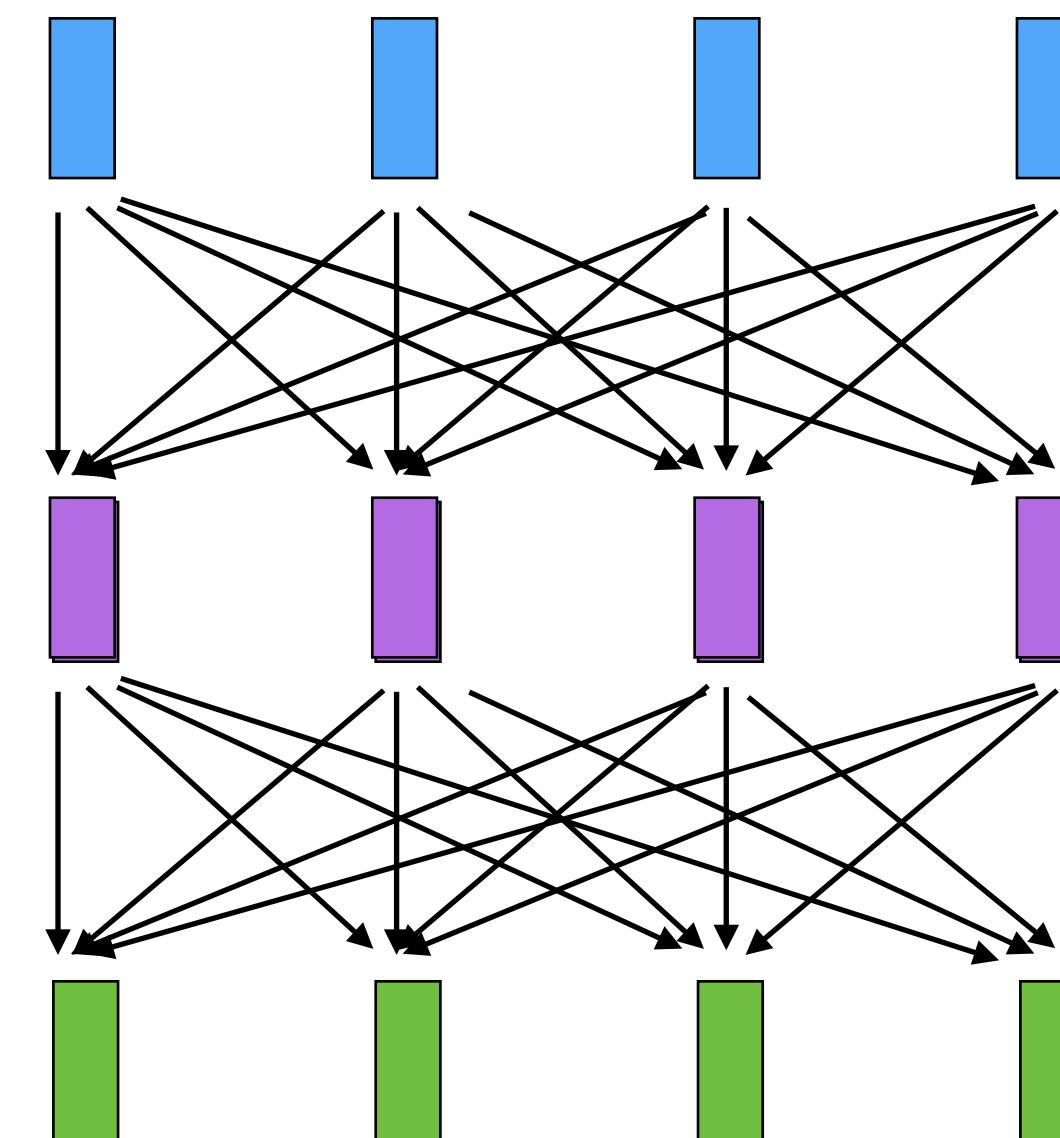


Нейронные сети для обработки текстов

Надежда Чиркова

Лаборатория компании Самсунг

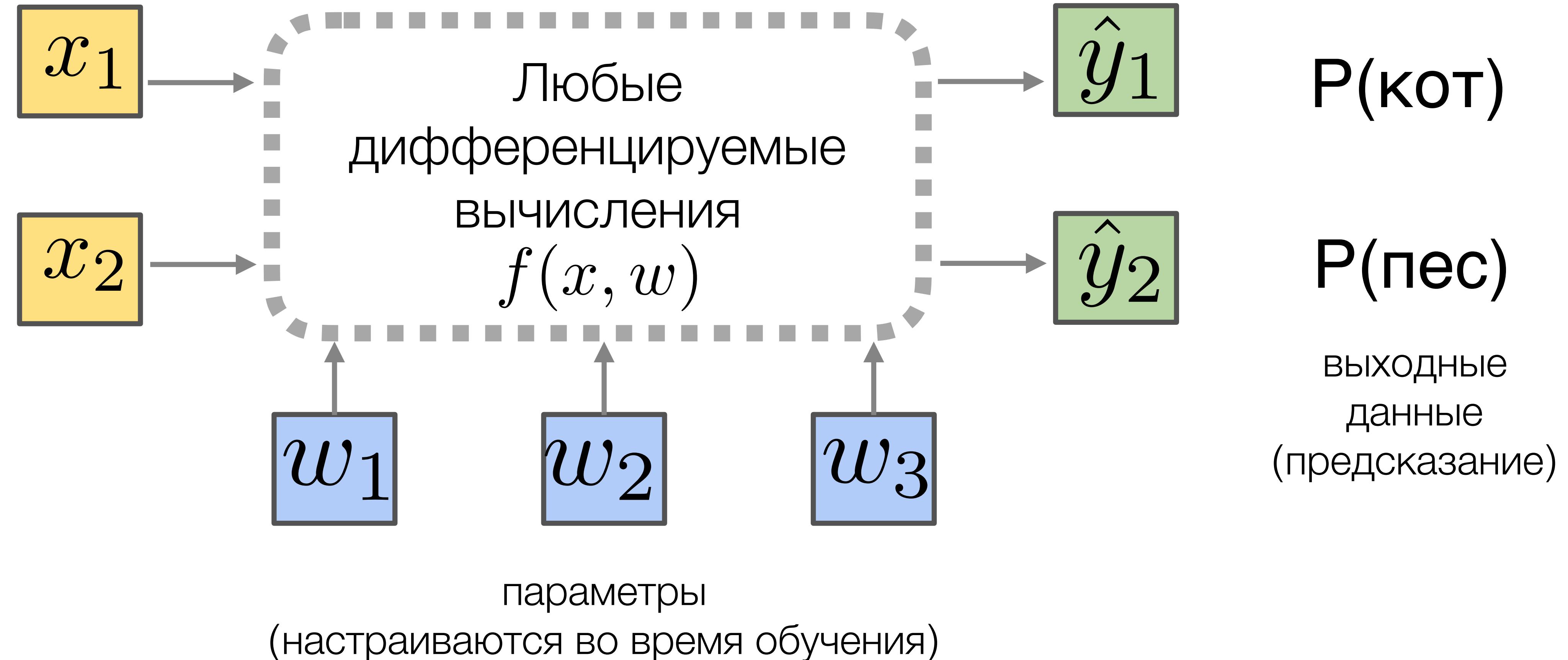
Высшая школа экономики



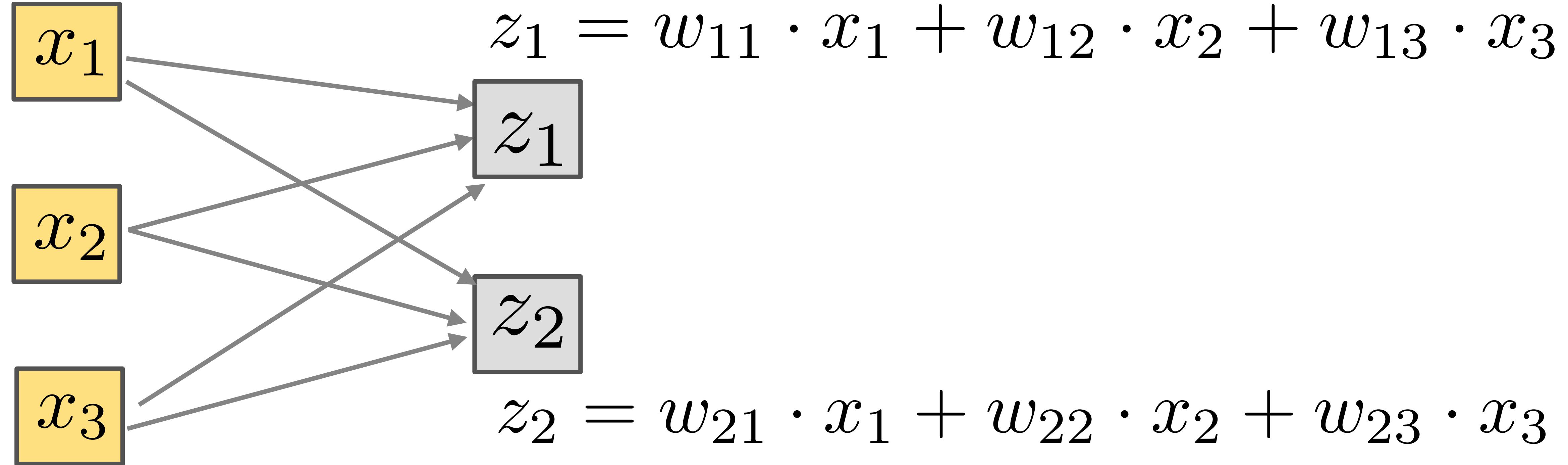
Общая идея графа вычислений



входные
даные
(объект)

