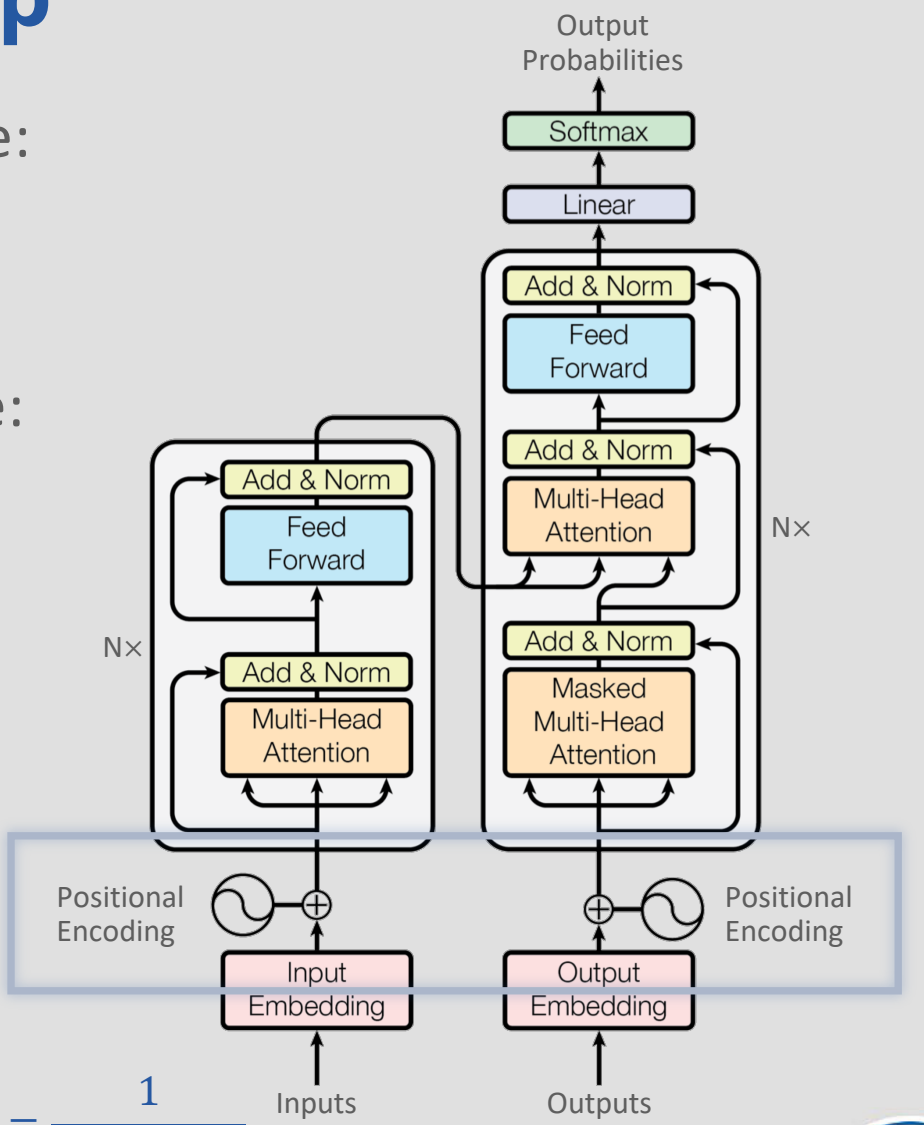


Модель Трансформер

- Позиционное кодирование:

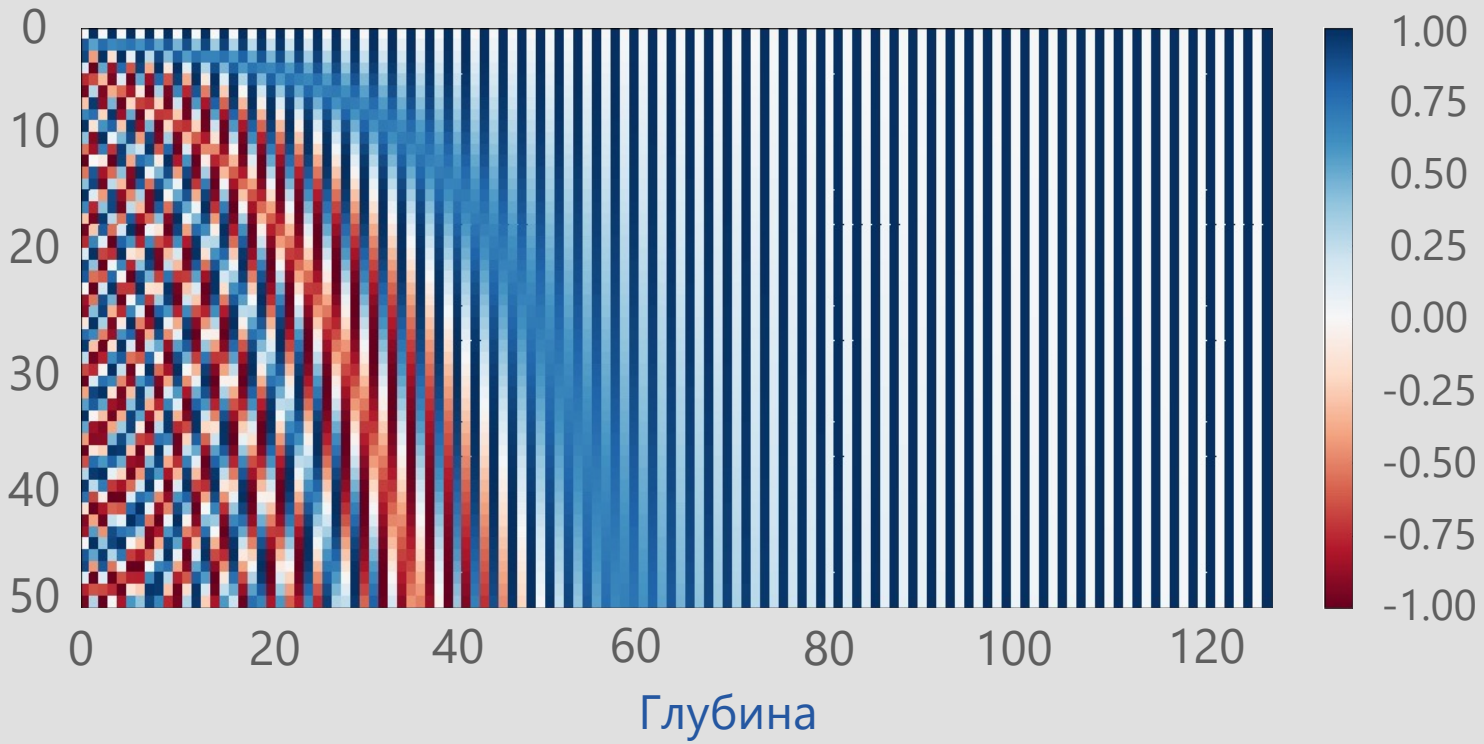
$$x_i = x_i + pe_i$$

- **Тригонометрическое** позиционное кодирование: заранее подсчитанные векторы, состоящие из набора синусов и косинусов с разными периодами от позиции слова в предложении



$$pe_t^{(i)} = \begin{cases} \sin(\omega_k, t), \text{если } i = 2k \\ \cos(\omega_k, t), \text{если } i = 2k+1 \end{cases}, \text{где } \omega_k = \frac{1}{10000^{2k/d}}$$

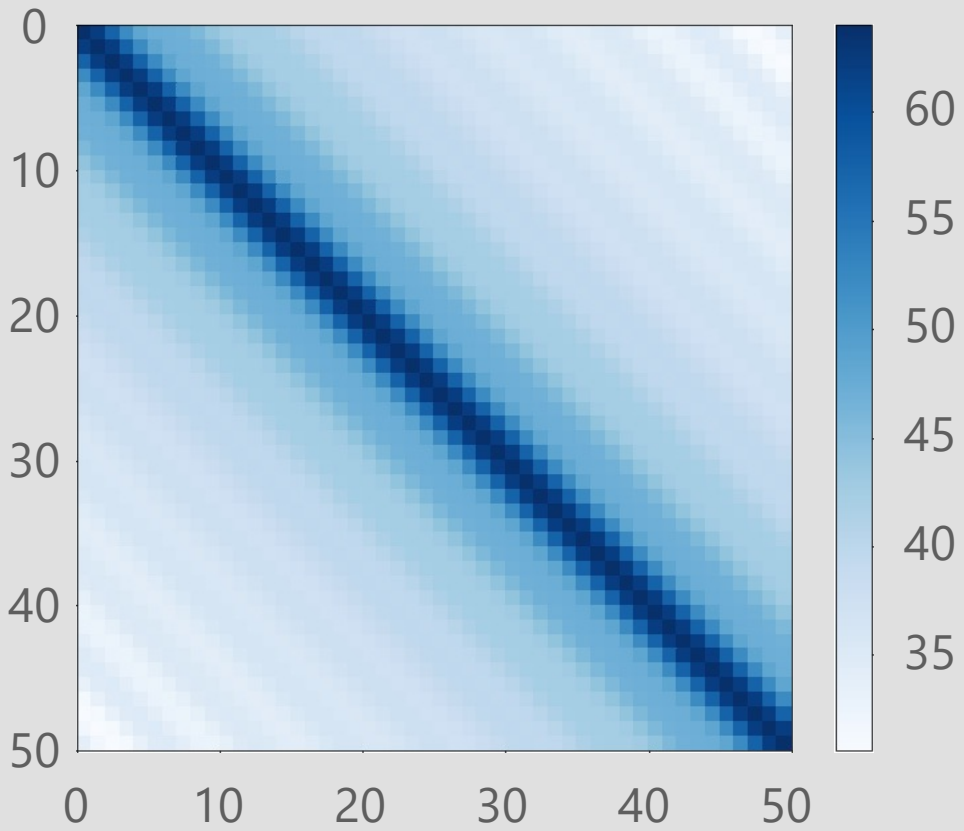
Тригонометрические позициональные эмбеддинги



128-размерные позициональные вектора для последовательности длины 50

➤ *https://kazemnejad.com/blog/transformer_architecture_positional_encoding/*

Тригонометрические позициональные эмбеддинги

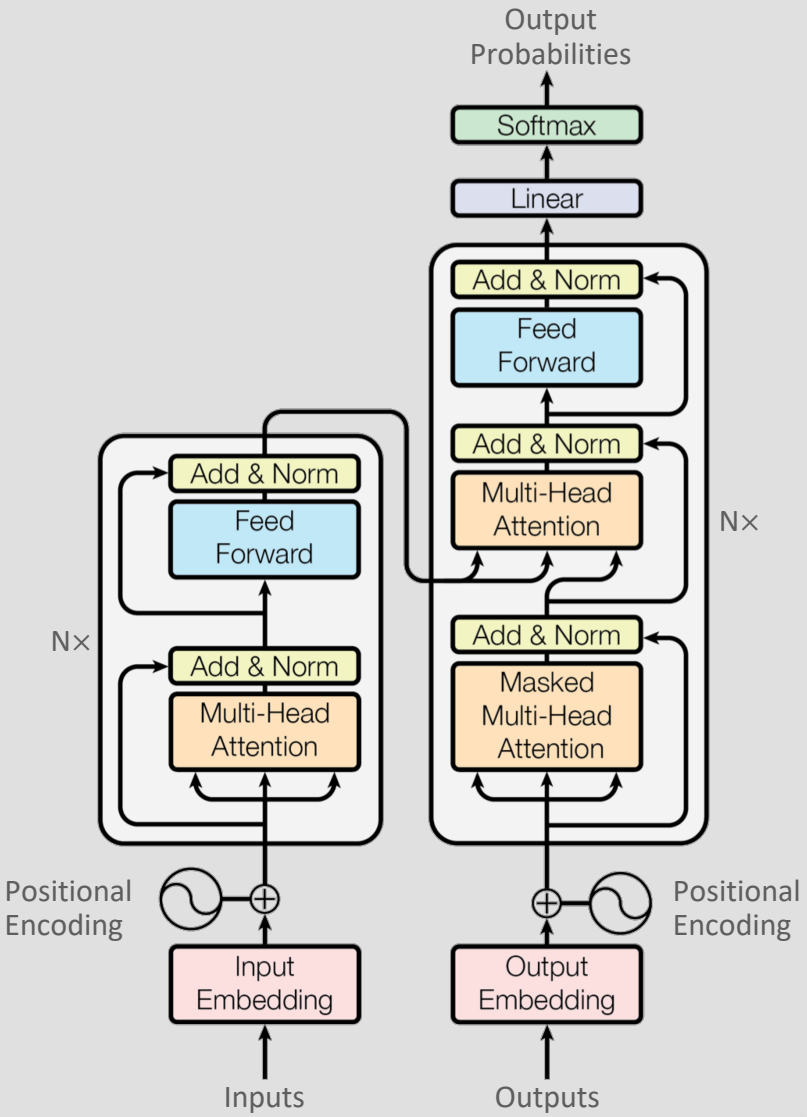


Расстояние между позициональными векторами

➤ *https://kazemnejad.com/blog/transformer_architecture_positional_encoding/*

Модель трансформер

- Доминирующая архитектура с середины 2018 года в большинстве задач автоматической обработки текстов
- По скорости и качеству превосходит существовавшие до нее архитектуры



Метрики качества в машинном переводе

Экспертная оценка

- Покажем предложение на исходном языке и перевод экспертам и попросим оценить перевод по шкале, например, от 0 до 5

Экспертная оценка

- Покажем предложение на исходном языке и перевод экспертам и попросим оценить перевод по шкале, например, от 0 до 5

Плюсы

- точная оценка

Минусы

- дорогая
- медленная
- трудоемкая оценка

Сравнение с золотым стандартом

- Система машинного перевода обучается по параллельному корпусу: предложения на исходном языке и несколько вариантов перевода на целевой язык

Сравнение с золотым стандартом

- Система машинного перевода обучается по параллельному корпусу: предложения на исходном языке и несколько вариантов перевода на целевой язык
- Разобьем параллельный корпус на обучающее и тестовое множество

Сравнение с золотым стандартом

- Система машинного перевода обучается по параллельному корпусу: предложения на исходном языке и несколько вариантов перевода на целевой язык
- Разобьем параллельный корпус на обучающее и тестовое множество
- Оценим качество системы машинного перевода на тестовом множестве

Сравнение с золотым стандартом

- Система машинного перевода обучается по параллельному корпусу: предложения на исходном языке и несколько вариантов перевода на целевой язык
- Разобьем параллельный корпус на обучающее и тестовое множество
- Оценим качество системы машинного перевода на тестовом множестве

Плюсы

- быстрая оценка (не учитываем затраты на создание корпуса)

Минусы

- корпусы, как правило, содержат тексты одного жанра, не содержат сленга и неологизмов

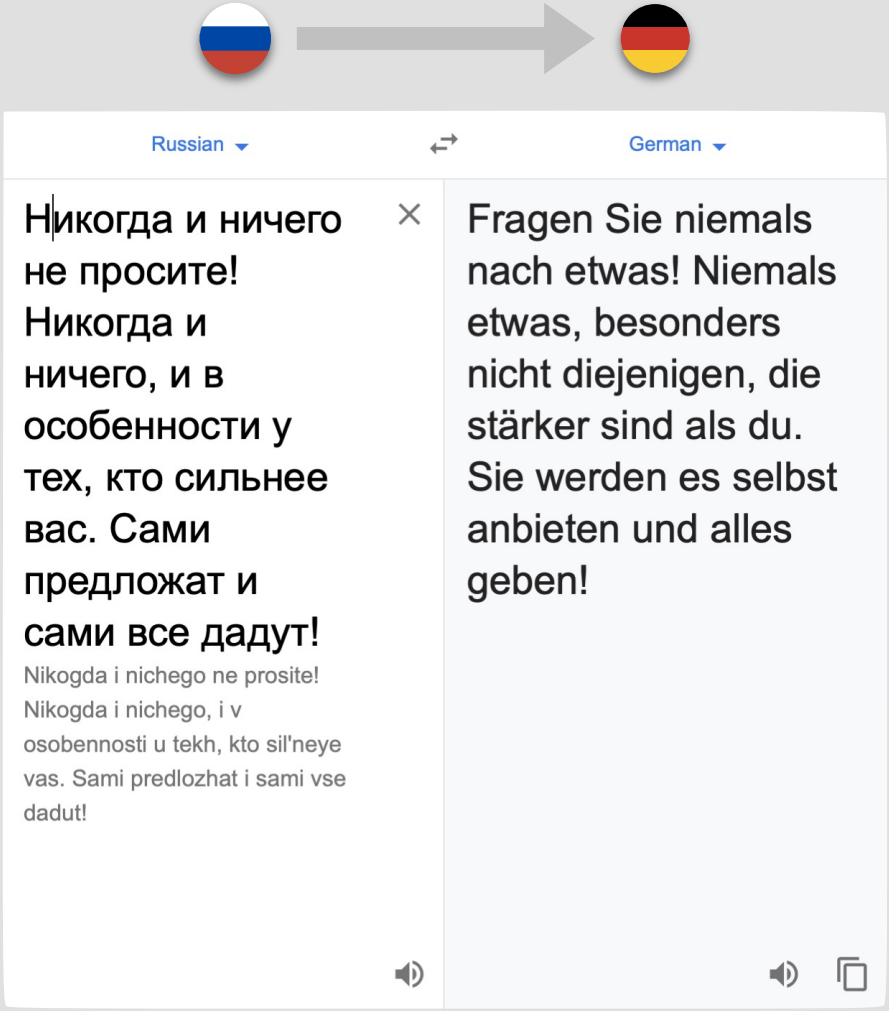
Перевод в обе стороны

- Используем машинный перевод для того, чтобы перевести текст на другой язык …

translate.google.com

Перевод в обе стороны

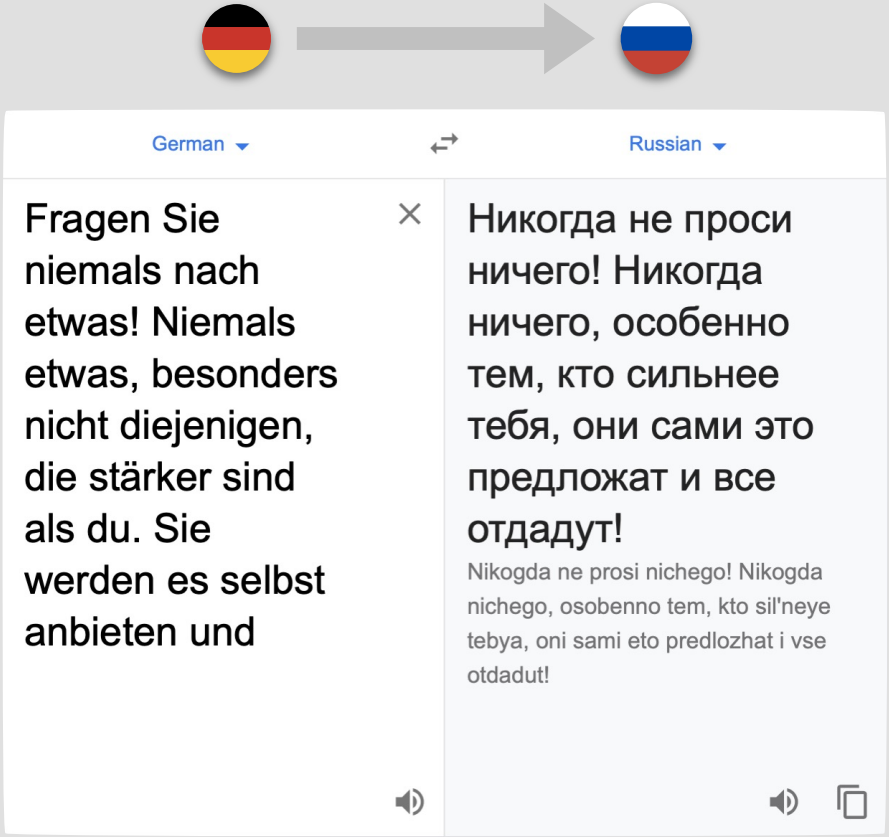
- Используем машинный перевод для того, чтобы перевести текст на другой язык …