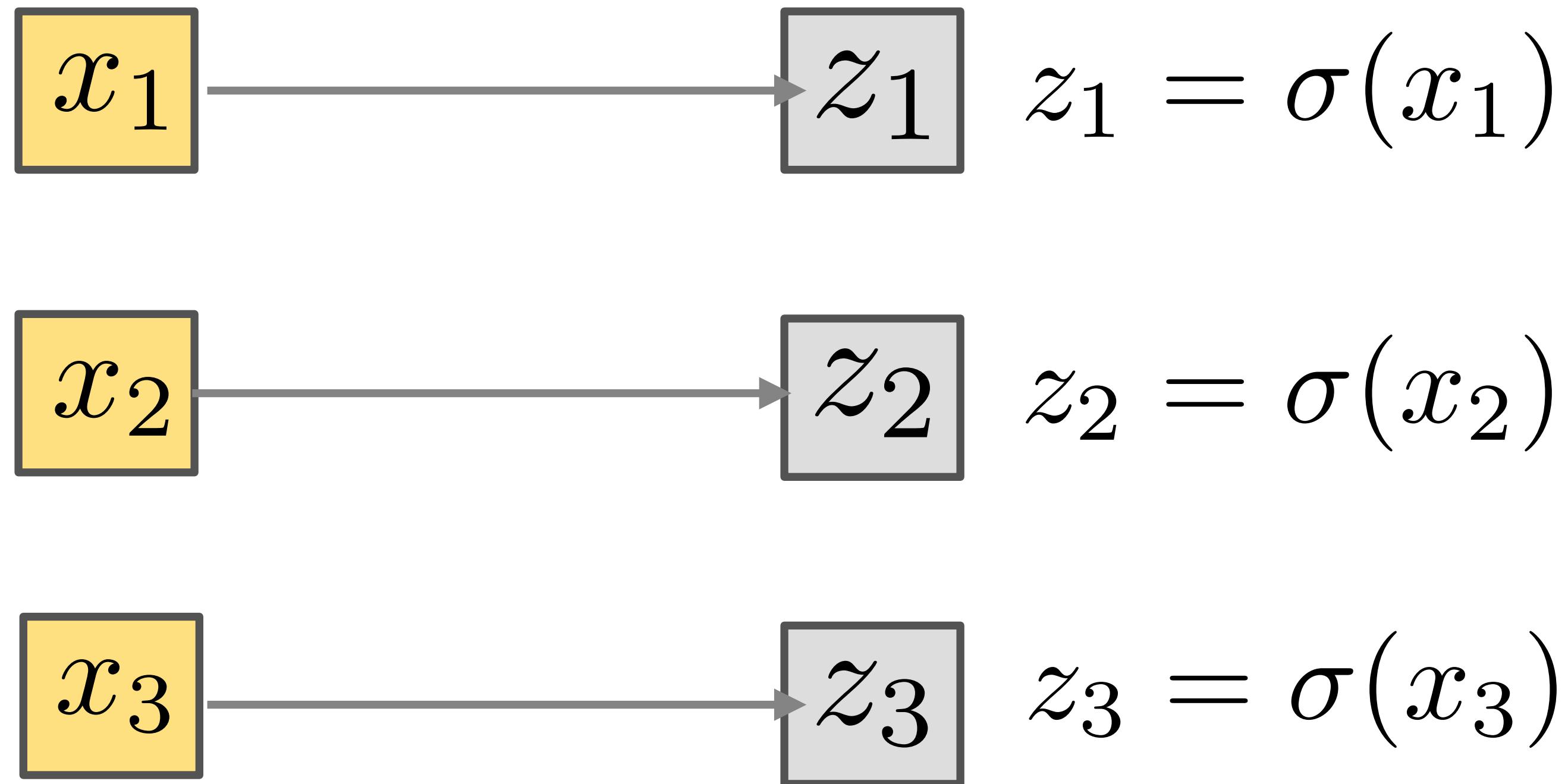


Полносвязный слой



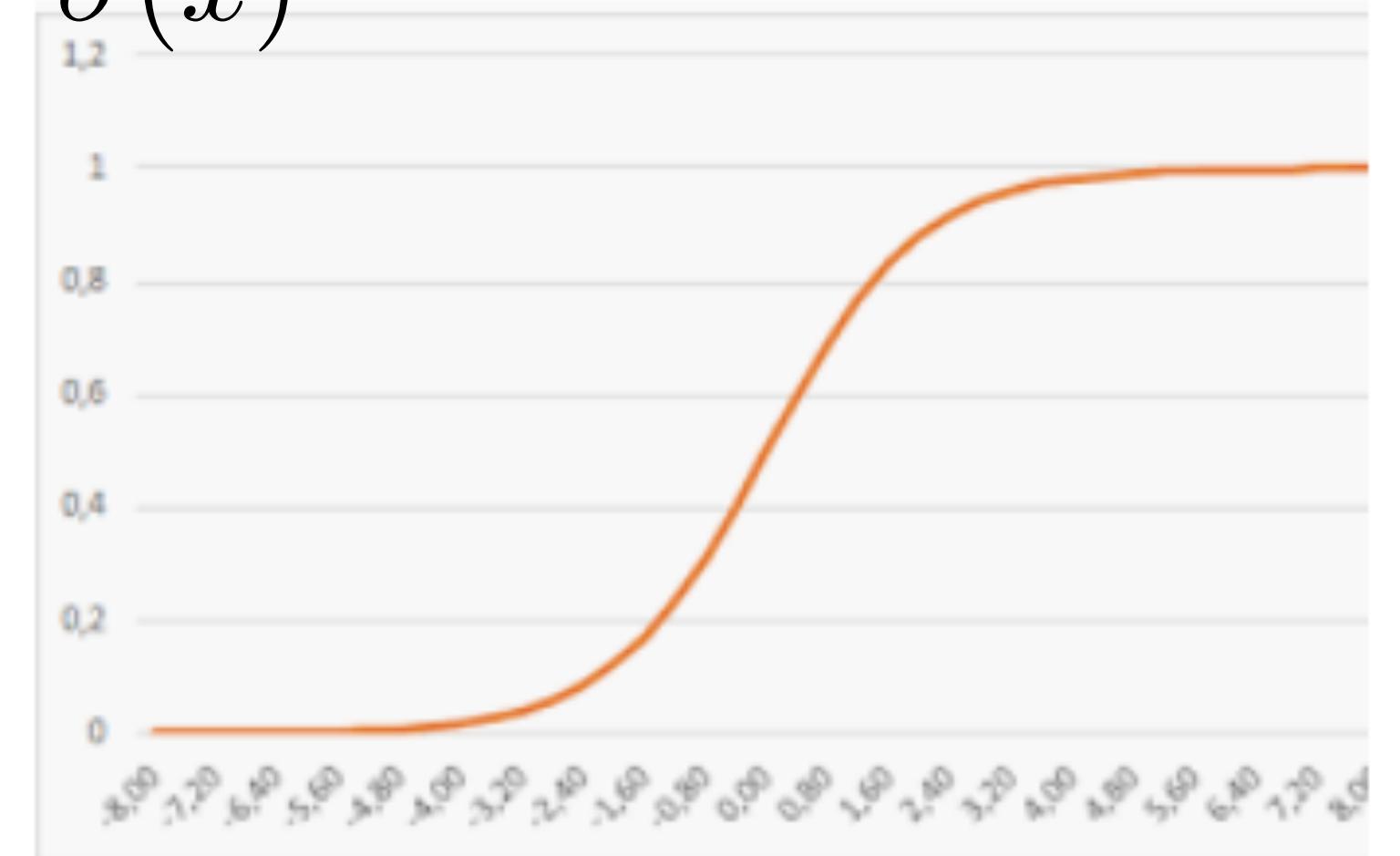
Параметры: w_{ij}

Слой нелинейности



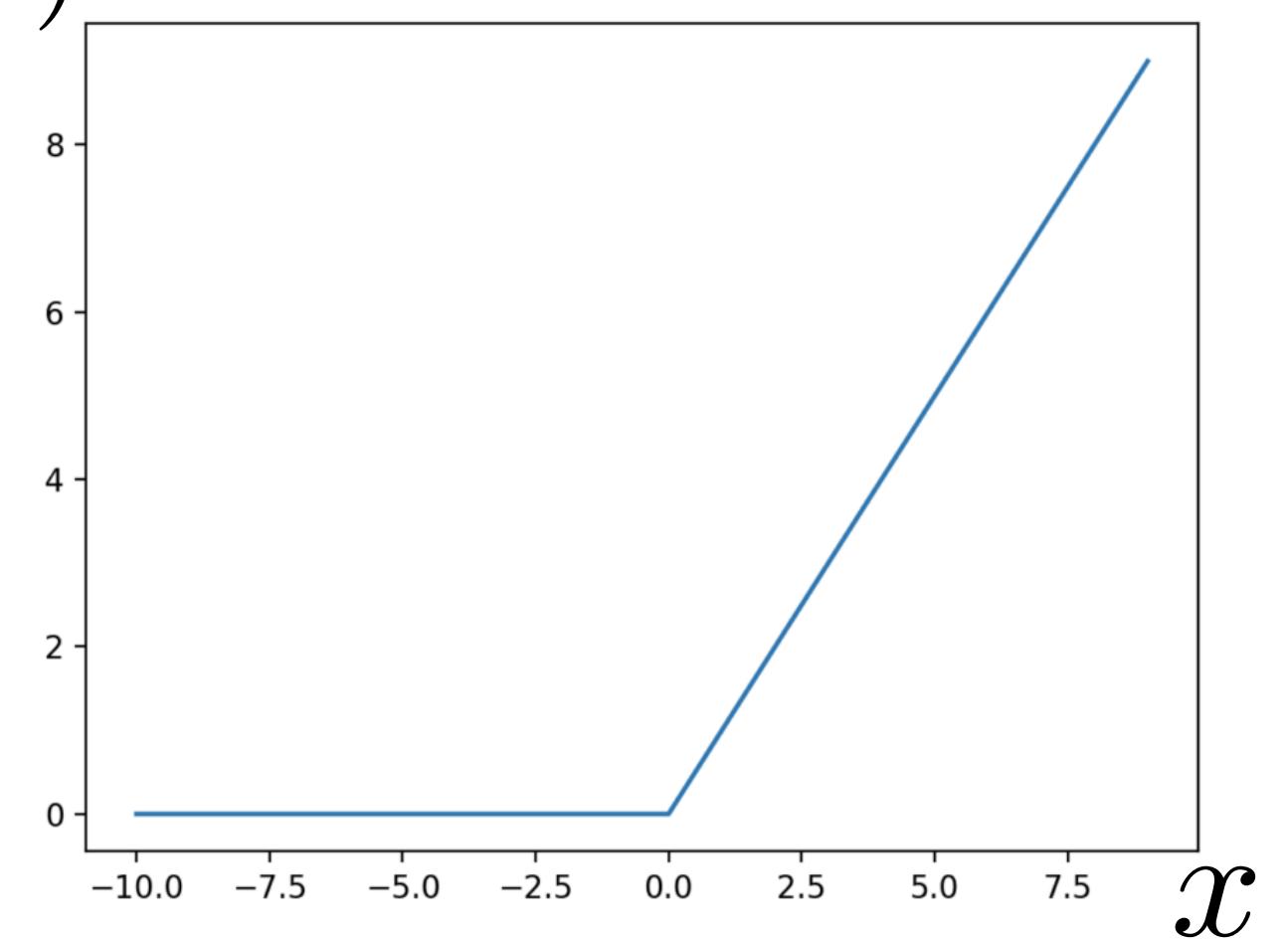
Параметров нет!

$$\sigma(x)$$