translate.google.com

Перевод в обе стороны

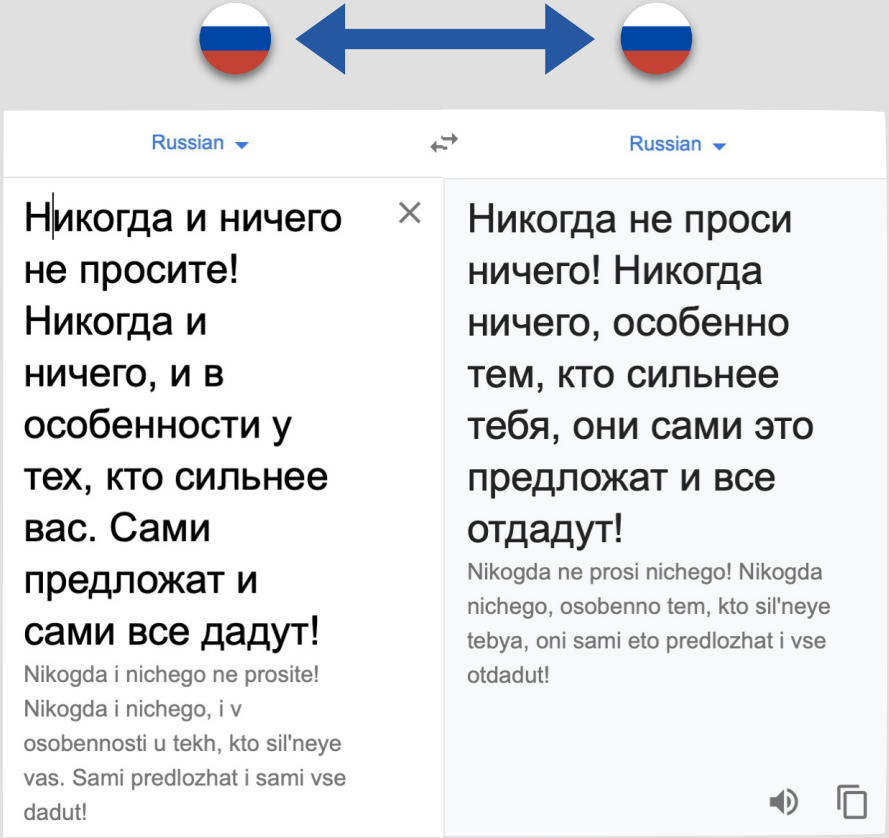
- Используем машинный перевод для того, чтобы перевести текст на другой язык
- … и обратно



translate.google.com

Перевод в обе стороны

- Используем машинный перевод для того, чтобы перевести текст на другой язык и обратно
- Сравним исходный текст и получившийся после перевода туда и обратно



translate.google.com

BLEU (bilingual evaluation understudy)

- Нужно уметь сравнивать предложения:
  - *Никогда и ничего не просите!*
  - *Никогда не проси ничего!*

translate.google.com

BLEU (bilingual evaluation understudy)

- Нужно уметь сравнивать предложения:
  - *Никогда и ничего не просите!*
  - *Никогда не проси ничего!*
- BLEU – алгоритм сравнения автоматически переведенного предложения с золотым стандартом

translate.google.com

BLEU (bilingual evaluation understudy)

- На вход подаётся перевод, сделанный машиной и перевод, который считается золотым стандартом

translate.google.com

BLEU (bilingual evaluation understudy)

- На вход подаётся перевод, сделанный машиной и перевод, который считается золотым стандартом
- Посчитаем для заданного N (обычно 3-4) точность, заданную как долю N-грамм в ответе системы, которые присутствуют в правильном ответе (1 к 1): $pr_i, i \in \{1,4\}$

translate.google.com

BLEU (bilingual evaluation understudy)

- На вход подаётся перевод, сделанный машиной и перевод, который считается золотым стандартом
- Посчитаем для заданного N (обычно 3-4) точность, заданную как долю N-грамм в ответе системы, которые присутствуют в правильном ответе (1 к 1): $pr_i, i \in \{1,4\}$
- Оценка точности перевода: $score = \sqrt[4]{\prod_{i=1}^{4} pr_i}$

translate.google.com

BLEU (bilingual evaluation understudy)

- На вход подаётся перевод, сделанный машиной и перевод, который считается золотым стандартом
- Посчитаем для заданного N (обычно 3-4) точность, заданную как долю N-грамм в ответе системы, которые присутствуют в правильном ответе (1 к 1): $pr_i, i \in \{1,4\}$
- Оценка точности перевода: $score = \sqrt[4]{\prod_{i=1}^{4} pr_i}$
- Введем штраф за краткость:

$$bp = min(1, \frac{mt-output-length}{gold-standard-length})$$

translate.google.com

BLEU (bilingual evaluation understudy)

- На вход подаётся перевод, сделанный машиной и перевод, который считается золотым стандартом
- Посчитаем для заданного N (обычно 3-4) точность, заданную как долю N-грамм в ответе системы, которые присутствуют в правильном ответе (1 к 1): $pr_i, i \in \{1,4\}$
- Оценка точности перевода: $score = \sqrt[4]{\prod_{i=1}^{4} pr_i}$
- Введем штраф за краткость:

$$bp = min(1, \frac{mt{-}output{-}length}{gold{-}standard{-}length})$$

- Итоговое значение: $BLEU = bp \times score$

translate.google.com

BLEU (bilingual evaluation understudy)

Цитата:

Remember that the most dangerous prison is the one in your head.

BLEU (bilingual evaluation understudy)

Цитата:

Remember that the most dangerous prison is the one in your head.

Золотой стандарт:

Помни, что самая опасная тюрьма – в твоей голове.

BLEU (bilingual evaluation understudy)

Цитата:

Remember that the most dangerous prison is the one in your head.

Золотой стандарт:

Помни, что самая опасная тюрьма – в твоей голове.

MT1:

Помните, что самая опасная тюрьма находится в вашей голове.

MT2:

Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

BLEU (bilingual evaluation understudy)

Цитата:

Remember that the most dangerous prison is the one in your head.

Золотой стандарт:

Помни, что самая опасная тюрьма – в твоей голове.

MT1:

Помните, что самая опасная тюрьма находится в вашей голове.

MT2:

Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

BLEU (bilingual evaluation understudy)

Цитата:

Remember that the most dangerous prison is the one in your head.

Золотой стандарт:

Помни, что самая опасная тюрьма – в твоей голове.

MT1:

Помните, что самая опасная тюрьма находится в вашей голове.

MT2:

Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
Помните, что самая опасная тюрьма находится в вашей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
Точность, n=1		
Точность, n=2		
Точность, n=3		
Точность, n=4		
Штраф за краткость		
BLEU		

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
Помните, что самая опасная тюрьма находится в вашей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
Точность, n=1	5/9	
Точность, n=2	3/8	
Точность, n=3	2/7	
Точность, n=4	1/6	
Штраф за краткость		
BLEU		

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
Помните, что самая опасная тюрьма находится в вашей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
Точность, n=1	5/9	
Точность, n=2	3/8	
Точность, n=3	2/7	
Точность, n=4	1/6	
Штраф за краткость	1	
BLEU		

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
Помните, что самая опасная тюрьма находится в вашей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
Точность, n=1	5/9	
Точность, n=2	3/8	
Точность, n=3	2/7	
Точность, n=4	1/6	
Штраф за краткость	1	
BLEU	0.3303	

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
Помните, что самая опасная тюрьма находится в вашей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
Точность, n=1	5/9	7/10
Точность, n=2	3/8	5/9
Точность, n=3	2/7	3/8
Точность, n=4	1/6	2/7
Штраф за краткость	1	1
BLEU	0.3303	0.6389

WER (word error rate)

- Минимальное количество операций, нужное для того, чтобы получить из перевода, сделанного машиной, золотой стандарт
- Операции: замена, удаление, добавление слова
- Итоговое значение:

$$WER = \frac{\#\,вставок + \#\,замен + \#\,удалений}{mt - output - length}$$

translate.google.com

BLEU (bilingual evaluation understudy)

Цитата:

Remember that the most dangerous prison is the one in your head.

Золотой стандарт:

Помни, что самая опасная тюрьма – в твоей голове.

MT1:

Помните, что самая опасная тюрьма находится в вашей голове.

MT2:

Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
Помните, что самая опасная тюрьма находится в вашей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
# вставок		
# замен		
# удалений		
WER		

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
~~Помните~~ помни, что самая опасная тюрьма находится в вашей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
# вставок		
# замен	1	
# удалений		
WER		

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
~~Помните~~ помни, что самая опасная тюрьма находится в вашей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
# вставок		
# замен	1	
# удалений		
WER		

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
~~Помните~~ помни, что самая опасная тюрьма ~~находится~~ в вашей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
# вставок		
# замен	1	
# удалений	1	
WER		

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
~~Помните~~ помни, что самая опасная тюрьма ~~находится~~ в ~~вашей~~ твоей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
# вставок		
# замен	2	
# удалений	1	
WER		

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
~~Помните~~ помни, что самая опасная тюрьма ~~находится~~ в ~~вашей~~ твоей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
# вставок	0	
# замен	2	
# удалений	1	
WER		

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
~~Помните~~ помни, что самая опасная тюрьма ~~находится~~ в ~~вашей~~ твоей голове.

MT2:
Помни, что самая опасная тюрьма – это тюрьма в твоей голове.

	MT1	MT2
# вставок	0	
# замен	2	
# удалений	1	
WER	0.33	

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
~~Помните~~ помни, что самая опасная тюрьма ~~находится~~ в ~~вашей~~ твоей голове.

MT2:
Помни, что самая опасная тюрьма – ~~это тюрьма~~ в твоей голове.

	MT1	MT2
# вставок	0	
# замен	2	
# удалений	1	
WER	0.33	

BLEU (bilingual evaluation understudy)

Цитата:
Remember that the most dangerous prison is the one in your head.

Золотой стандарт:
Помни, что самая опасная тюрьма – в твоей голове.

MT1:
~~Помните~~ помни, что самая опасная тюрьма ~~находится~~ в ~~вашей~~ твоей голове.

MT2:
Помни, что самая опасная тюрьма – ~~это тюрьма~~ в твоей голове.

	MT1	MT2
# вставок	0	0
# замен	2	0
# удалений	1	2
WER	0.33	0.2

Анализ WER и BLEU

Плюсы

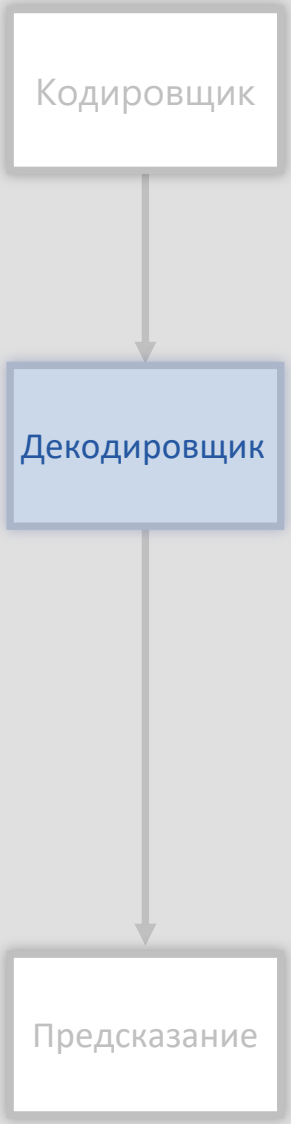
- Простые и быстрые оценки качества перевода (и не только!)
- Коррелируют с экспертной оценкой

Минусы

- Оперируют короткими фрагментами, не позволяет оценить общую корректность предложения
- Не позволяют оценить жанровую специфику
- Оценки носят относительный характер, сложно сравнивать между собой
- Важная техническая проблема: оценки не дифференцируемы