$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

$$\sigma(x)$$



$$\text{ReLU: } \sigma(x) = \max(0, x)$$

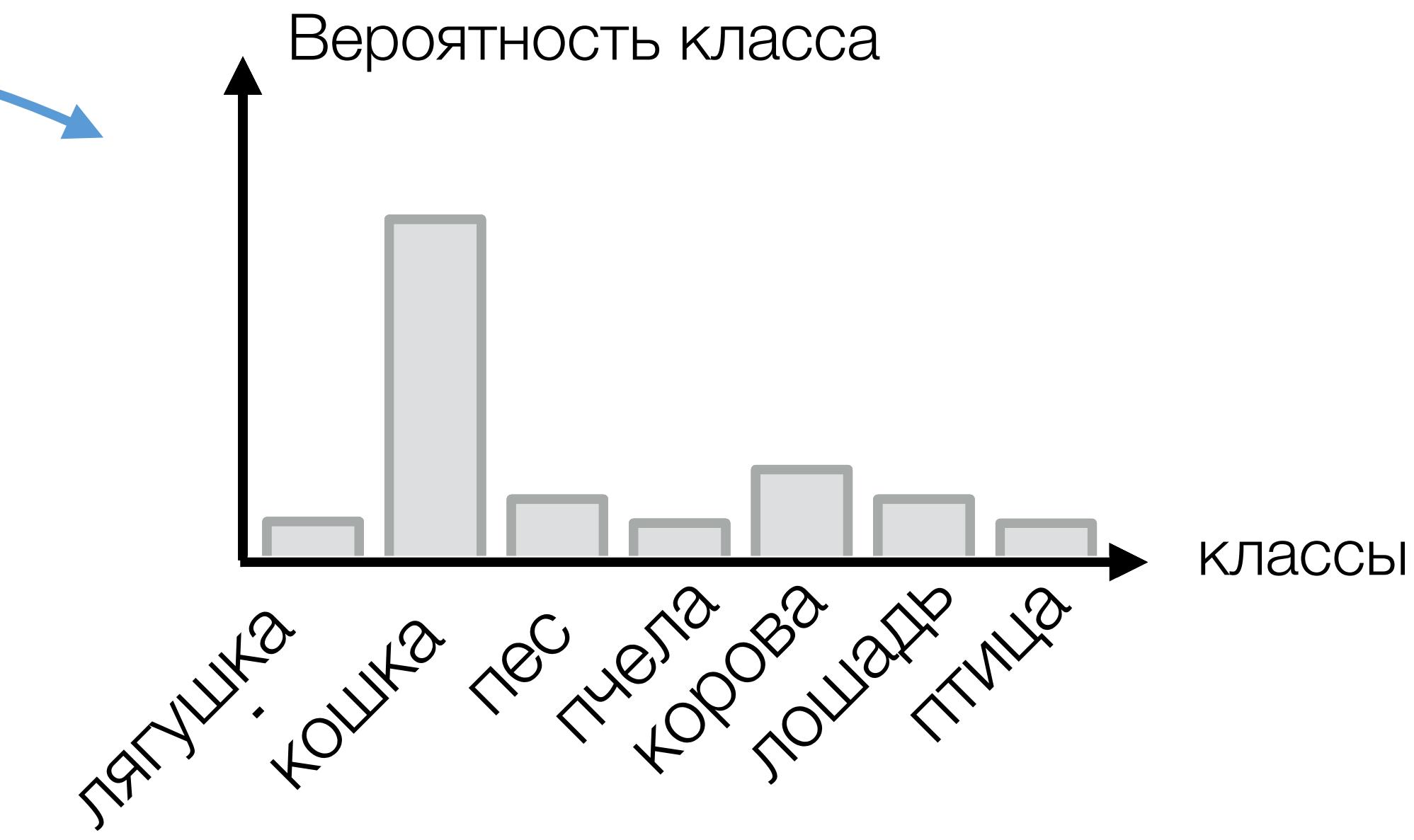
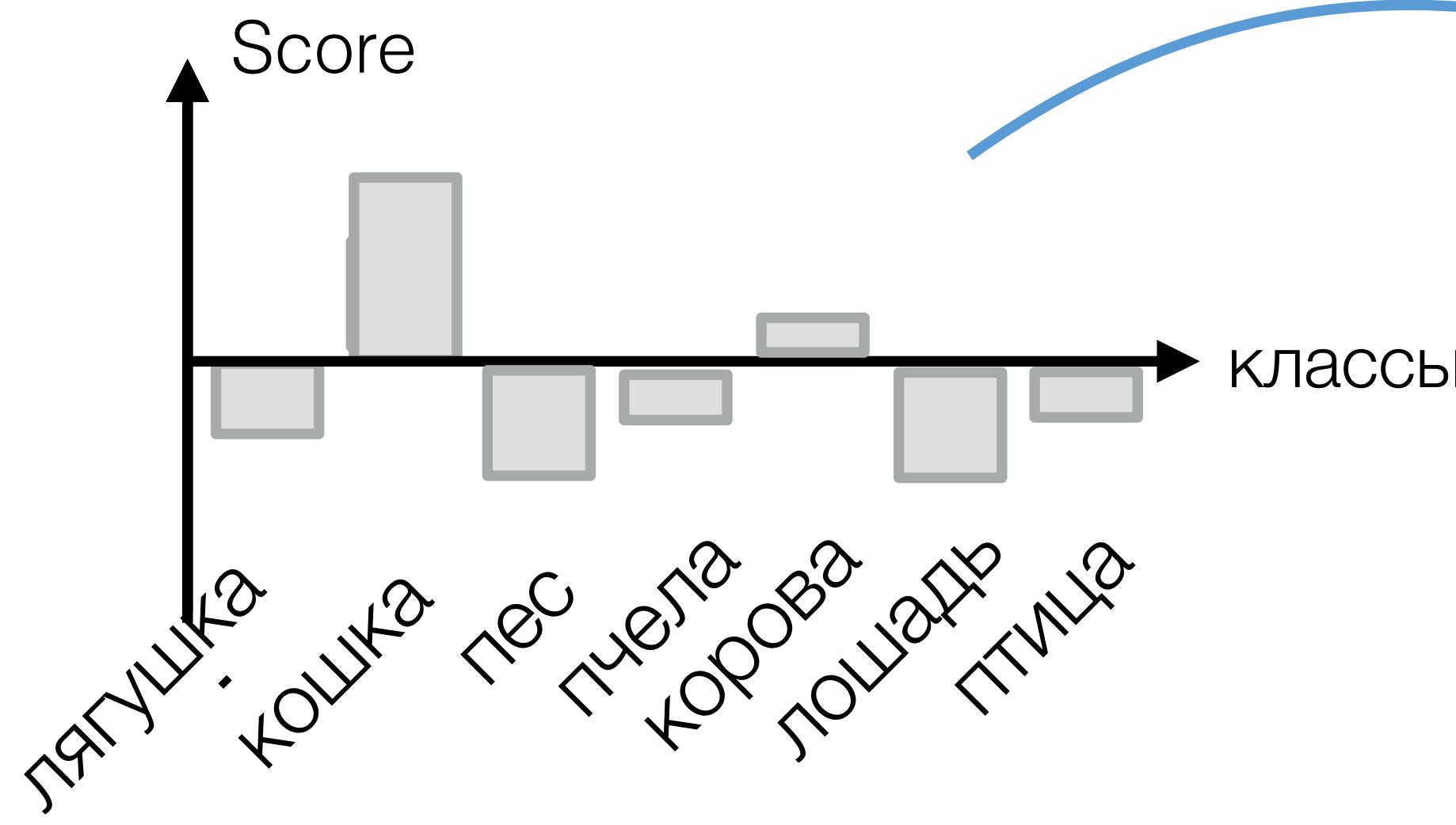
Слой Softmax

Величины z суммируются в 1!

- Трансформирует произвольный набор чисел в вероятности

$$z_i = \frac{\exp(x_i)}{\exp(x_1) + \dots + \exp(x_c)}$$

c — Число классов



План

- Задачи анализа текстов
- Векторные представления слов
- Архитектуры нейронных сетей
 - Рекуррентные нейронные сети
 - Трансформер
- Измерение качества в задачах генерации текста
- Перенос обучения в анализе текстов

План

- Задачи анализа текстов
- Векторные представления слов
- Архитектуры нейронных сетей
 - Рекуррентные нейронные сети
 - Трансформер
- Измерение качества в задачах генерации текста
- Перенос обучения в анализе текстов

Задачи анализа текстов

Дискриминативные

Текст → ?

Классификация текстов

А. С. Пушкин → **литература**
писал стихи

Предсказание заработной
платы по тексту объявления

Стажер в отдел
продаж, опыт не нужен → **25**

Задачи анализа текстов

Дискриминативные

Текст → ?

Классификация текстов

А. С. Пушкин → **литература**
писал стихи

Предсказание заработной платы по тексту объявления

Стажер в отдел
продаж, опыт не нужен → **25**

Генеративные

? → Текст

Автодополнение

Мы поехали в Италию... →
в La Scuole. Там все то же самое.

Генерация подписи к изображению



→ **группа людей
на пляже**

seq2seq

Текст → Текст

Машинный перевод

London is the capital of
Great Britain



**Лондон – столица
Великобритании**

Суммаризация текстов

Этим летом мы поехали в Италию,
мы ели пасту и загорали на пляже,
а по вечерам гуляли вдоль
побережья.



**Этим летом мы
отдыхали в
Италии.**

План

- Задачи анализа текстов
- Векторные представления слов
- Архитектуры нейронных сетей
 - Рекуррентные нейронные сети
 - Трансформер
- Измерение качества в задачах генерации текста
- Перенос обучения в анализе текстов

Предобработка текстов

Лето закончится через девятнадцать дней

2 6 5 0 5

- Словарь наиболее частотных слов
- Редкие слова заменяются на “Неизвестно”
- Можно кодировать по буквам или буквосочетаниям (ВРЕ)

<Неизвестно>	0
я	1
лето	2
мы	3
через	4
дней	5
закончится	6
зима	7

Предобработка текстов

Лето закончится через девятнадцать дней

2 6 5 0 5

- Словарь наиболее частотных слов
- Редкие слова заменяются на “Неизвестно”
- Можно кодировать по буквам или буквосочетаниям (BPE)

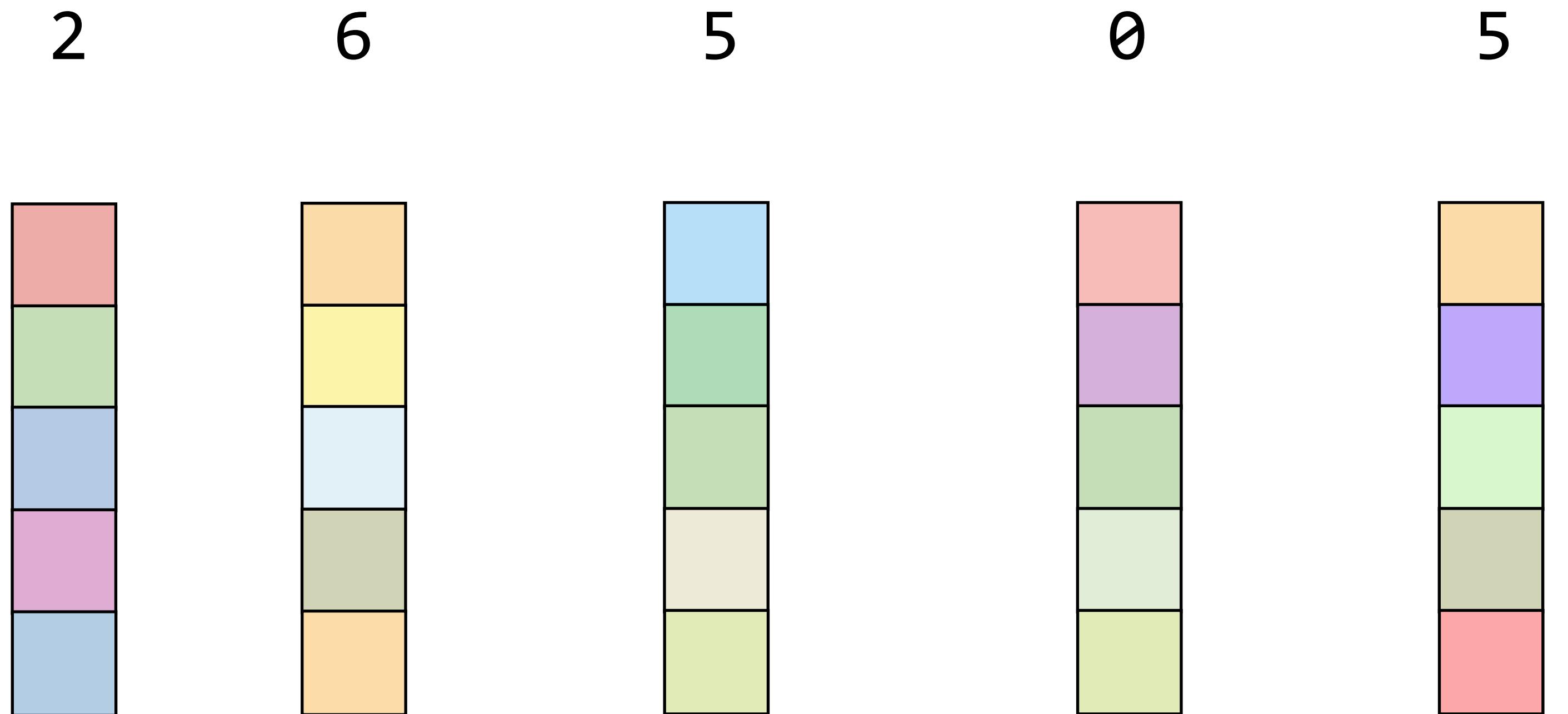
<Неизвестно>	0
я	1
лето	2
мы	3
через	4
дней	5
закончится	6
зима	7

Byte-pair encoding (BPE): Лето закончится через **девят над ца ть** дней

2 6 5 124 10 109 31 5

Векторные представления слов (embeddings)