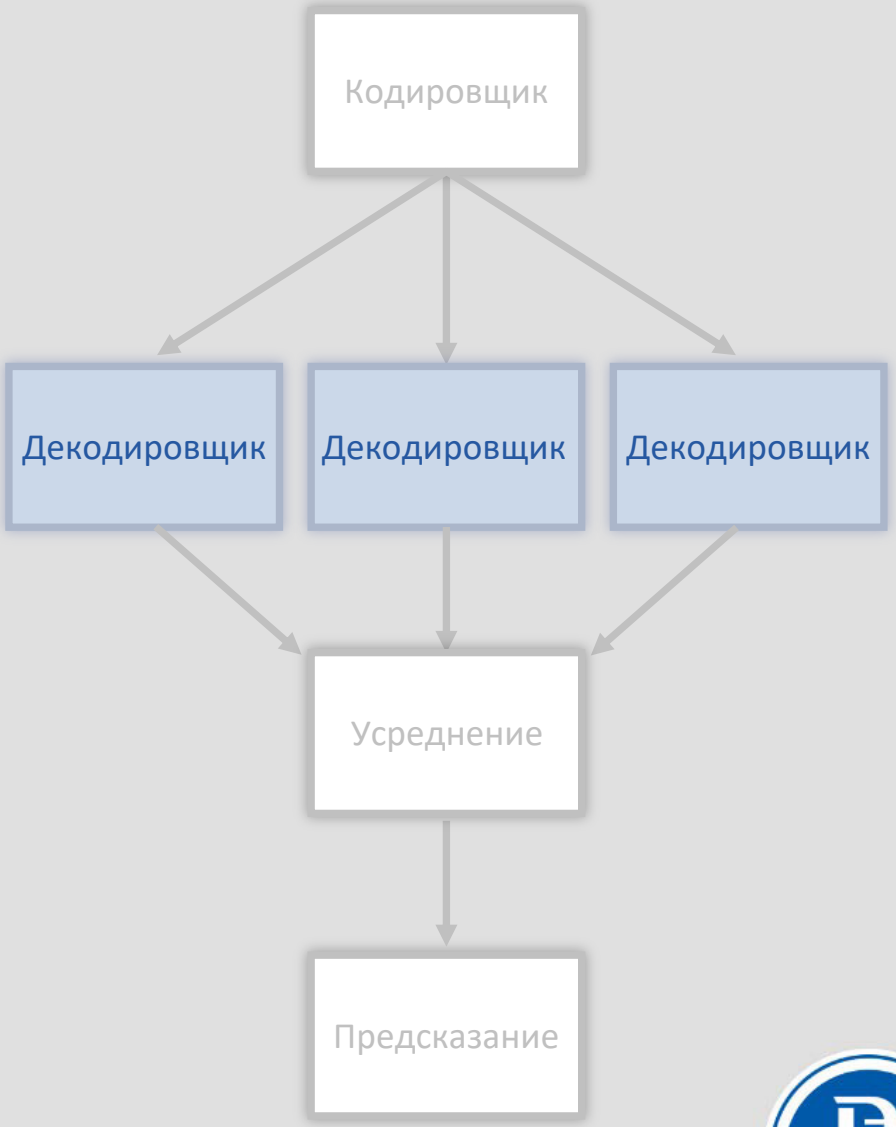
Улучшение качества машинного перевода

Ансамбль декодировщиков



Ансамбль декодировщиков

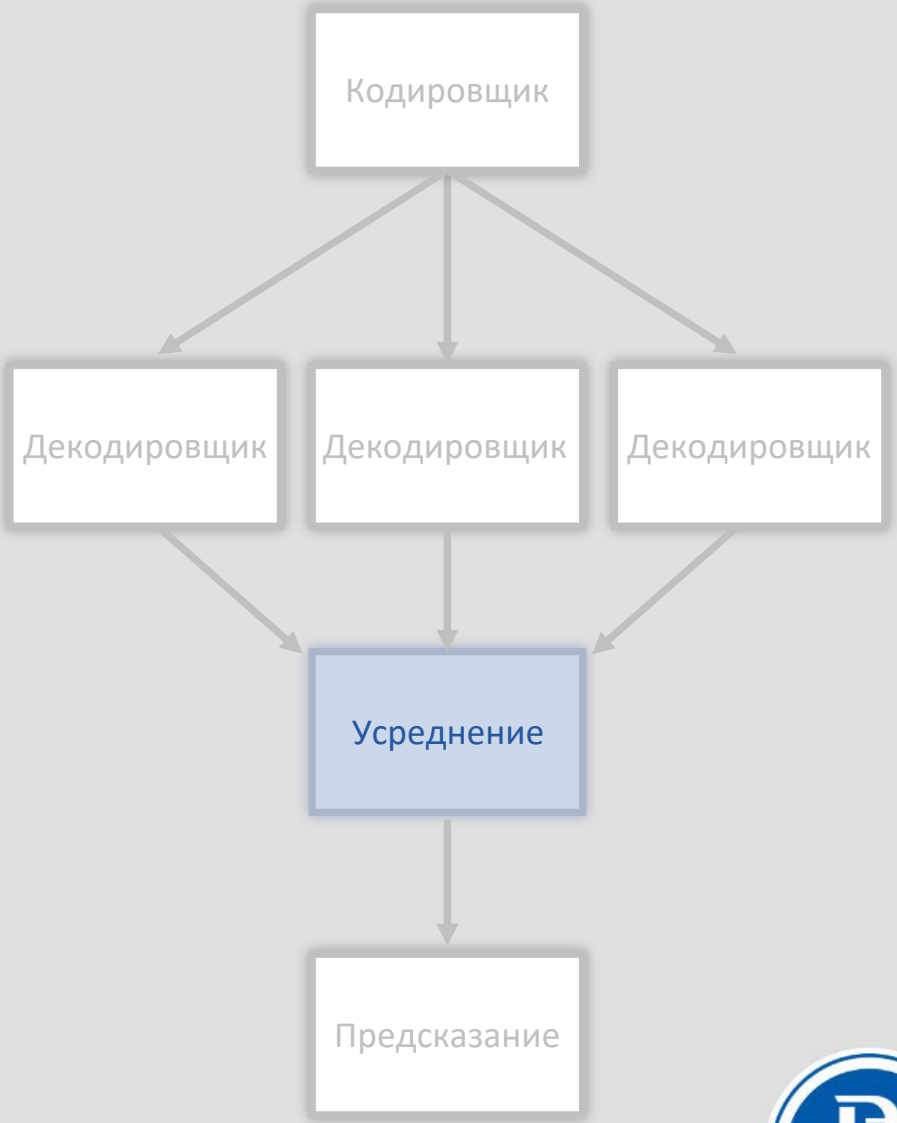
- Обучаем одновременно несколько декодировщиков



Ансамбль декодировщиков

- Обучаем одновременно несколько декодировщиков
- На этапе декодирования усредняем предсказания всех декодировщиков:

$$P(y_j|y_{j-1},\mathbf{x}) = \\ = \frac{1}{K}\sum_{k=1}^{K} P_k(y_j|y_{j-1},\mathbf{x})$$



Направление перевода

Исходный язык

The	cat	sat	on	the	mat

Направление перевода

Кодировщик

Слева – направо (прямой порядок слов)

The	cat	sat	on	the	mat

The

Направление перевода

Кодировщик

Слева – направо (прямой порядок слов)

The	cat	sat	on	the	mat

The cat

Направление перевода

Кодировщик

Слева – направо (прямой порядок слов)

The	cat	sat	on	the	mat

The cat sat

Направление перевода

Кодировщик

Слева – направо (прямой порядок слов)

The	cat	sat	on	the	mat

The cat sat on

Направление перевода

Кодировщик

Слева – направо (прямой порядок слов)

The	cat	sat	on	the	mat

The cat sat on the

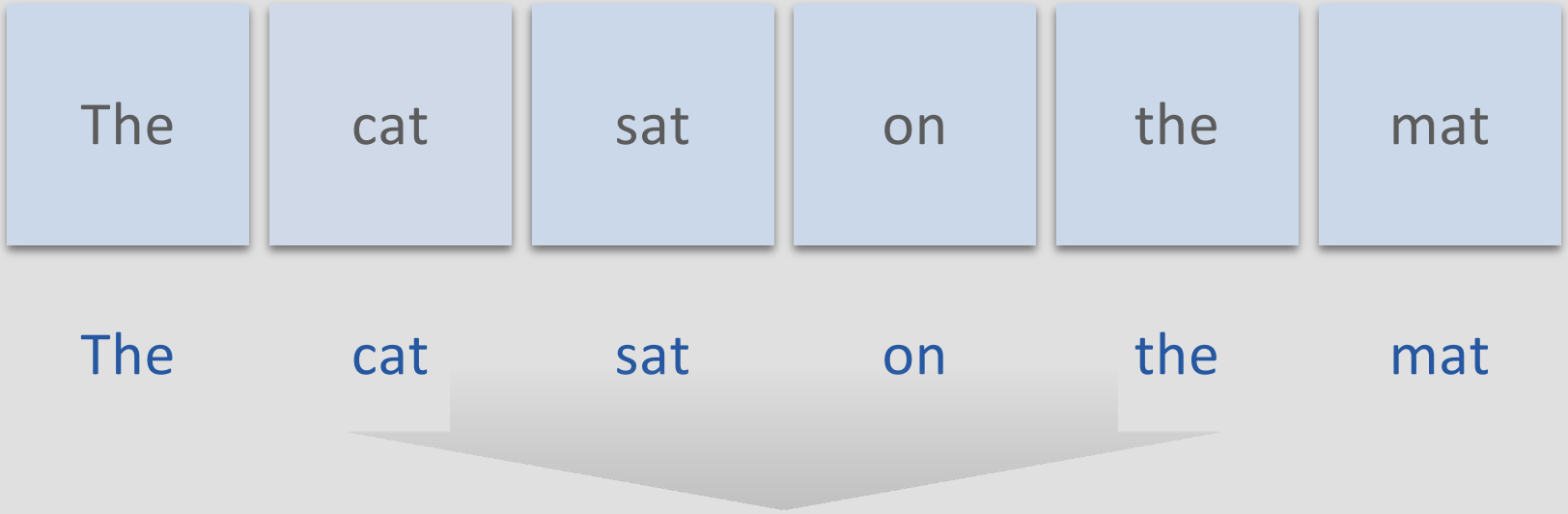
Направление перевода

Кодировщик

Слева – направо (прямой порядок слов)

The	cat	sat	on	the	mat
The	cat	sat	on	the	mat

Направление перевода



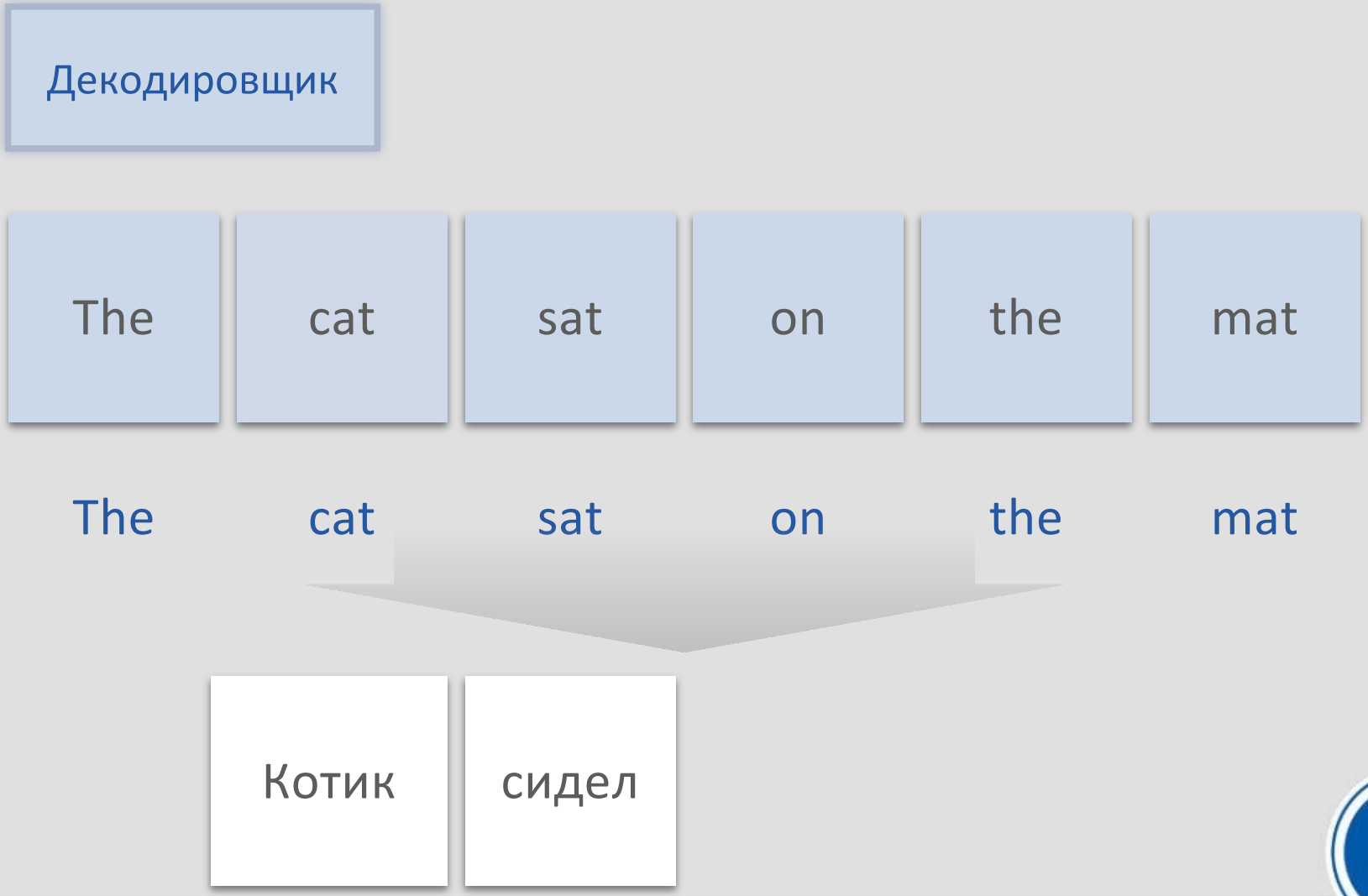
Направление перевода

Декодировщик

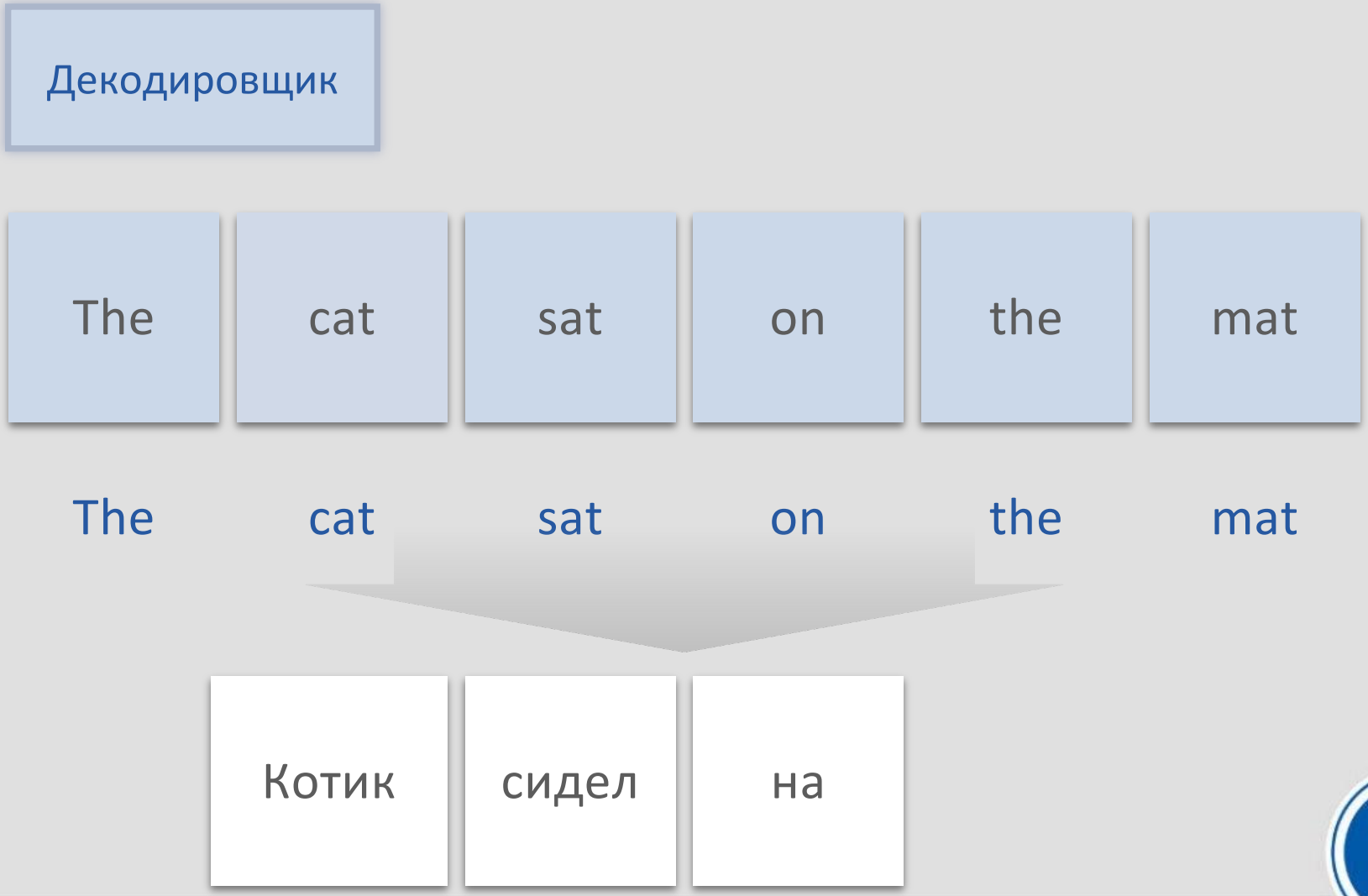
The	cat	sat	on	the	mat
The	cat	sat	on	the	mat

Котик

Направление перевода



Направление перевода



Направление перевода

Декодировщик

The	cat	sat	on	the	mat
The	cat	sat	on	the	mat

Котик	сидел	на	коврике

Направление перевода

Исходный язык

The	cat	sat	on	the	mat

Направление перевода

Кодировщик

Справа – налево (обратный порядок слов)

The	cat	sat	on	the	mat

mat

Направление перевода

Кодировщик

Справа – налево (обратный порядок слов)

The	cat	sat	on	the	mat

mat the

Направление перевода

Кодировщик

Справа – налево (обратный порядок слов)

The	cat	sat	on	the	mat

mat the on

Направление перевода

Кодировщик

Справа – налево (обратный порядок слов)

The	cat	sat	on	the	mat

mat the on sat

Направление перевода

Кодировщик

Справа – налево (обратный порядок слов)

The	cat	sat	on	the	mat

mat the on sat cat

Направление перевода

Кодировщик

Справа – налево (обратный порядок слов)

The	cat	sat	on	the	mat
mat	the	on	sat	cat	The

Направление перевода

Кодировщик

Справа – налево (обратный порядок слов)