Лето закончится через девятнадцать дней



размер представления

число слов в словаре

Таблица
векторных
представлений

Визуализация векторных представлений

**номер слова
в словаре**

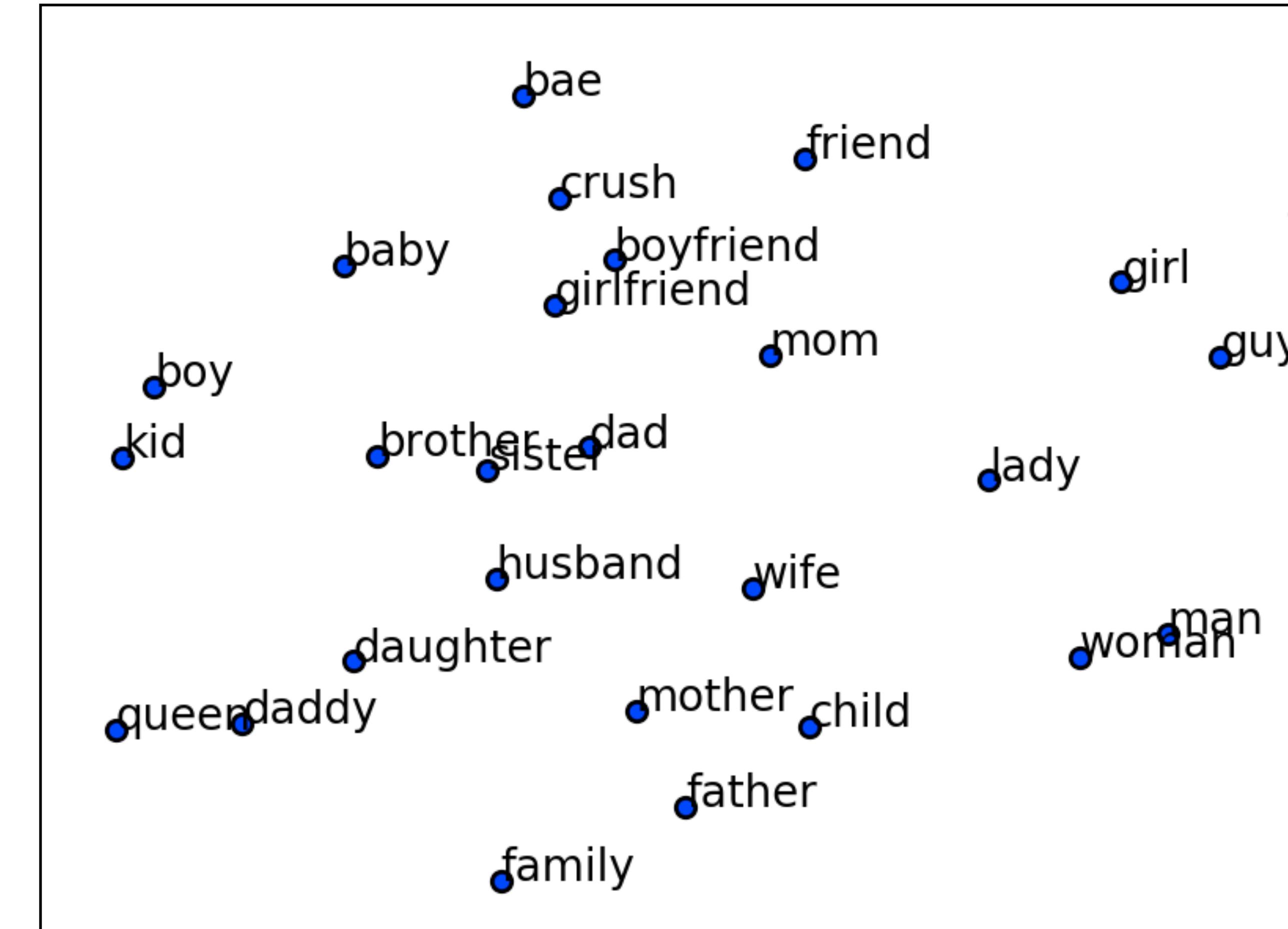
word2vec

**вектор из
256 элементов**

tSNE

**вектор из
2 элементов**

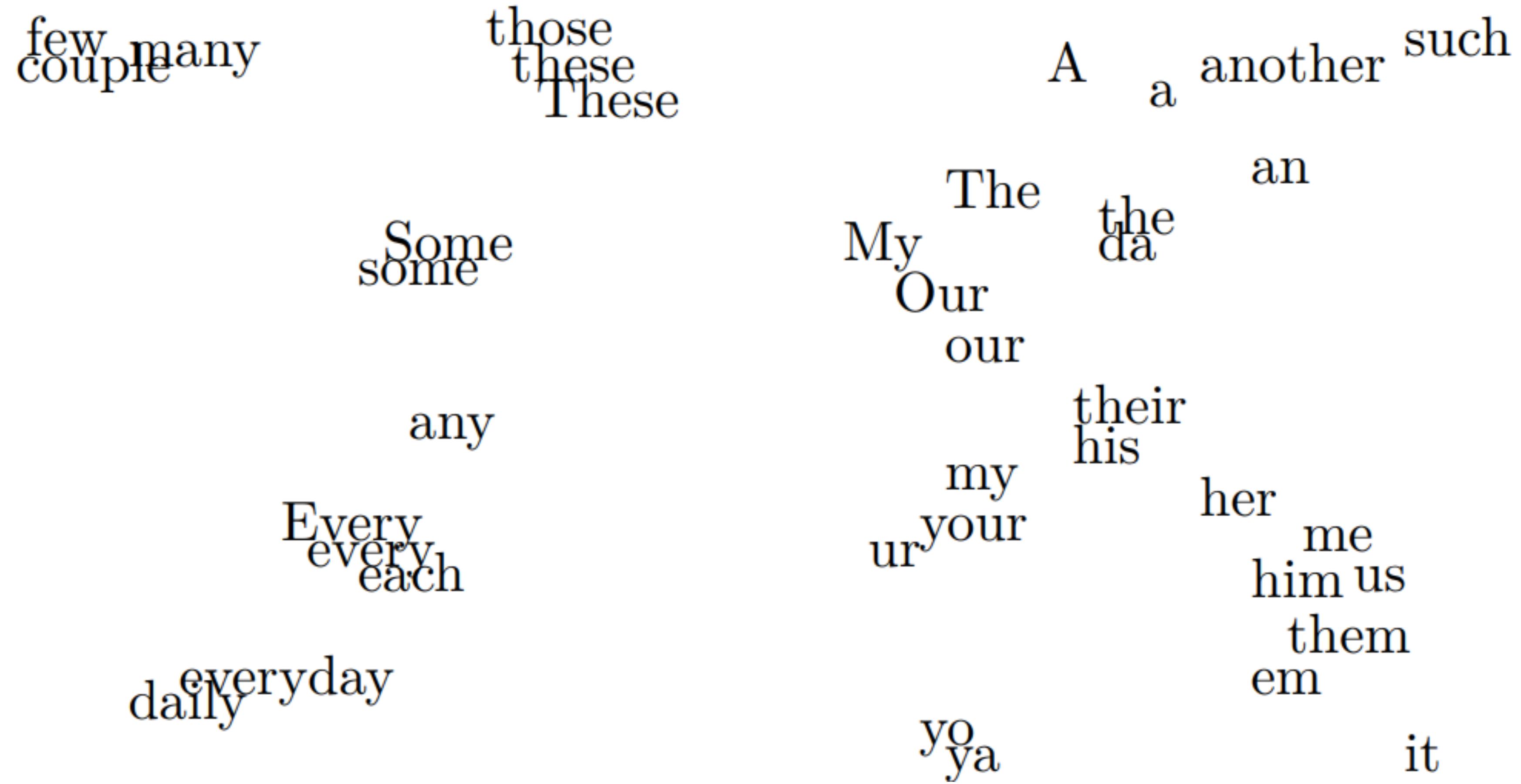
2 элемент
вектора



1 элемент вектора

Визуализация векторных представлений

Данные Twitter:

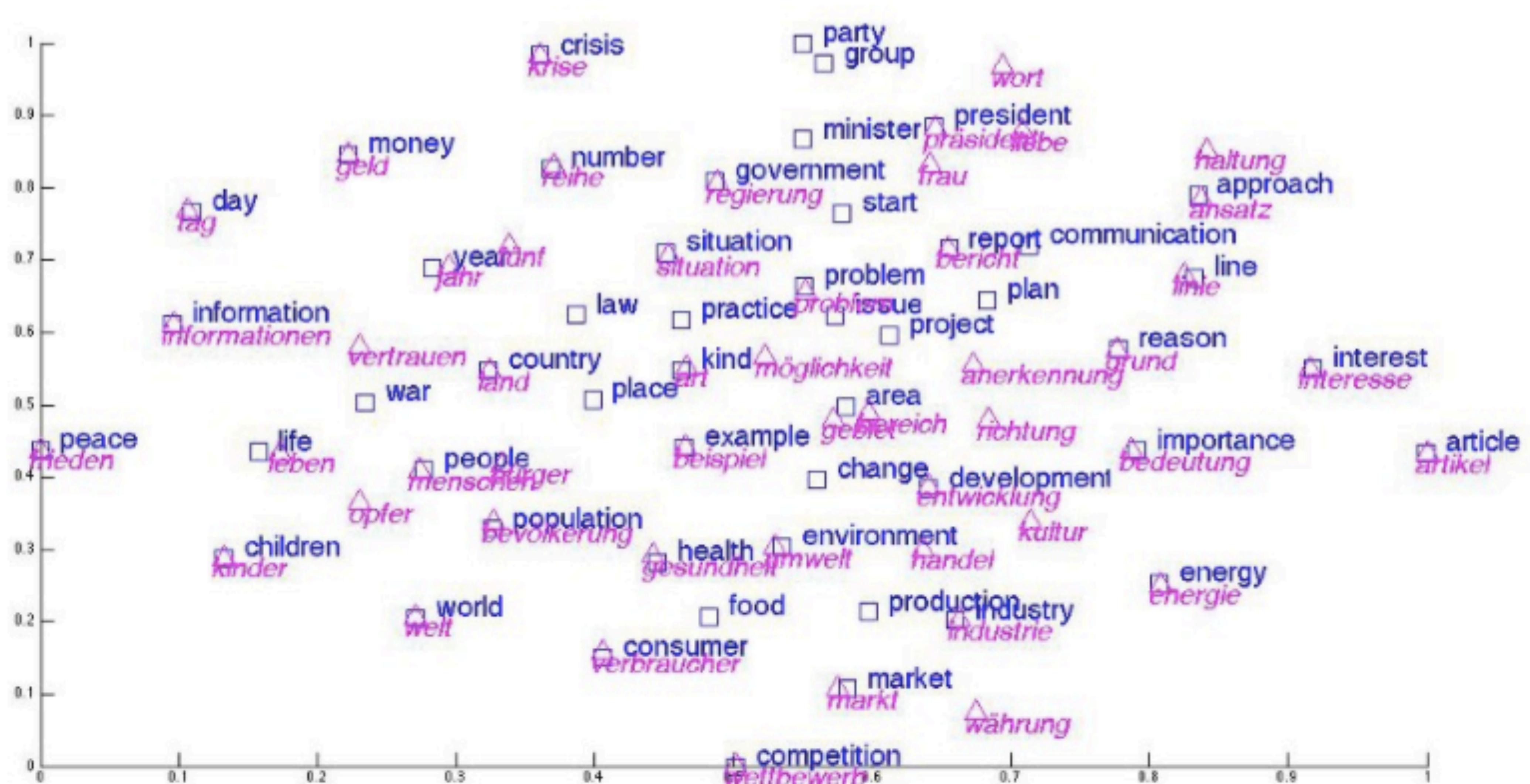


Визуализация векторных представлений

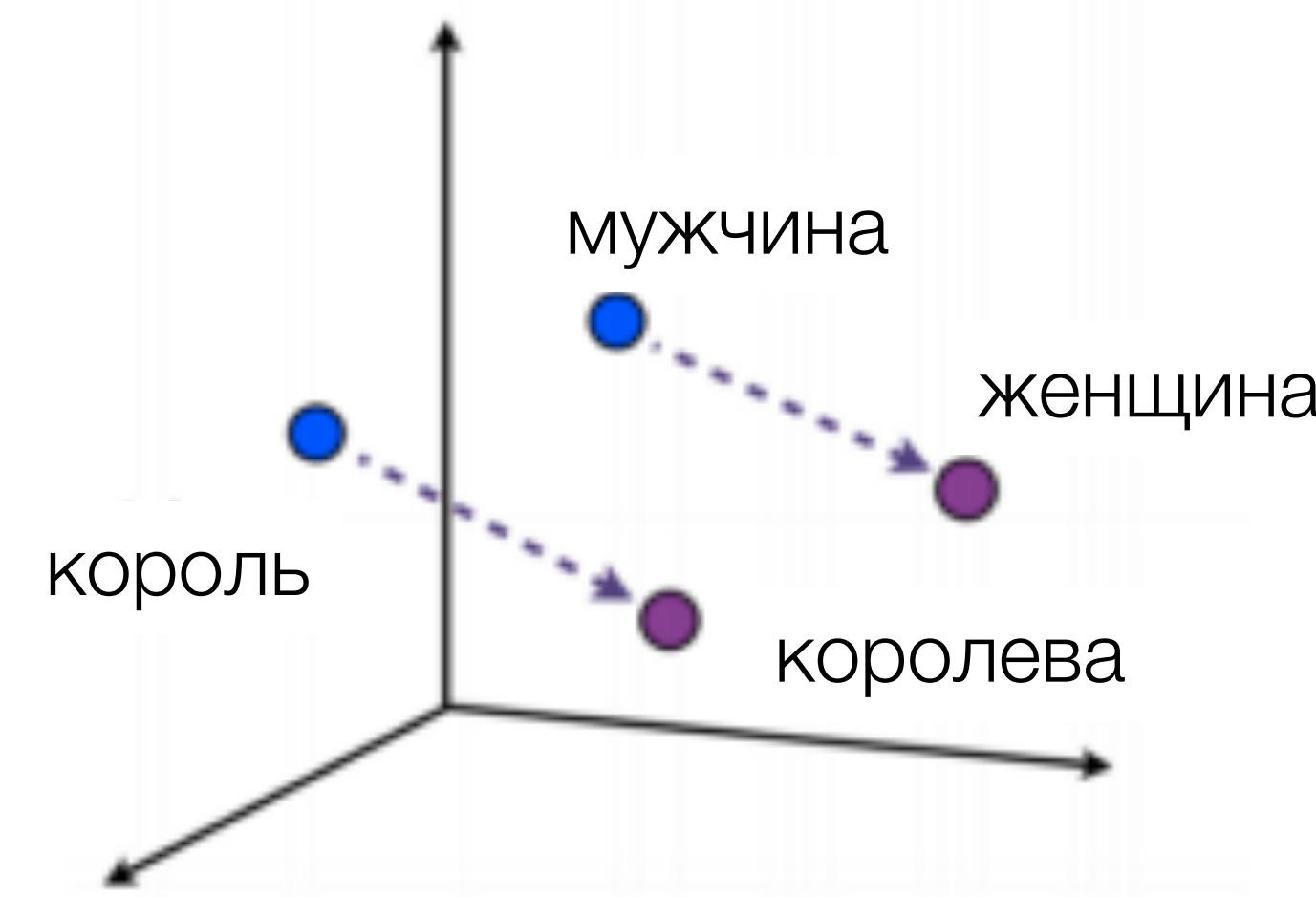
Данные кода программ:



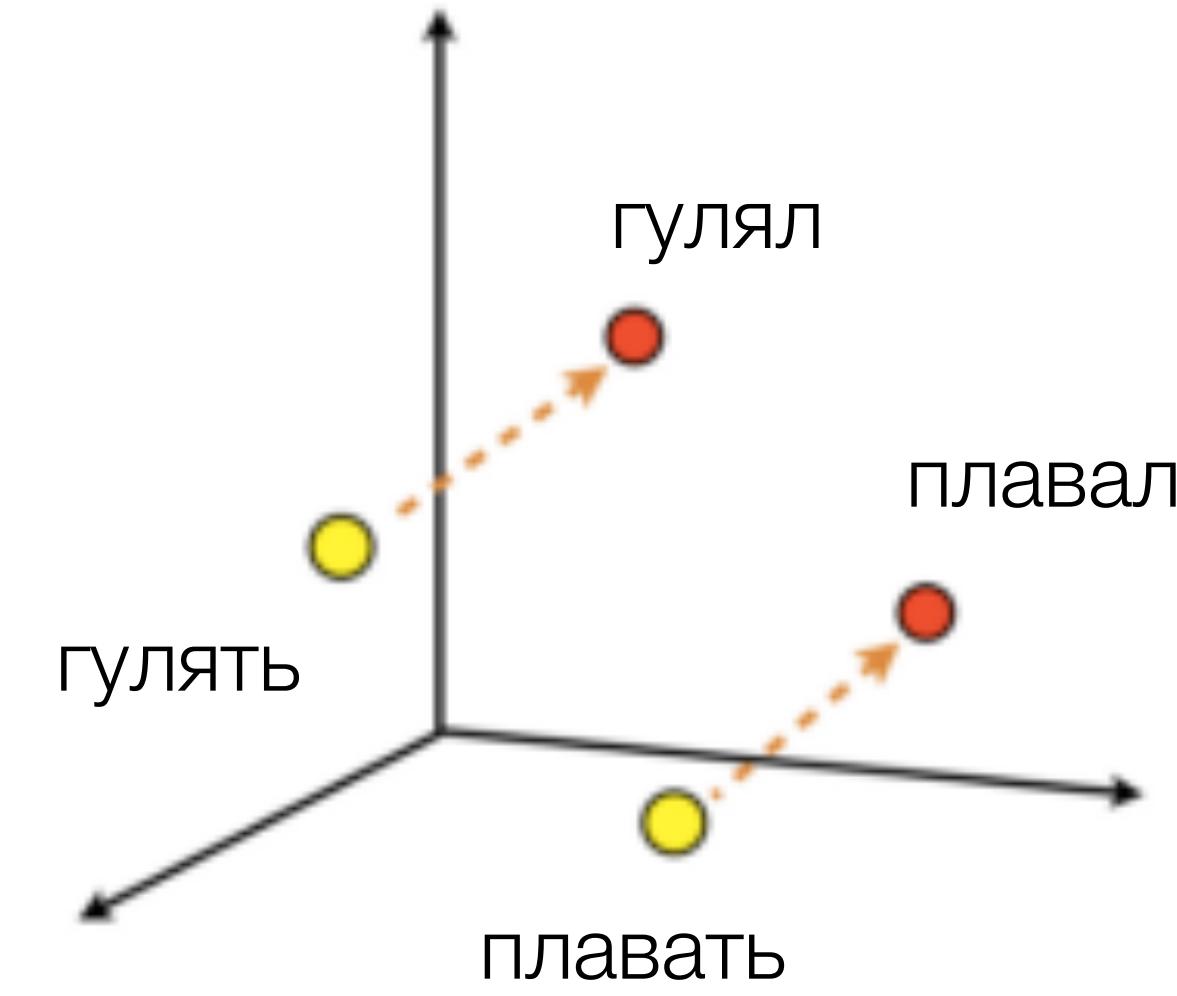
Векторные представления в машинном переводе



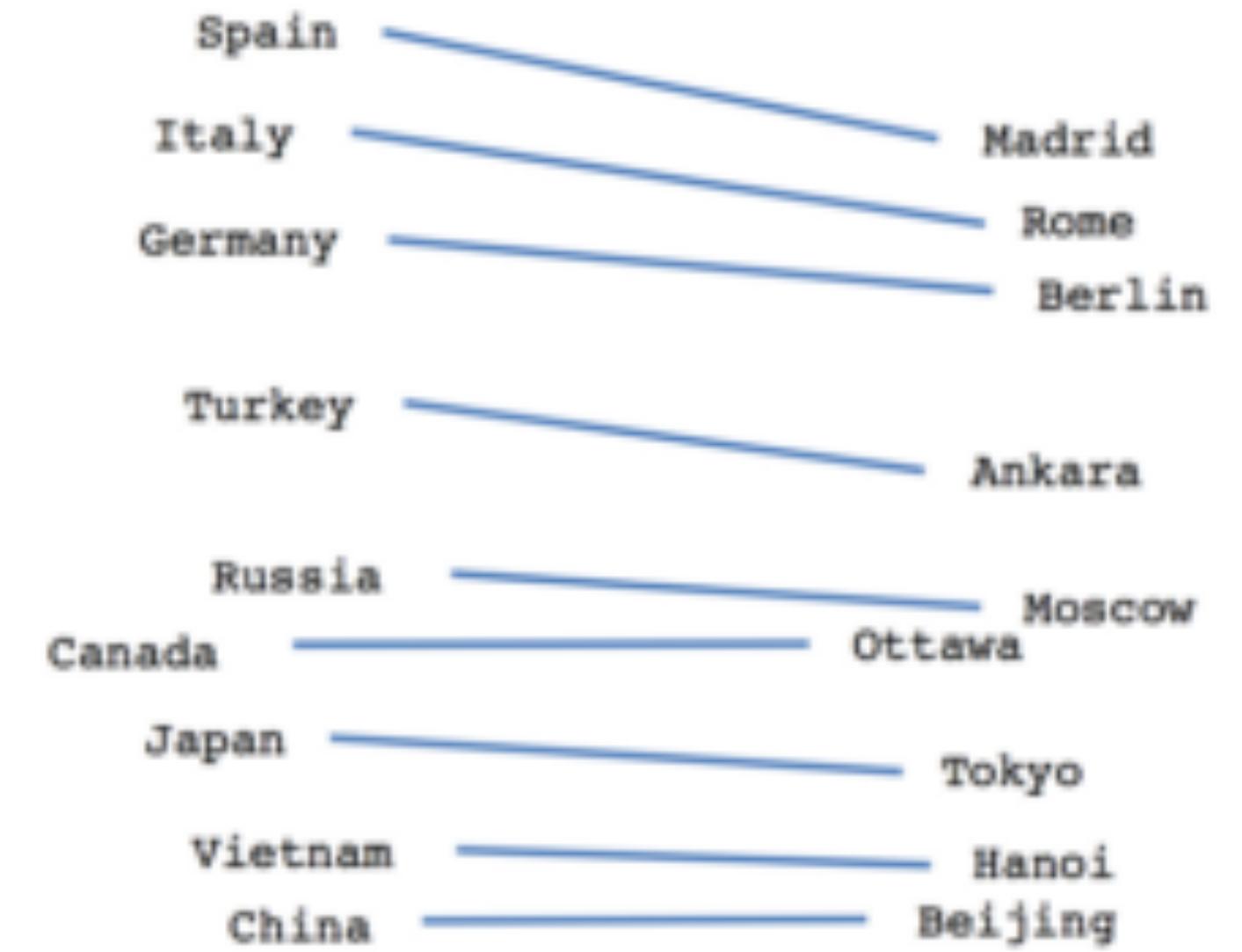
king - man + woman = queen



женский / мужской род



время глагола

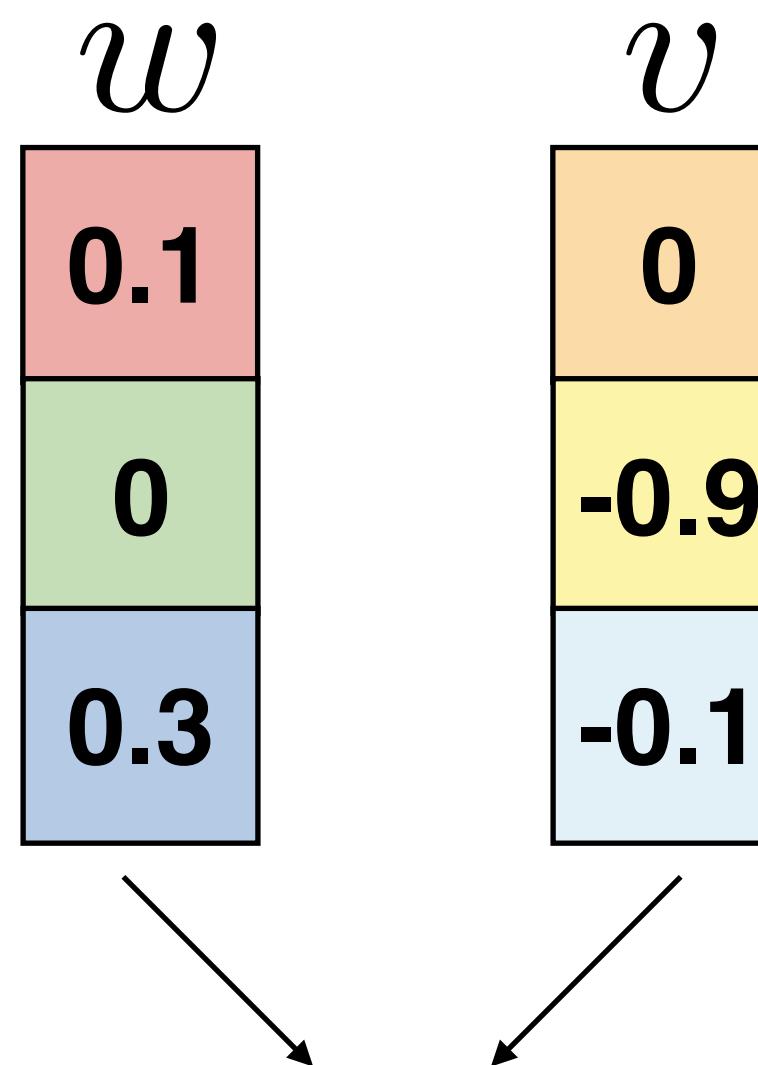


страна — столица

word2vec

Лето закончится

Насколько семантически похожи два слова?



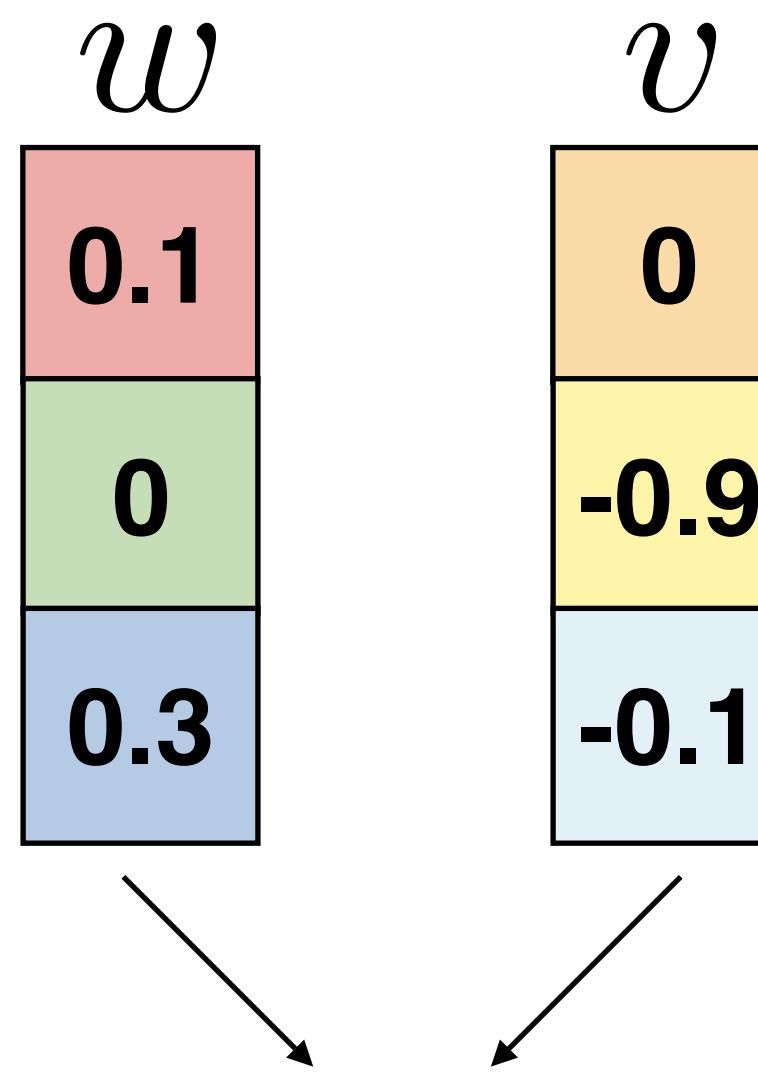
$$\langle w, v \rangle = w_1 \cdot v_1 + \cdots + w_d \cdot v_d$$

$$= 0.1 \cdot 0 - 0 \cdot 0.9 - 0.3 \cdot 0.1 = 0.03$$

word2vec

Лето закончится

Насколько семантически похожи два слова?