The	cat	sat	on	the	mat
mat	the	on	sat	cat	The

Направление перевода

The	cat	sat	on	the	mat
mat	the	on	sat	cat	The

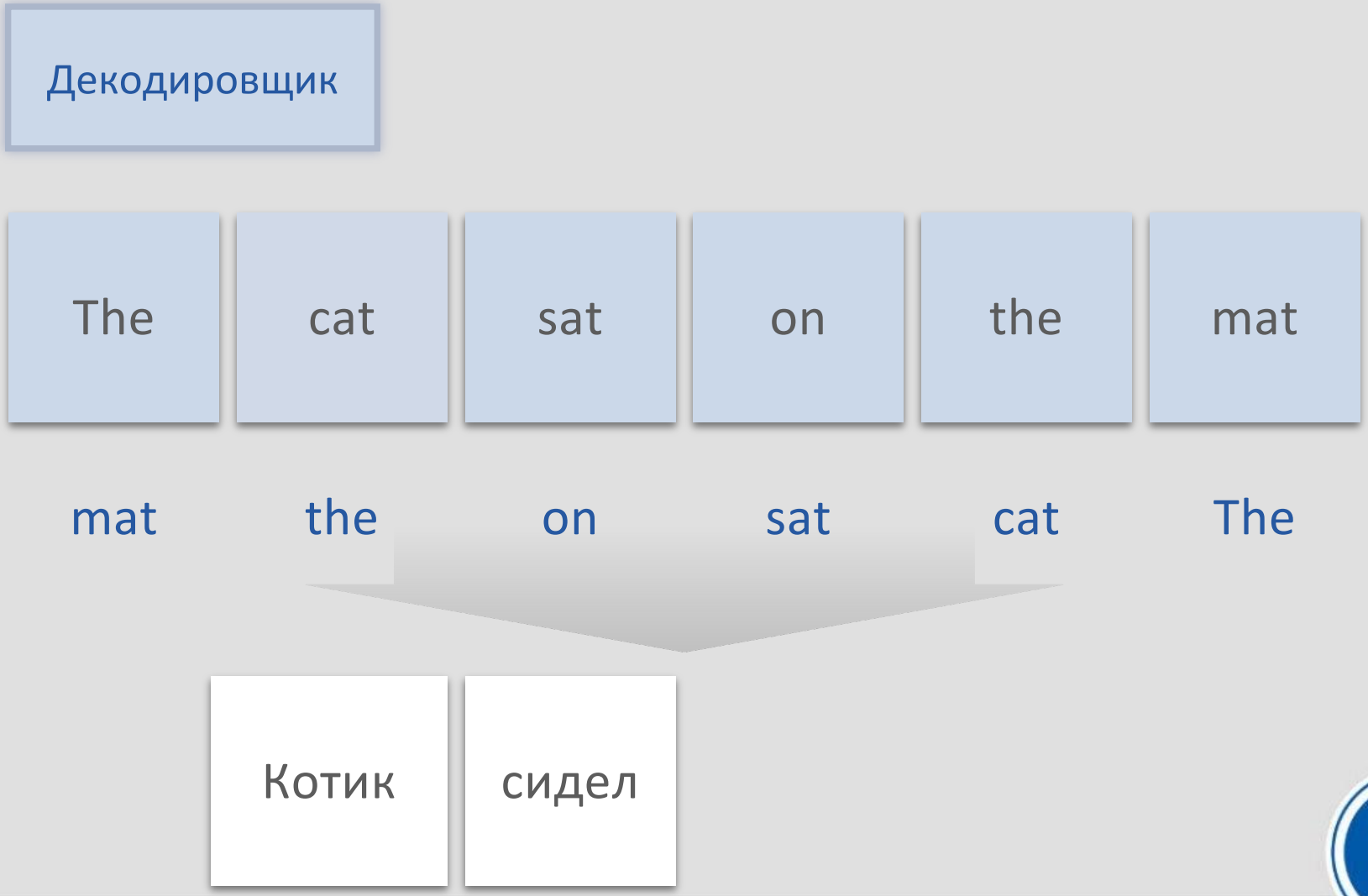
Направление перевода

Декодировщик

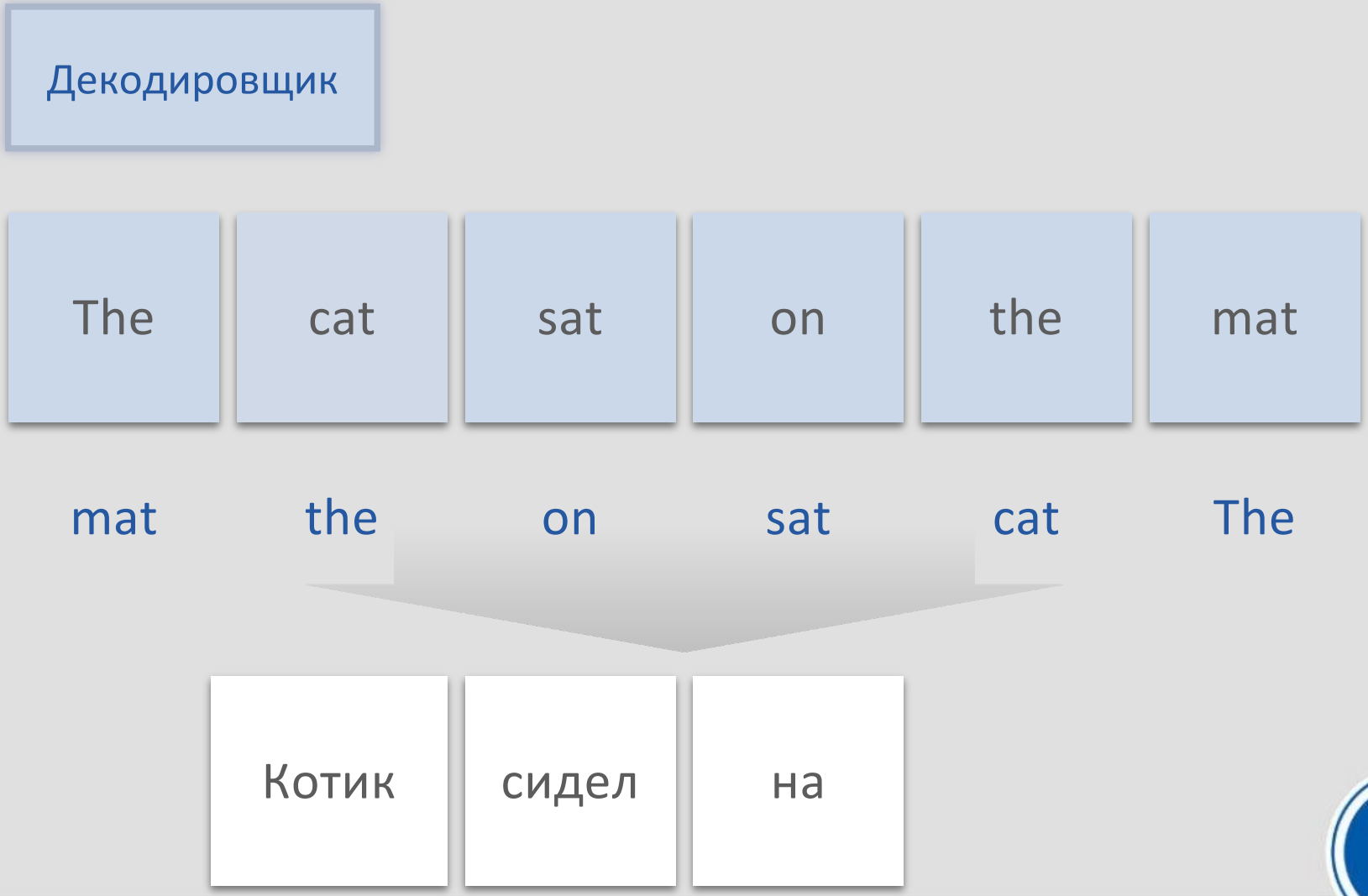
The	cat	sat	on	the	mat
mat	the	on	sat	cat	The

Котик

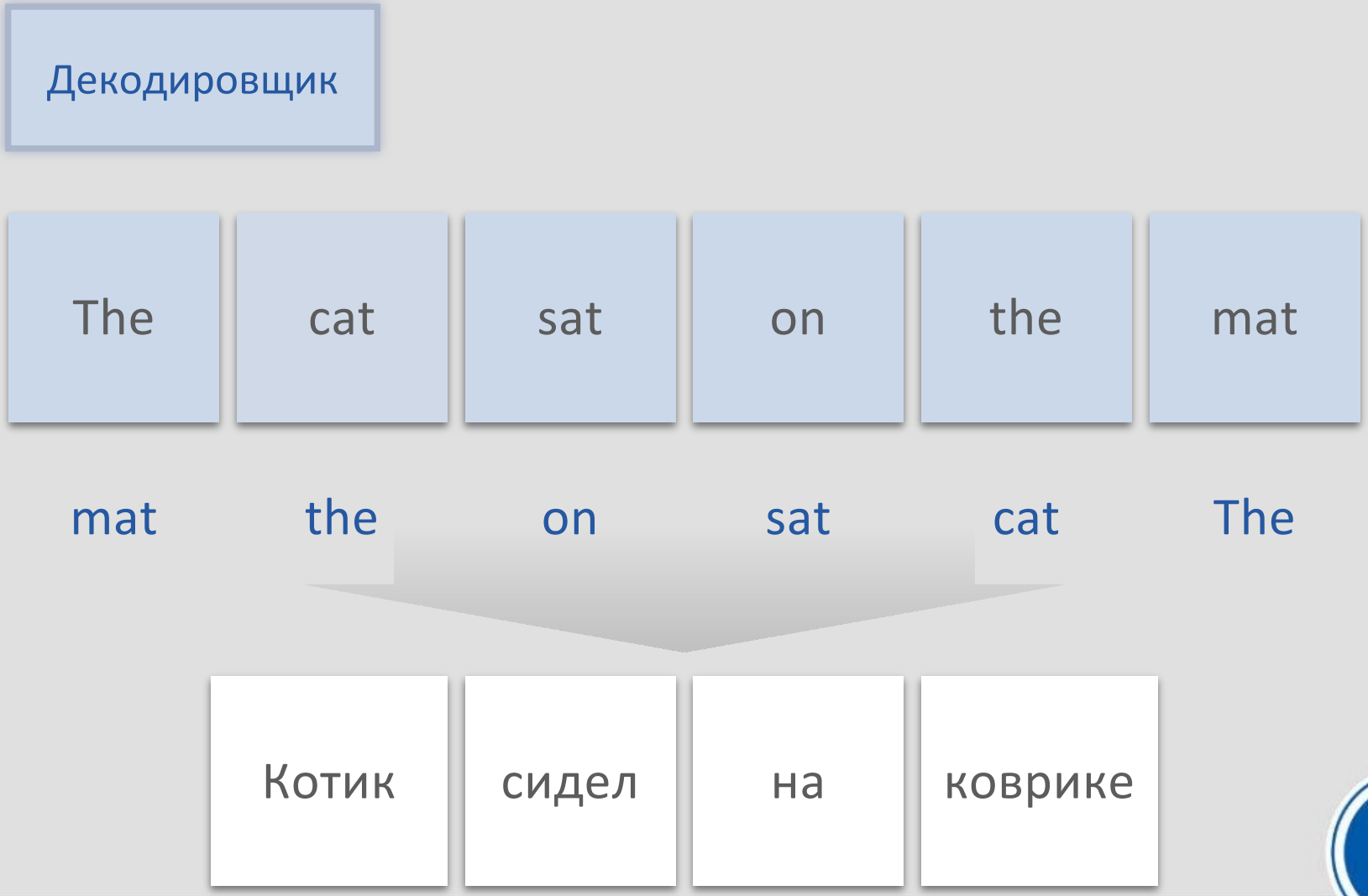
Направление перевода



Направление перевода



Направление перевода



Синтетические данные

Обратный перевод (back translation):

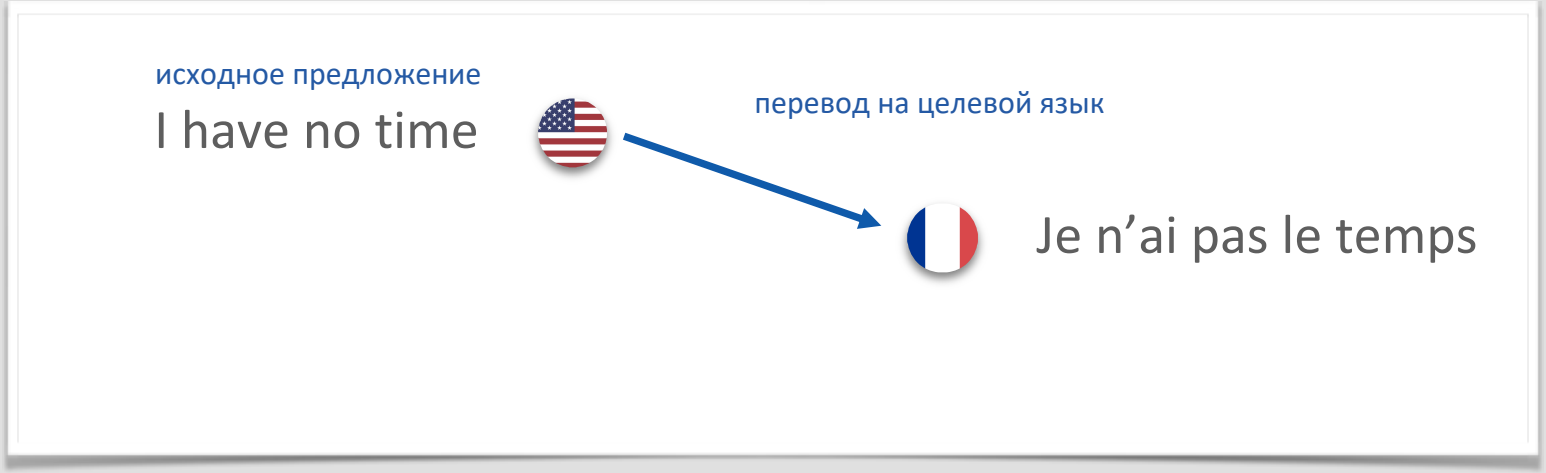
Синтетические данные

исходное предложение

I have no time

Обратный перевод (back translation):

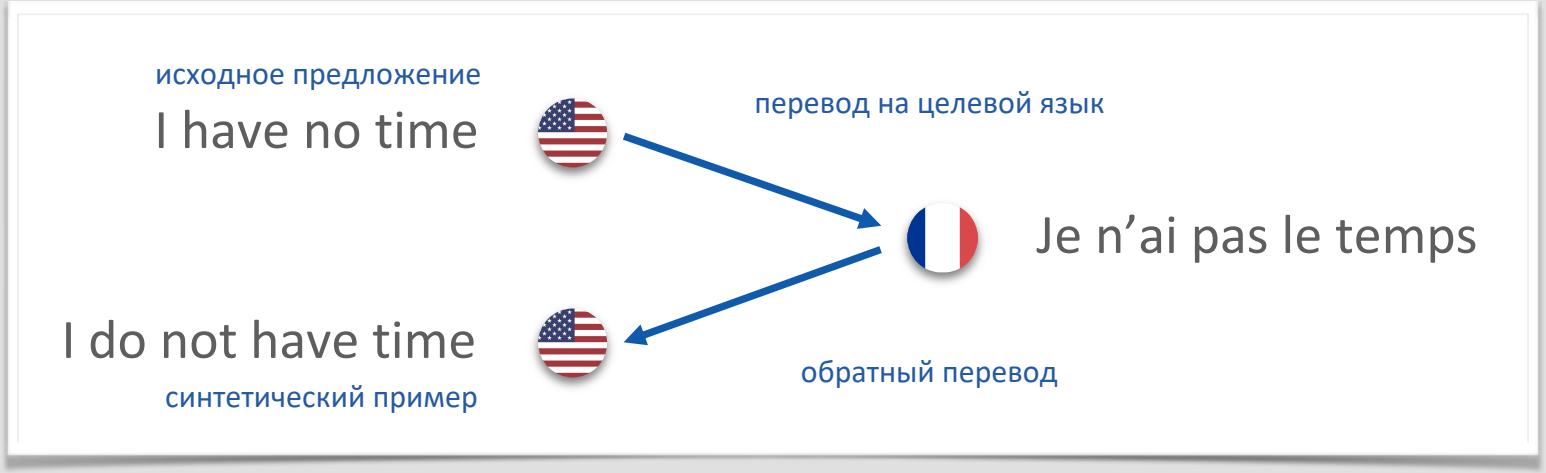
Синтетические данные



Обратный перевод (back translation):

- Переведем предложение с исходного языка на целевой

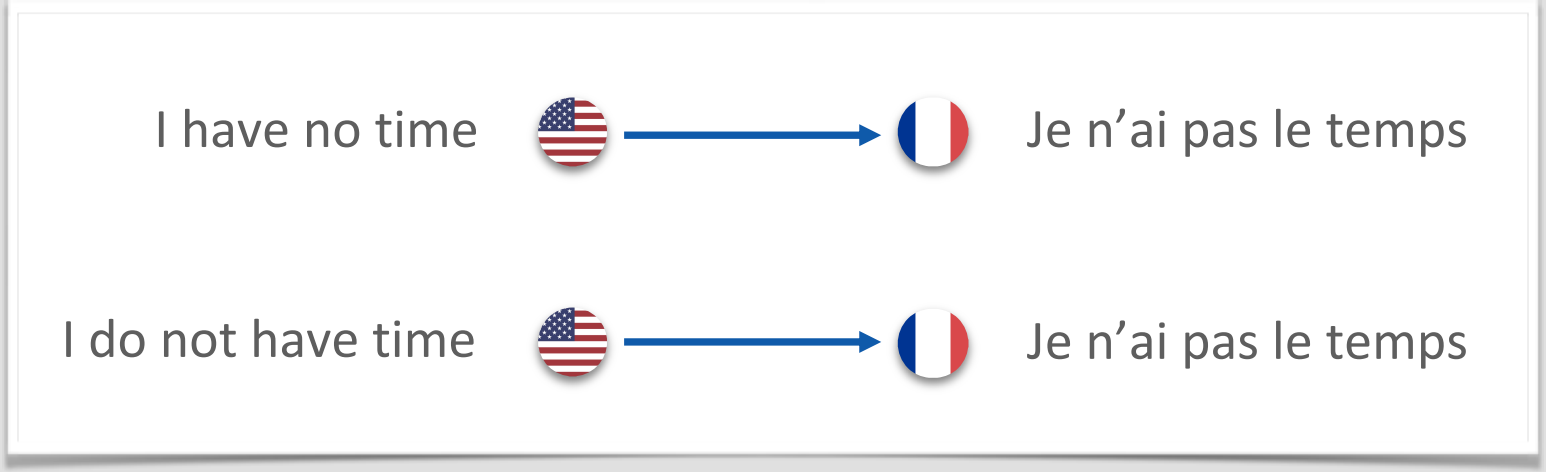
Синтетические данные



Обратный перевод (back translation):

- Переведем предложение с исходного языка на целевой
- Переведем обратно предложение с целевого языка на исходный

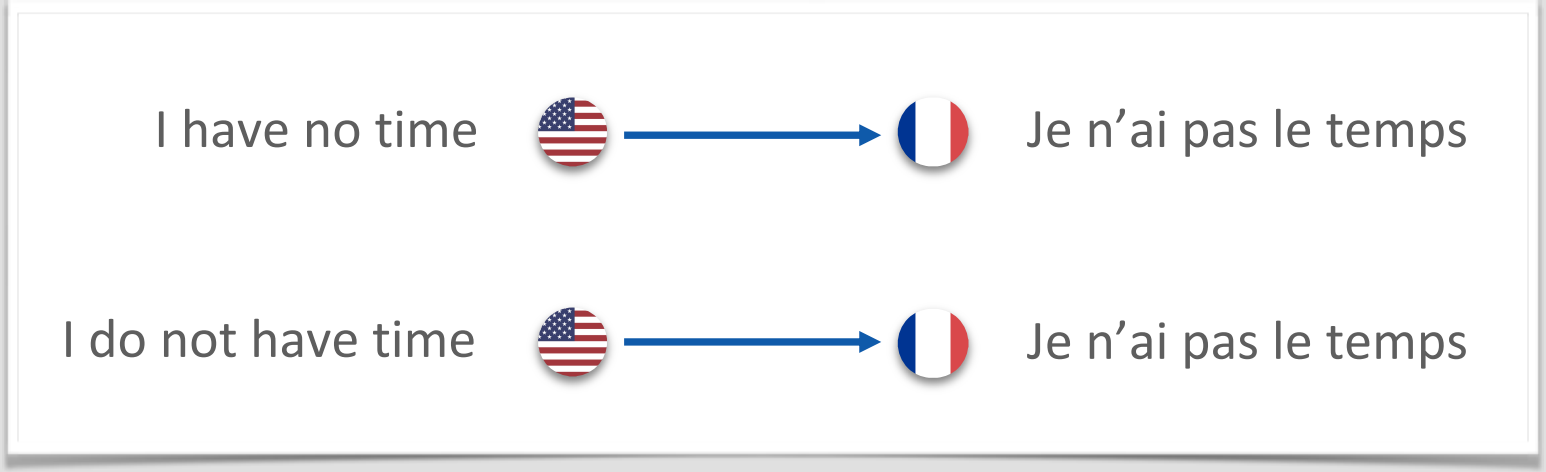
Синтетические данные



Обратный перевод (back translation):

- Переведем предложение с исходного языка на целевой
- Переведем обратно предложение с целевого языка на исходный
- Получим новую пару предложений для обучения

Синтетические данные



Обратный перевод (back translation):

- Переведем предложение с исходного языка на целевой
- Переведем обратно предложение с целевого языка на исходный
- Получим новую пару предложений для обучения

Нужно соблюдать баланс между реальными и синтетическими данными

Синхронный перевод

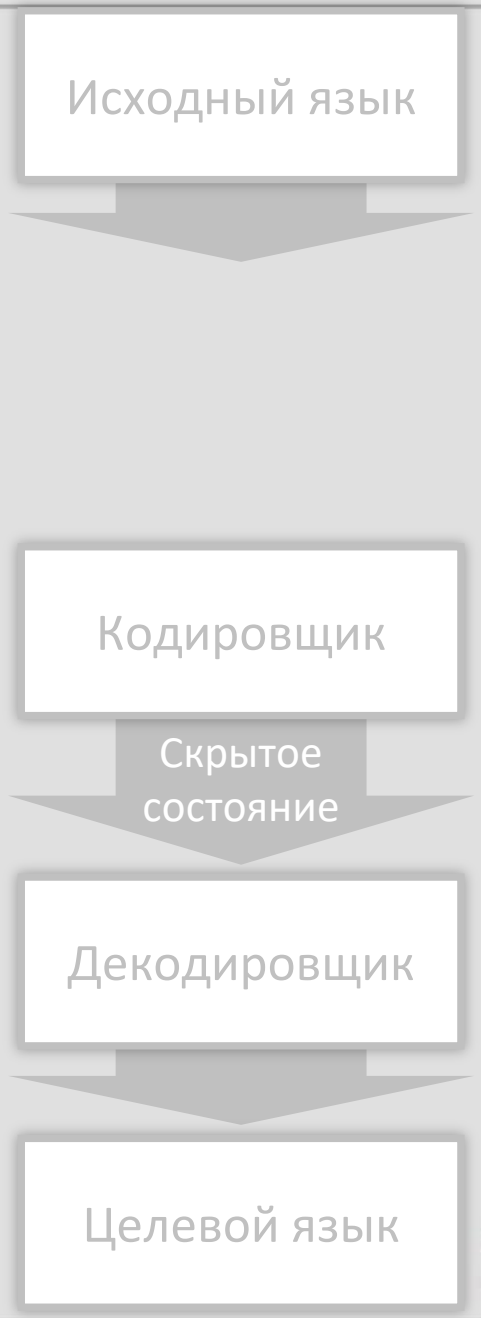
Дополнительная модель – сегментатор – определяет, в какой момент отправить в кодировщик последовательность слов

Синхронный перевод

Дополнительная модель – сегментатор – определяет, в какой момент отправить в кодировщик последовательность слов

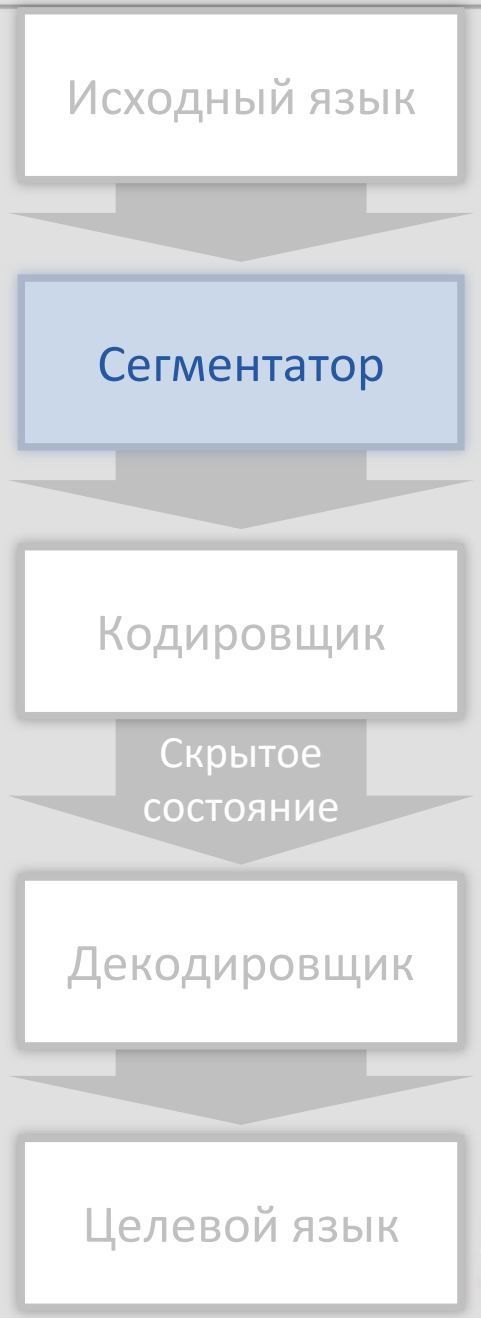
Синхронный перевод

Дополнительная модель – сегментатор – определяет, в какой момент отправить в кодировщик последовательность слов



Синхронный перевод

Дополнительная модель – сегментатор – определяет, в какой момент отправить в кодировщик последовательность слов

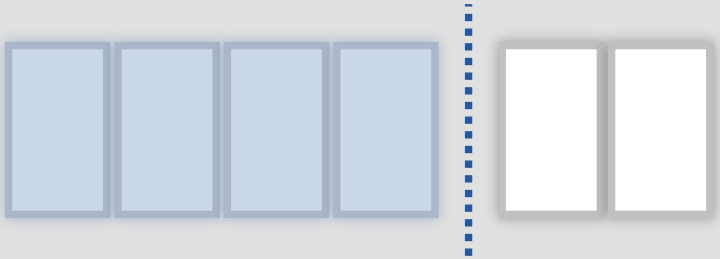


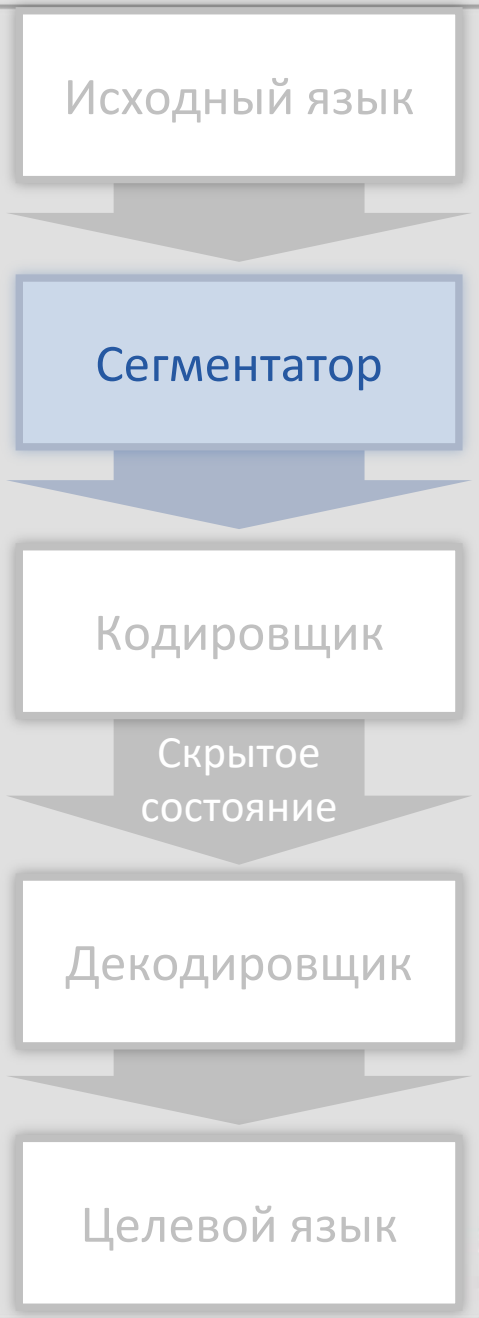
Синхронный перевод

Дополнительная модель – сегментатор – определяет, в какой момент отправить в кодировщик последовательность слов

Синхронный перевод

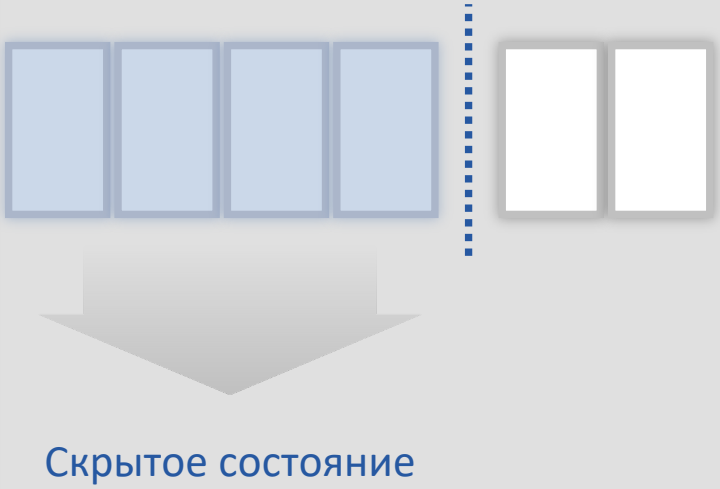
Дополнительная модель – сегментатор – определяет, в какой момент отправить в кодировщик последовательность слов





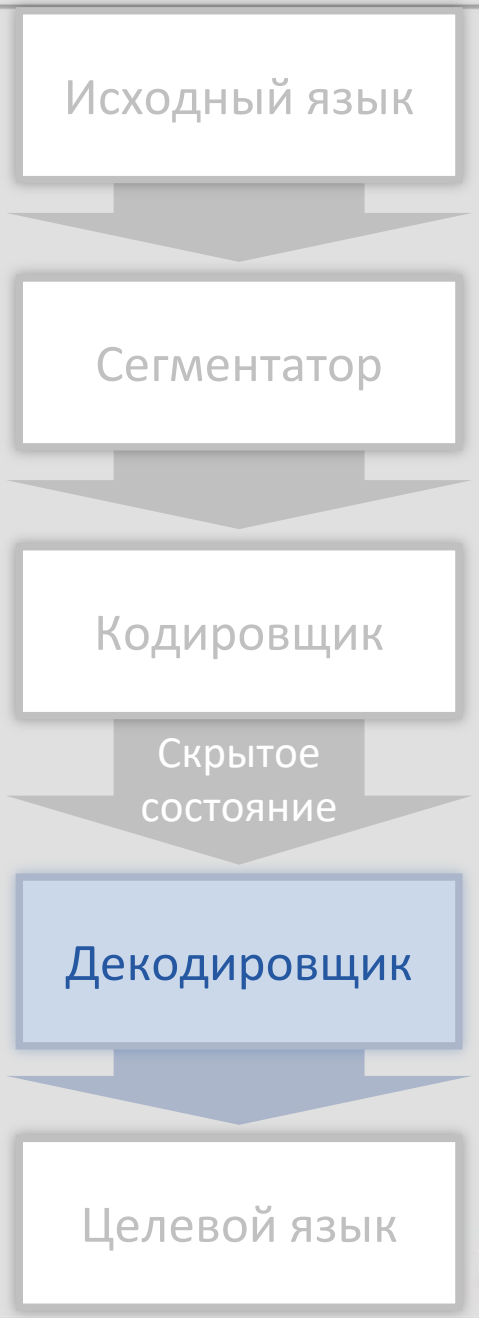
Синхронный перевод

Дополнительная модель – сегментатор – определяет, в какой момент отправить в кодировщик последовательность слов



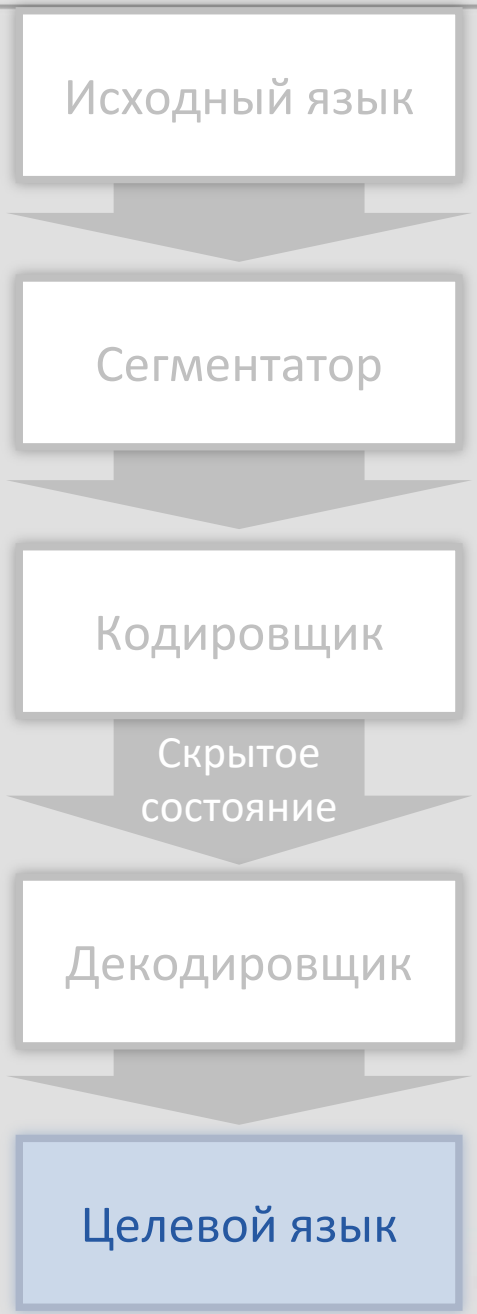
Синхронный перевод

Дополнительная модель – сегментатор – определяет, в какой момент отправить в кодировщик последовательность слов



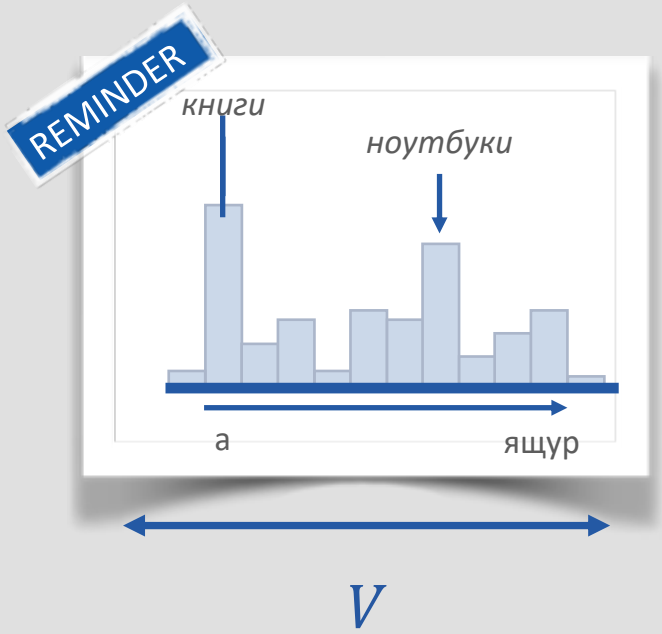
Синхронный перевод

Дополнительная модель – сегментатор – определяет, в какой момент отправить в кодировщик последовательность слов



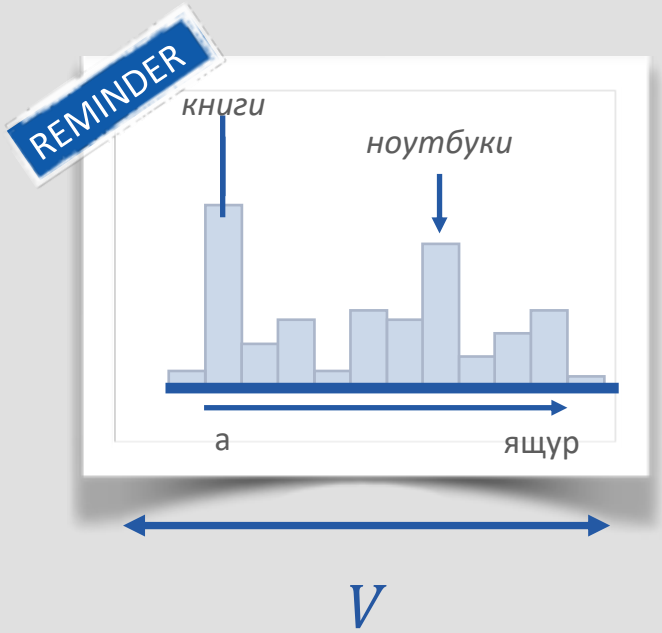
Открытый словарь

- Softmax на последнем слое – сильно ограничение: для добавления нового слова надо полностью переобучать последний слой



Открытый словарь

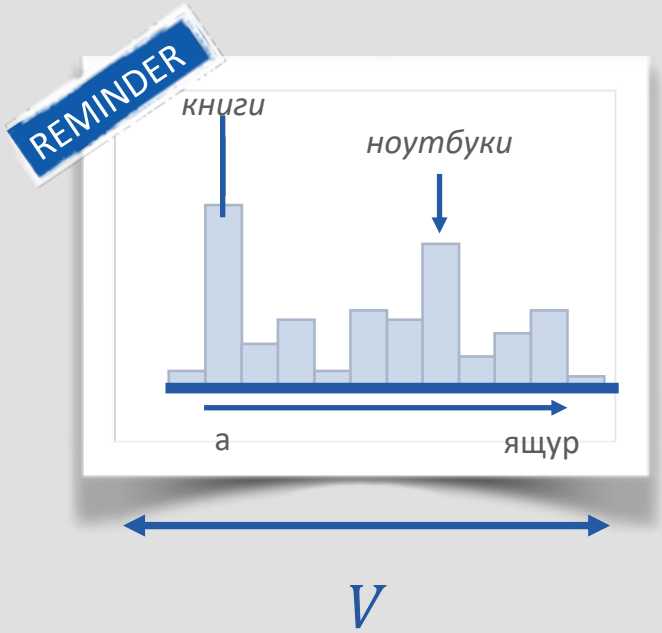
- Softmax на последнем слое – сильно ограничение: для добавления нового слова надо полностью переобучать последний слой



- Добавим специальный токен <unk> для слов, которых нет в словаре и будем их переводить по обычному словарю

Открытый словарь

- Softmax на последнем слое – сильно ограничение: для добавления нового слова надо полностью переобучать последний слой



- Добавим специальный токен <unk> для слов, которых нет в словаре и будем их переводить по обычному словарю
- Контрастная оценка по отношению к шуму (Noise-contrastive estimation, NCE): обучим $|V|$ классификаторов, отделяющих слово от шума

Слова, фрагменты слов, или символы?

Символьные
модели
машинного перевода

Слова, фрагменты слов, или символы?

Символьные модели машинного перевода

- Обрабатываем слова как последовательности символов

у н … е т

(университет)

Слова, фрагменты слов, или символы?

Символьные модели машинного перевода

- Обрабатываем слова как последовательности символов
- Потребуется токенизации на слова

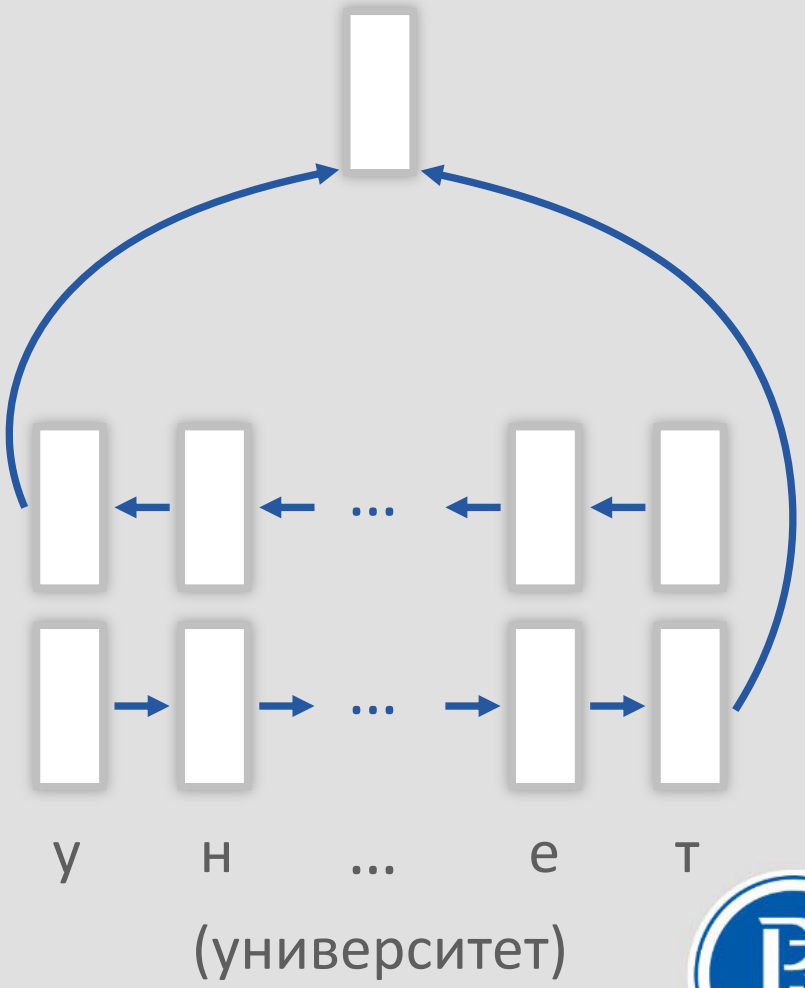
у н … е т

(университет)

Слова, фрагменты слов, или символы?

Символьные модели машинного перевода

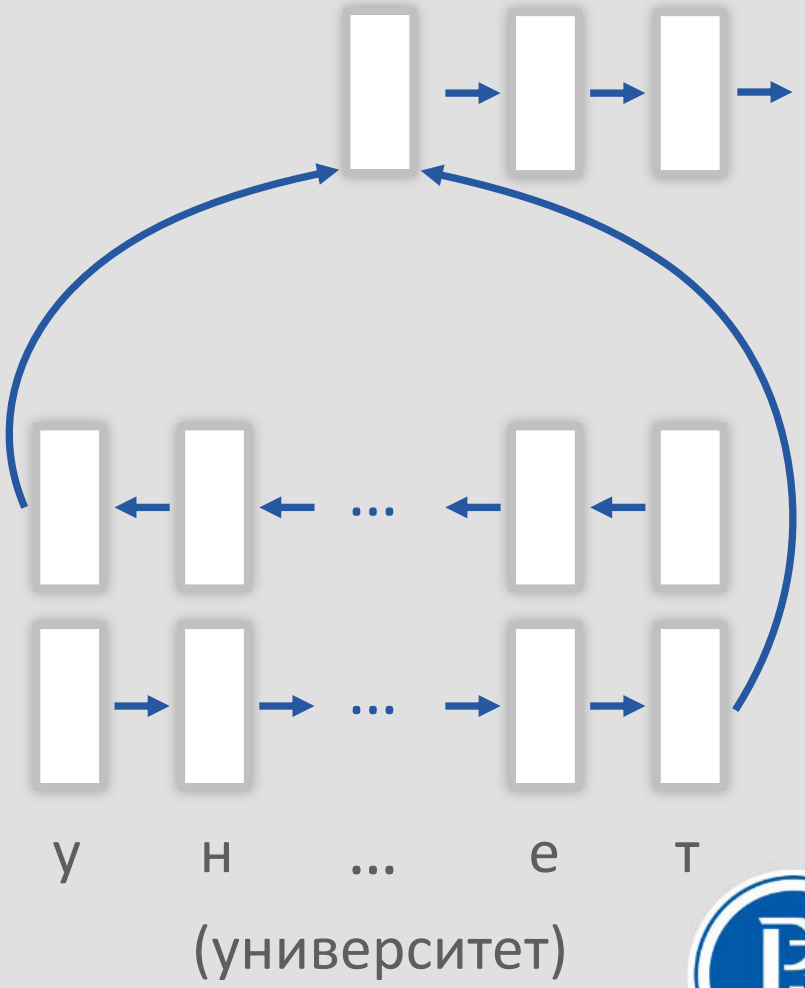
- Обрабатываем слова как последовательности символов
- Потребуется токенизации на слова
- Добавим специальные подсети, обрабатывающие последовательности символов



Слова, фрагменты слов, или символы?

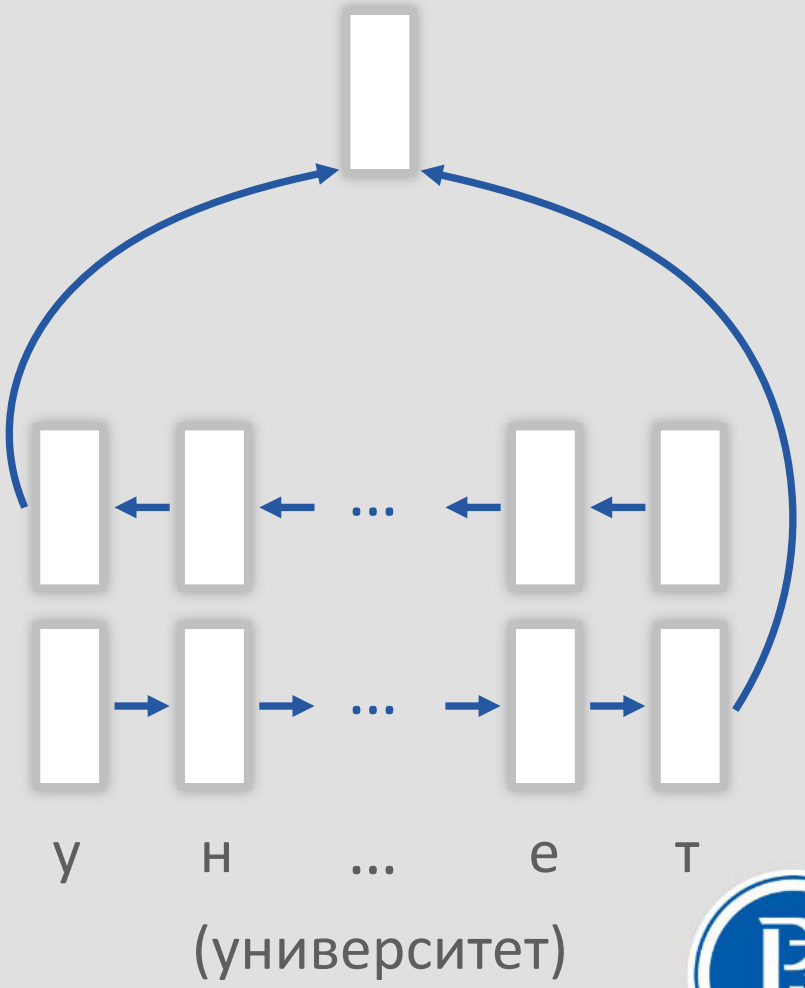
Символьные модели машинного перевода

- Обрабатываем слова как последовательности символов
- Потребуется токенизации на слова
- Добавим специальные подсети, обрабатывающие последовательности символов



Слова, фрагменты слов, или символы?

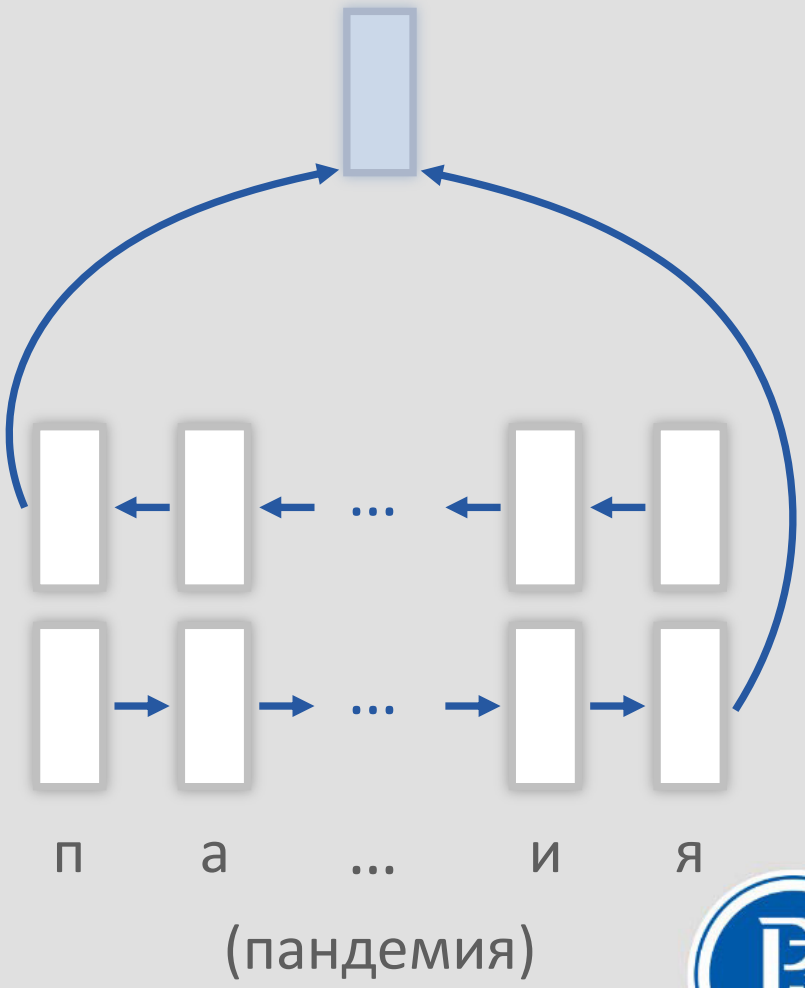
Комбинированное решение:
два вектора слова



Слова, фрагменты слов, или символы?

Комбинированное решение:
два вектора слова

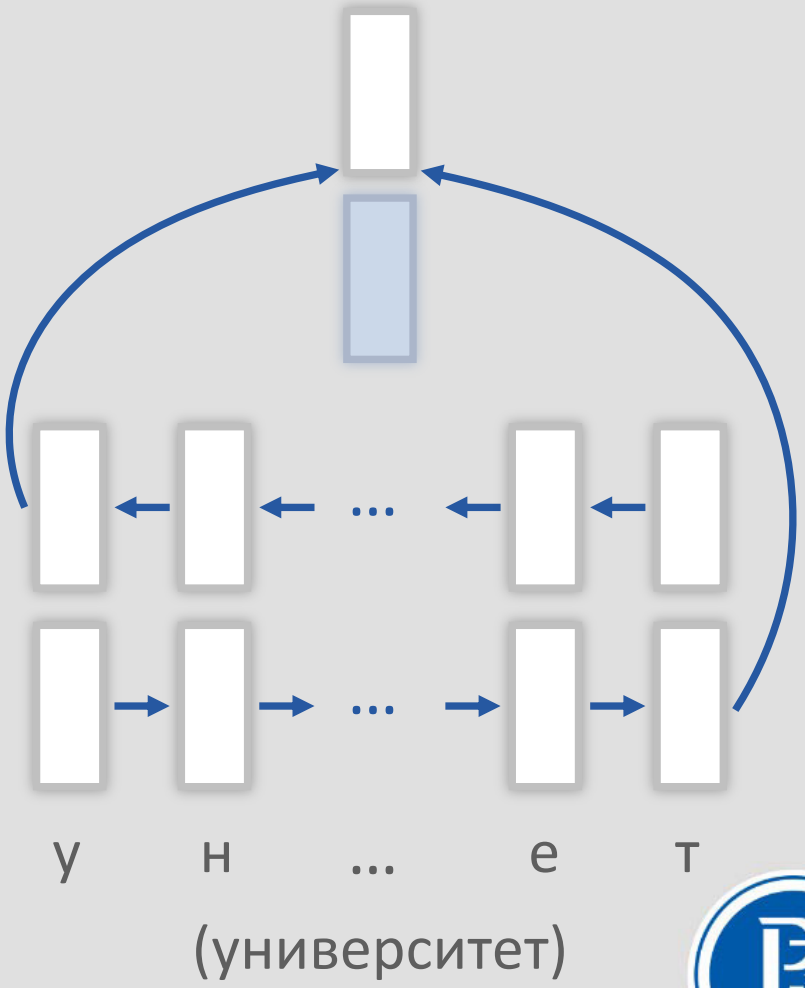
- Представление слова на уровне символов



Слова, фрагменты слов, или символы?

Комбинированное решение:
два вектора слова

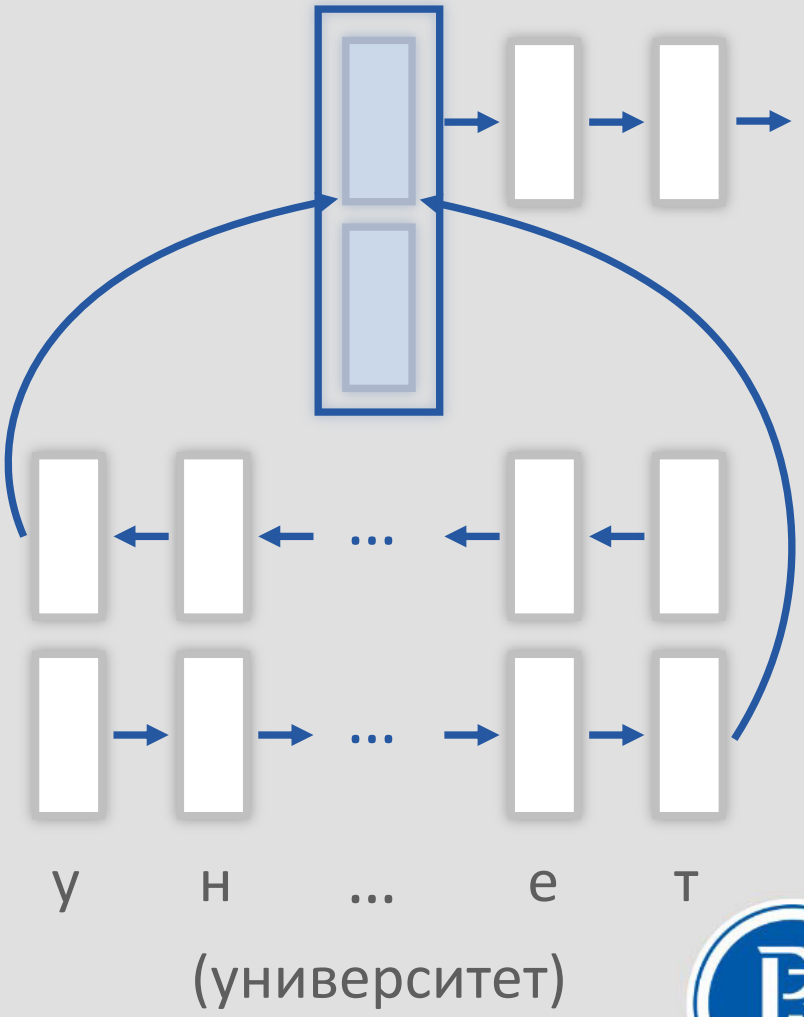
- Представление слова на уровне символов
- Стандартный вектор слова



Слова, фрагменты слов, или символы?

Комбинированное решение:
два вектора слова

- Представление слова на уровне символов
- Стандартный вектор слова



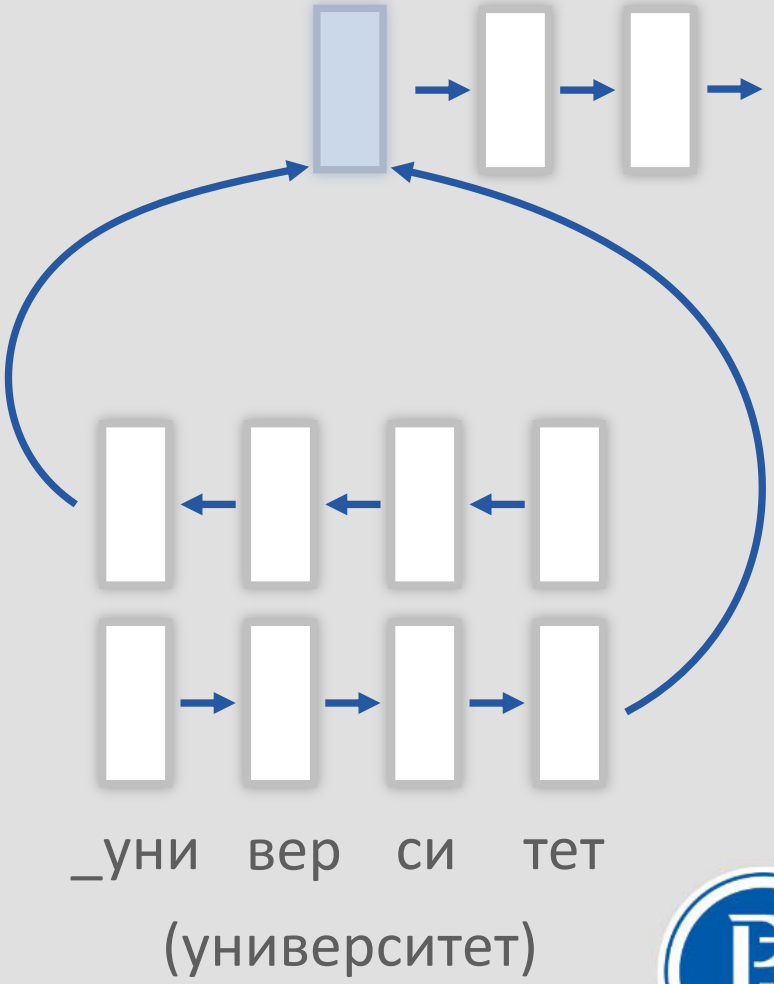
Слова, фрагменты слов, или символы?

Поделим слова на фрагменты

Будем подавать кодировщику на вход фрагменты слов.

Алгоритмы сегментации:

- Деление на морфемы
- Попарное битовое кодирование



Попарное битовое кодирование [byte pair encoding, bpe]

- Объединяем частые пары символов
- Выполняем фиксированное количество операций объединения

Попарное битовое кодирование [byte pair encoding, bpe]

- Объединяем частые пары символов
- Выполняем фиксированное количество операций объединения

Итерация	BPE
0	_г р ё з о б л а ж е н с т в у ю щ и й
1 000	
3 000	
5 000	
10 000	
25 000	
50 000	
100 000	
200 000	

Попарное битовое кодирование [byte pair encoding, bpe]

- Объединяем частые пары символов
- Выполняем фиксированное количество операций объединения

Итерация	BPE
0	_г р ё з о б л а ж е н с т в у ю щ и й
1 000	_г , р , ё , зо , б , ла , же , н , ству , ющи , й ,
3 000	
5 000	
10 000	
25 000	
50 000	
100 000	
200 000	

Попарное битовое кодирование [byte pair encoding, bpe]

- Объединяем частые пары символов
- Выполняем фиксированное количество операций объединения

Итерация	BPE
0	_г р ё з о б л а ж е н с т в у ю щ и й
1 000	_г , р , ё , зо , б , ла , же , н , ству , ющи , й ,
3 000	_г , рё , зо , б , ла , жен , ству , ющий ,
5 000	
10 000	
25 000	
50 000	
100 000	
200 000	

Попарное битовое кодирование [byte pair encoding, bpe]

- Объединяем частые пары символов
- Выполняем фиксированное количество операций объединения

Итерация	BPE
0	_г р ё з о б л а ж е н с т в у ю щ и й
1 000	_г , р , ё , зо , б , ла , же , н , ству , ющи , й ,
3 000	_г , рё , зо , б , ла , жен , ству , ющий ,
5 000	_г , рё , зо , б , ла , жен , ству , ющий ,
10 000	
25 000	
50 000	
100 000	
200 000	

Попарное битовое кодирование [byte pair encoding, bpe]

- Объединяем частые пары символов
- Выполняем фиксированное количество операций объединения

Итерация	BPE
0	_г р ё з о б л а ж е н с т в у ю щ и й
1 000	_г , р , ё , зо , б , ла , же , н , ству , ющи , й ,
3 000	_г , рё , зо , б , ла , жен , ству , ющий ,
5 000	_г , рё , зо , б , ла , жен , ству , ющий ,
10 000	_г , рё , зоб , ла , жен , ству , ющий ,
25 000	
50 000	
100 000	
200 000	

Попарное битовое кодирование [byte pair encoding, bpe]

- Объединяем частые пары символов
- Выполняем фиксированное количество операций объединения

Итерация	BPE
0	_г р ё з о б л а ж е н с т в у ю щ и й
1 000	_г , р , ё , зо , б , ла , же , н , ству , ющи , й ,
3 000	_г , рё , зо , б , ла , жен , ству , ющий ,
5 000	_г , рё , зо , б , ла , жен , ству , ющий ,
10 000	_г , рё , зоб , ла , жен , ству , ющий ,
25 000	_г , рё , зоб , ла , жен , ствующий ,
50 000	
100 000	
200 000	

Попарное битовое кодирование [byte pair encoding, bpe]

- Объединяем частые пары символов
- Выполняем фиксированное количество операций объединения

Итерация	BPE
0	_г р ё з о б л а ж е н с т в у ю щ и й
1 000	_г , р , ё , зо , б , ла , же , н , ству , ющи , й ,
3 000	_г , рё , зо , б , ла , жен , ству , ющий ,
5 000	_г , рё , зо , б , ла , жен , ству , ющий ,
10 000	_г , рё , зоб , ла , жен , ству , ющий ,
25 000	_г , рё , зоб , ла , жен , ствующий ,
50 000	_г , рё , зоб , ла , жен , ствующий ,
100 000	
200 000	

Попарное битовое кодирование [byte pair encoding, bpe]

- Объединяем частые пары символов
- Выполняем фиксированное количество операций объединения

Итерация	BPE
0	_г р ё з о б л а ж е н с т в у ю щ и й
1 000	_г , р , ё , зо , б , ла , же , н , ству , ющи , й ,
3 000	_г , рё , зо , б , ла , жен , ству , ющий ,
5 000	_г , рё , зо , б , ла , жен , ству , ющий ,
10 000	_г , рё , зоб , ла , жен , ству , ющий ,
25 000	_г , рё , зоб , ла , жен , ствующий ,
50 000	_г , рё , зоб , ла , жен , ствующий ,
100 000	_грё , зоб , ла , жен , ствующий ,
200 000	

Попарное битовое кодирование [byte pair encoding, bpe]

- Объединяем частые пары символов
- Выполняем фиксированное количество операций объединения

Итерация	BPE
0	_г р ё з о б л а ж е н с т в у ю щ и й
1 000	_г , р , ё , зо , б , ла , же , н , ству , ющи , й ,
3 000	_г , рё , зо , б , ла , жен , ству , ющий ,
5 000	_г , рё , зо , б , ла , жен , ству , ющий ,
10 000	_г , рё , зоб , ла , жен , ству , ющий ,
25 000	_г , рё , зоб , ла , жен , ствующий ,
50 000	_г , рё , зоб , ла , жен , ствующий ,
100 000	_грё , зоб , ла , жен , ствующий ,
200 000	_грё , зобла , жен , ствующий

Представление по символам, подсловам и словам

Модель
на уровне символов

Представление по символам, подсловам и словам

Модель
на уровне символов

в	е	с	е	л	ы	й
3	6	19	6	13	29	11

в	е	с	е	л	е	е
3	6	19	6	13	6	6

б	о	д	р	ы	й
2	16	5	18	29	11

б	о	д	р	е	е
2	16	5	18	6	6

Представление по символам, подсловам и словам

Модель
на уровне символов

размер
словаря

в	е	с	е	л	ы	й
3	6	19	6	13	29	11

в	е	с	е	л	е	е
3	6	19	6	13	6	6

б	о	д	р	ы	й
2	16	5	18	29	11

б	о	д	р	е	е
2	16	5	18	6	6

Представление по символам, подсловам и словам

Модель
на уровне символов

размер
словаря

в	е	с	е	л	ы	й	весел	ый
3	6	19	6	13	29	11	17835	284
в	е	с	е	л	е	е	весел	ее
3	6	19	6	13	6	6	17835	791
б	о	д	р	ы	й		бодр	ый
2	16	5	18	29	11		14502	284
б	о	д	р	е	е		бодр	ее
2	16	5	18	6	6		14502	791

Представление по символам, подсловам и словам

Модель на уровне символов

размер словаря

Модель на уровне слов

в	е	с	е	л	ы	й
3	6	19	6	13	29	11

весел	ый
17835	284

в	е	с	е	л	е	е
3	6	19	6	13	6	6

весел	ее
17835	791

б	о	д	р	ы	й
2	16	5	18	29	11

бодр	ый
14502	284

б	о	д	р	е	е
2	16	5	18	6	6

бодр	ее
14502	791

Представление по символам, подсловам и словам

Модель на уровне символов

размер словаря

Модель на уровне слов

в (3)	е (6)	с (19)	е (6)	л (13)	ы (29)	й (11)	весел (17835)	ый (284)	веселый (23941)
в (3)	е (6)	с (19)	е (6)	л (13)	е (6)	е (6)	весел (17835)	ее (791)	веселее (13008)
б (2)	о (16)	д (5)	р (18)	ы (29)	й (11)		бодр (14502)	ый (284)	бодрый (1472)
б (2)	о (16)	д (5)	р (18)	е (6)	е (6)		бодр (14502)	ее (791)	бодрее (18772)

Слова, фрагменты слов, или символы?

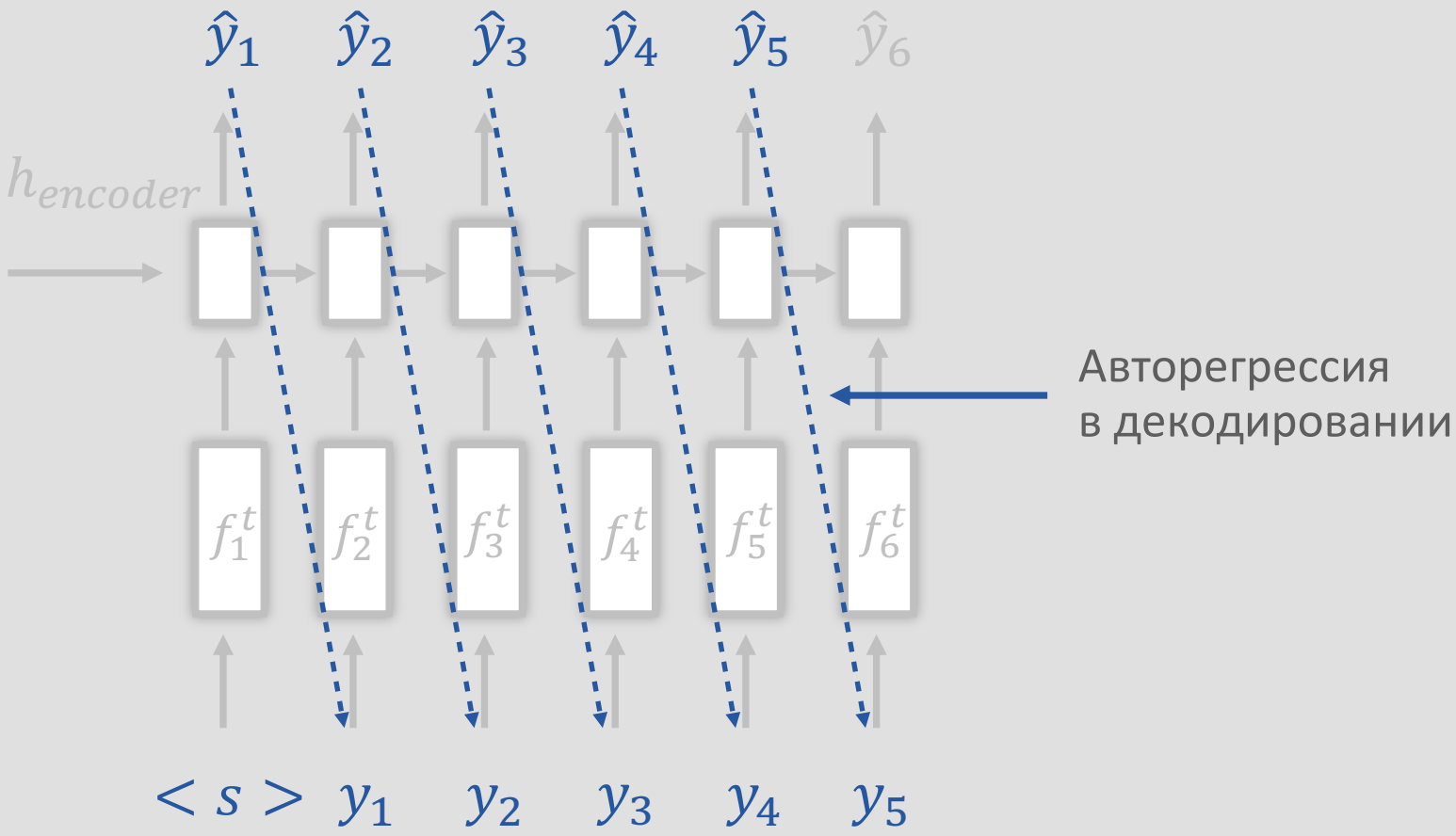
Перевод по символам

+ лучше транслитерирует
+ лучше понимает морфологию
+ более устойчив к шуму
− требует большего объема вычислений

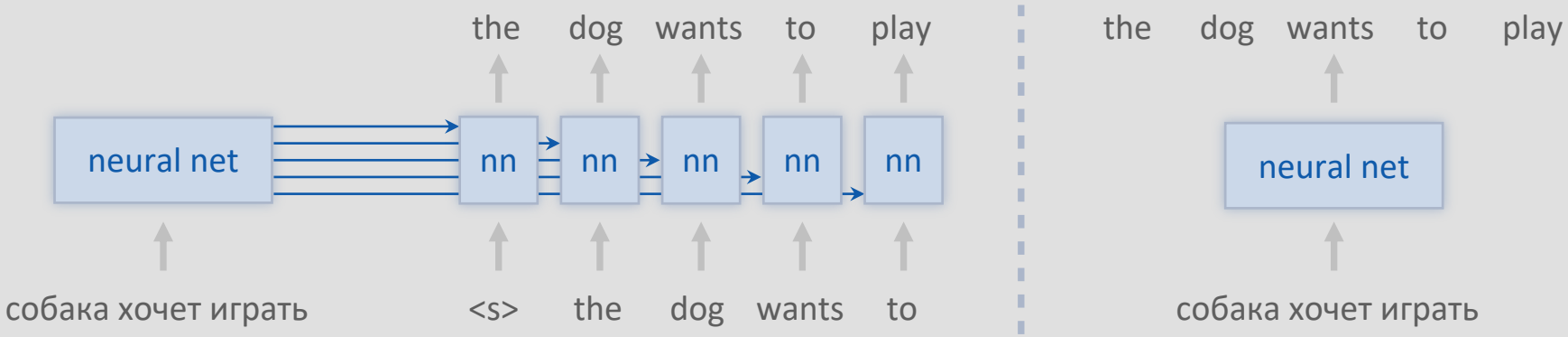
Перевод по фрагментам слов

+ получаются грамматически корректные переводы
+ поддерживает больший объем словаря
+ лучше понимает синтаксис
+ дает лучшие показатели качества

Медленный авторегрессионный перевод



Неавторегрессионный перевод



Авторегрессионный

Неавторегрессионный

Машинный перевод без учителя

Машинный перевод без учителя

- Если нет параллельных данных или параллельный корпус очень маленький, можно использовать технологии машинного перевода без учителя

Facebook unsupervised MT

Машинный перевод без учителя

- Если нет параллельных данных или параллельный корпус очень маленький, можно использовать технологии машинного перевода без учителя
- Три ключевые компоненты:

Facebook unsupervised MT

Машинный перевод без учителя

- Если нет параллельных данных или параллельный корпус очень маленький, можно использовать технологии машинного перевода без учителя
- Три ключевые компоненты:
  - Алгоритм перевода слов (автоматический словарь русский ↔ иврит)

Facebook unsupervised MT

Машинный перевод без учителя

- Если нет параллельных данных или параллельный корпус очень маленький, можно использовать технологии машинного перевода без учителя
- Три ключевые компоненты:
  - Алгоритм перевода слов (автоматический словарь русский ↔ иврит)
  - Языковая модель для русского и языковая модель для иврита

Facebook unsupervised MT

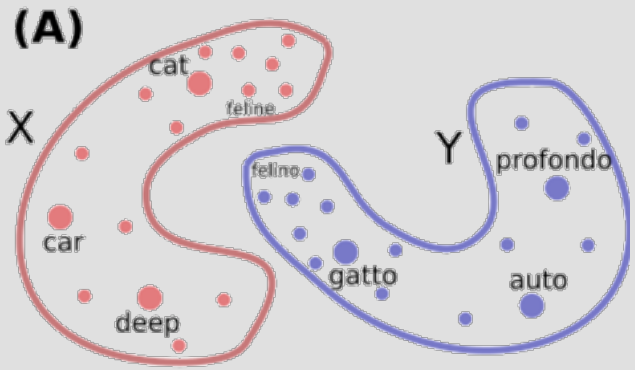
Машинный перевод без учителя

- Если нет параллельных данных или параллельный корпус очень маленький, можно использовать технологии машинного перевода без учителя
- Три ключевые компоненты:
  - Алгоритм перевода слов (автоматический словарь русский ↔ иврит)
  - Языковая модель для русского и языковая модель для иврита
  - Модель машинного перевода

Facebook unsupervised MT

Перевод по словам

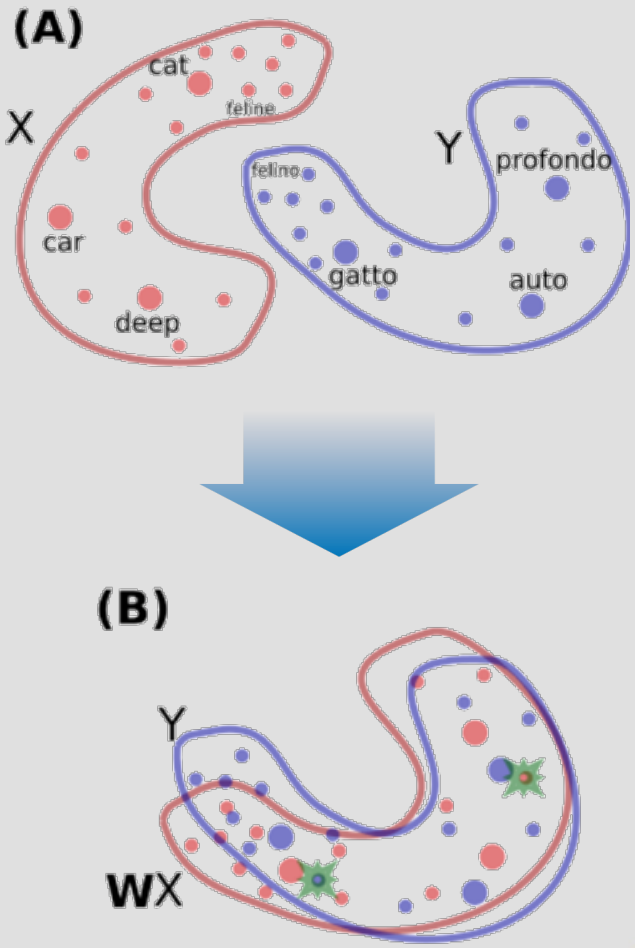
- Поворачиваем пространство векторов слов на одном языке в пространство слов на другом языке в несколько шагов
- Шаг А. Задаем независимые пространства векторов слов на каждом языке, X и Y



➤ *Conneau, Alexis, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. "Word translation without parallel data." ICLR 2018*

Перевод по словам

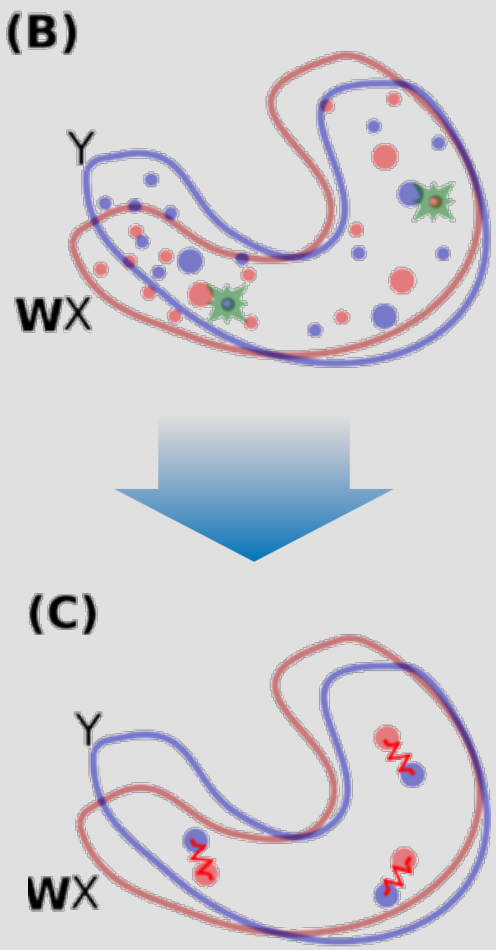
- Поворачиваем пространство векторов слов на одном языке в пространство слов на другом языке в несколько шагов
- Шаг B. Выбираем опорные точки – несколько пар слов, перевод которых известен.



➤ *Conneau, Alexis, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. "Word translation without parallel data." ICLR 2018*

Перевод по словам

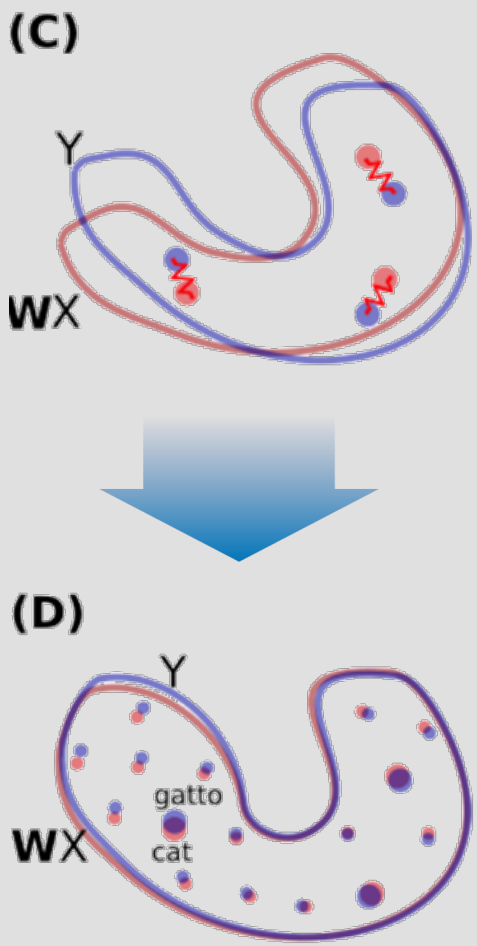
- Поворачиваем пространство векторов слов на одном языке в пространство слов на другом языке в несколько шагов
- Шаг C. Используем преобразование Прокруста и получаем приблизительное выравнивание для остальных слов
- Преобразование Прокруста растягивает и поворачивает пространство X так, чтобы опорные точки в X совпали с опорными точками Y



➤ *Conneau, Alexis, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. "Word translation without parallel data." ICLR 2018*

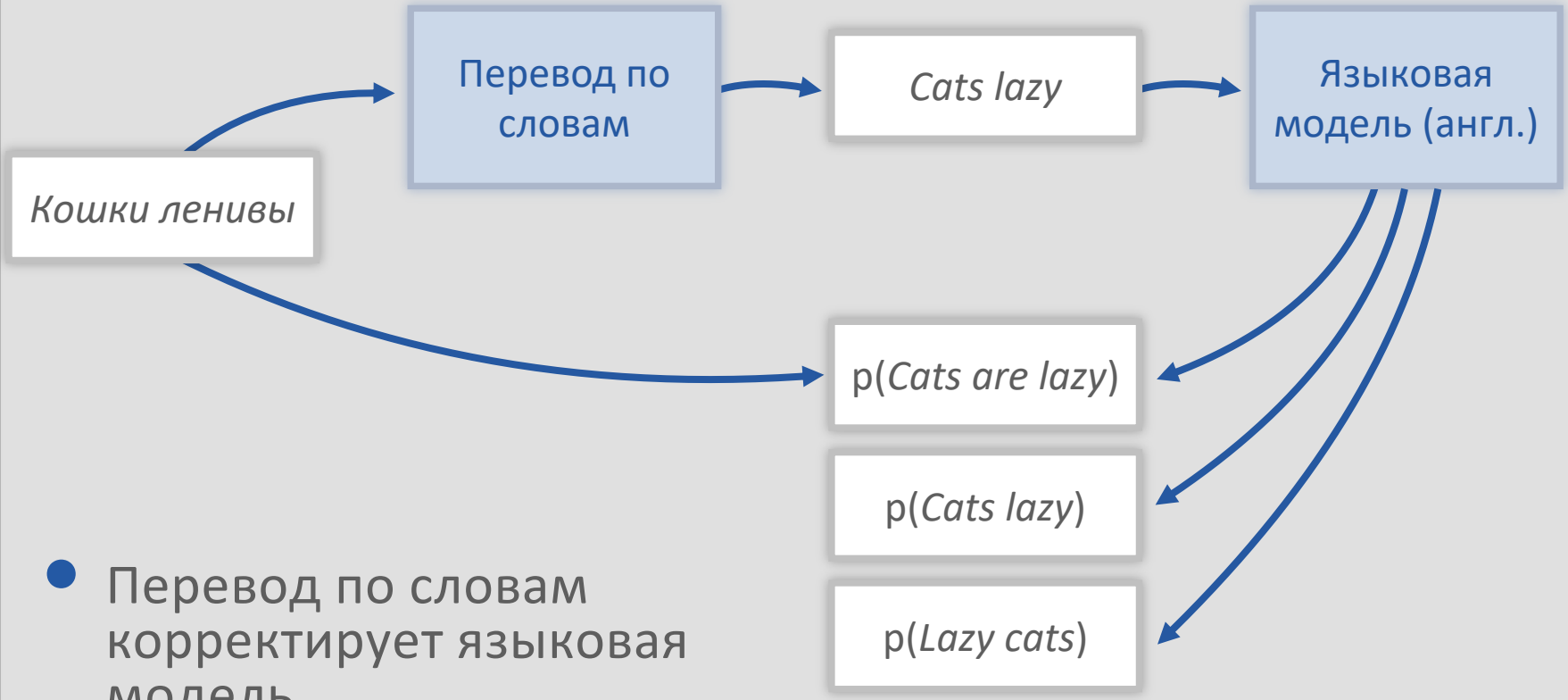
Перевод по словам

- Поворачиваем пространство векторов слов на одном языке в пространство слов на другом языке в несколько шагов
- Шаг D. С помощью метода ближайших соседей доуточняем выравнивание вокруг частых слов.



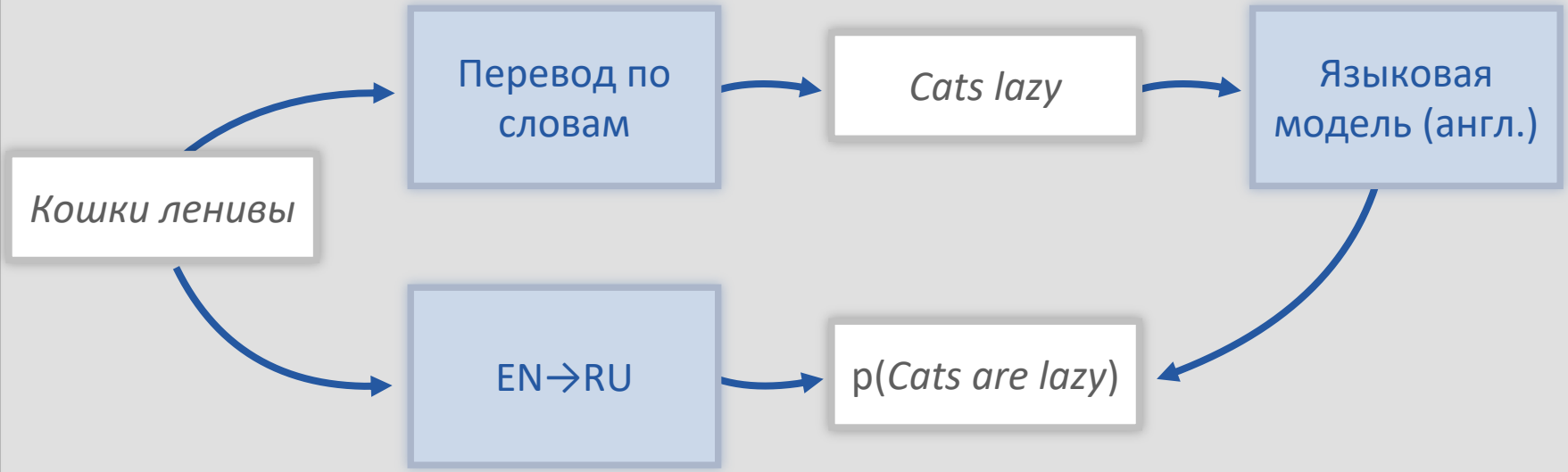
➤ *Conneau, Alexis, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. "Word translation without parallel data." ICLR 2018*

Машинный перевод без учителя



- Перевод по словам корректирует языковая модель
- Получаем пары предложений для обучения модели машинного перевода

Машинный перевод без учителя



Машинный перевод без учителя

- Используем обратный перевод:

Facebook unsupervised MT

Машинный перевод без учителя

- Используем обратный перевод:
  - Из исходного предложения на русском, получаем пару предложений на английском и русском

Facebook unsupervised MT

Машинный перевод без учителя

- Используем обратный перевод:
  - Из исходного предложения на русском, получаем пару предложений на английском и русском
  - Обучаем модель перевода с английского на русский

Facebook unsupervised MT

Машинный перевод без учителя

- Используем обратный перевод:
  - Из исходного предложения на русском, получаем пару предложений на английском и русском
  - Обучаем модель перевода с английского на русский
  - То же самое повторяем для исходных предложений на английском

Facebook unsupervised MT

Машинный перевод без учителя

- Используем обратный перевод:
  - Из исходного предложения на русском, получаем пару предложений на английском и русском
  - Обучаем модель перевода с английского на русский
  - То же самое повторяем для исходных предложений на английском
  - Используем обе модели машинного перевода, получаем новые синтетические данные для обучения

Facebook unsupervised MT

Машинный перевод без учителя

- Мы обсудили метод генерации параллельных данных для обучения моделей машинного перевода
- Исследователи из Facebook достигли существенного улучшения качества перевода с помощью этого метода

Facebook unsupervised MT

Заключение

- Машинный перевод – одна из самых сложных, популярных и востребованных задач

Заключение

- Машинный перевод – одна из самых сложных, популярных и востребованных задач
- Для обучения моделей требуются, как правило, параллельные корпуса

Заключение

- Машинный перевод – одна из самых сложных, популярных и востребованных задач
- Для обучения моделей требуются, как правило, параллельные корпуса
- Основной класс архитектур: кодировщик-декодировщик

Заключение

- Машинный перевод – одна из самых сложных, популярных и востребованных задач
- Для обучения моделей требуются, как правило, параллельные корпуса
- Основной класс архитектур: кодировщик-декодировщик
- Направление развития нейросетевого машинного перевода:

Заключение

- Машинный перевод – одна из самых сложных, популярных и востребованных задач
- Для обучения моделей требуются, как правило, параллельные корпуса
- Основной класс архитектур: кодировщик-декодировщик
- Направление развития нейросетевого машинного перевода:
  - Отказ от параллельных корпусов и использование синтетических данных

Заключение

- Машинный перевод – одна из самых сложных, популярных и востребованных задач
- Для обучения моделей требуются, как правило, параллельные корпуса
- Основной класс архитектур: кодировщик-декодировщик
- Направление развития нейросетевого машинного перевода:
  - Отказ от параллельных корпусов и использование синтетических данных
  - Использование обучения с подкреплением для оптимизации BLUE

Заключение

- Машинный перевод – одна из самых сложных, популярных и востребованных задач
- Для обучения моделей требуются, как правило, параллельные корпуса
- Основной класс архитектур: кодировщик-декодировщик
- Направление развития нейросетевого машинного перевода:
  - Отказ от параллельных корпусов и использование синтетических данных
  - Использование обучения с подкреплением для оптимизации BLUE
  - Ускорение декодирования с помощью неавторегрессионного вывода

Заключение

- Машинный перевод – одна из самых сложных, популярных и востребованных задач
- Для обучения моделей требуются, как правило, параллельные корпуса
- Основной класс архитектур: кодировщик-декодировщик
- Направление развития нейросетевого машинного перевода:
  - Отказ от параллельных корпусов и использование синтетических данных
  - Использование обучения с подкреплением для оптимизации BLUE
  - Ускорение декодирования с помощью неавторегрессионного вывода
  - Перевод со многих языков на многие