похожи $\longrightarrow \langle w, v \rangle$ далеко от 0

$$0.1*0.1 + 0.9*0.8 - 0.5*0.4 = 0.53$$

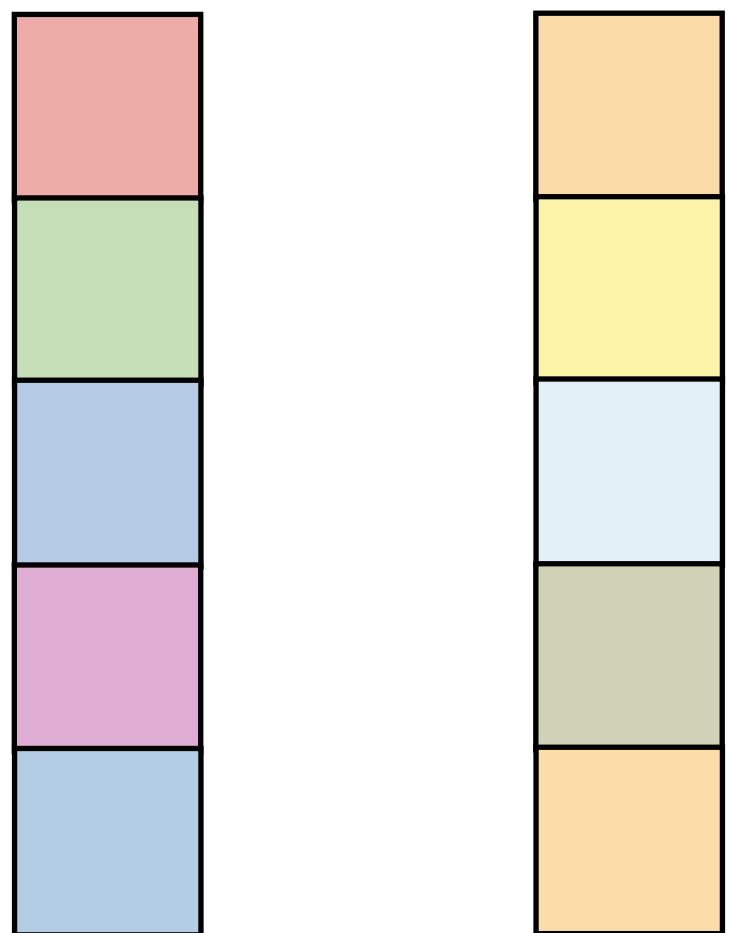
не похожи $\longrightarrow \langle w, v \rangle = 0$

$$\langle w, v \rangle = w_1 \cdot v_1 + \cdots + w_d \cdot v_d$$

$$= 0.1*0 - 0*0.9 - 0.3*0.1 = 0.03$$

word2vec

Лето закончится



<закончится, лето>

**Насколько вероятно
встретить одно слово рядом с другим?**

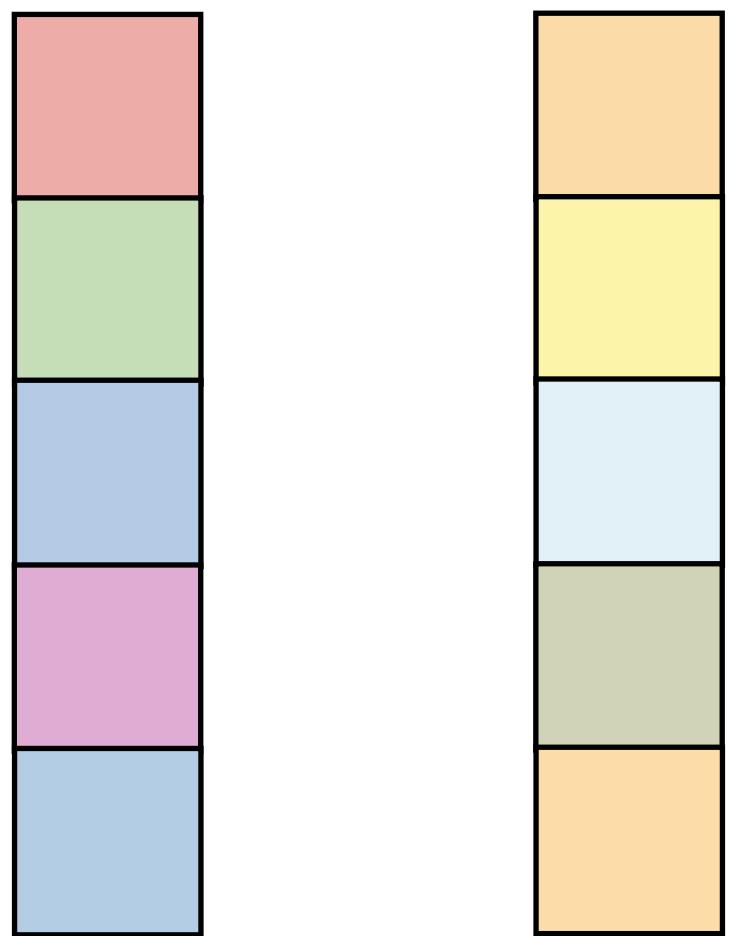
$$P(\text{закончится} \mid \text{лето}) =$$

$$= \frac{\langle \text{закончится}, \text{лето} \rangle}{\langle \text{я, лето} \rangle + \langle \text{через, лето} \rangle + \langle \text{мы, лето} \rangle + \langle \text{лето, лето} \rangle + \dots}$$

все слова из словаря

word2vec

Лето закончится



<закончится, лето>

**Насколько вероятно
встретить одно слово рядом с другим?**

$P(\text{закончится} \mid \text{лето}) =$

$$= \frac{\exp(\langle\text{закончится, лето}\rangle)}{\exp(\langle\text{я, лето}\rangle) + \exp(\langle\text{через, лето}\rangle) + \exp(\langle\text{мы, лето}\rangle) + \exp(\langle\text{лето, лето}\rangle) + \dots}$$

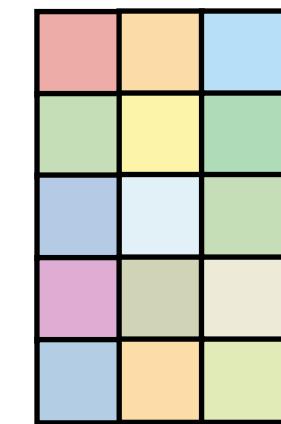
все слова из словаря

word2vec

Лето **закончится** ? **девятнадцать** дней

контекст

$\log P(\text{через} \mid \text{закончится}) + \log P(\text{через} \mid \text{девятнадцать}) \longrightarrow \max$

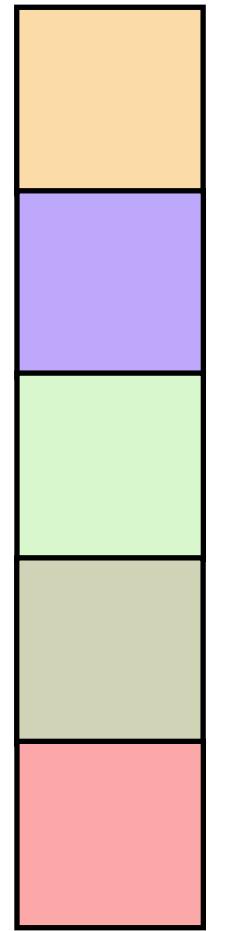
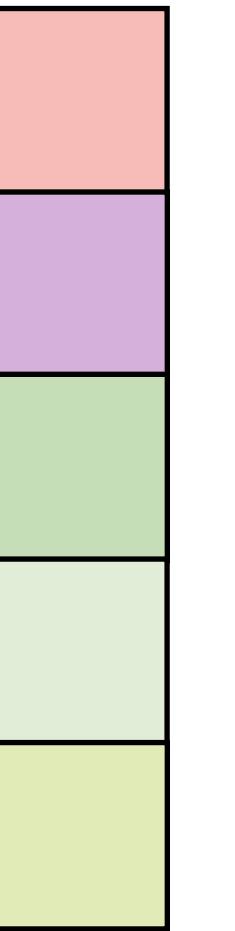
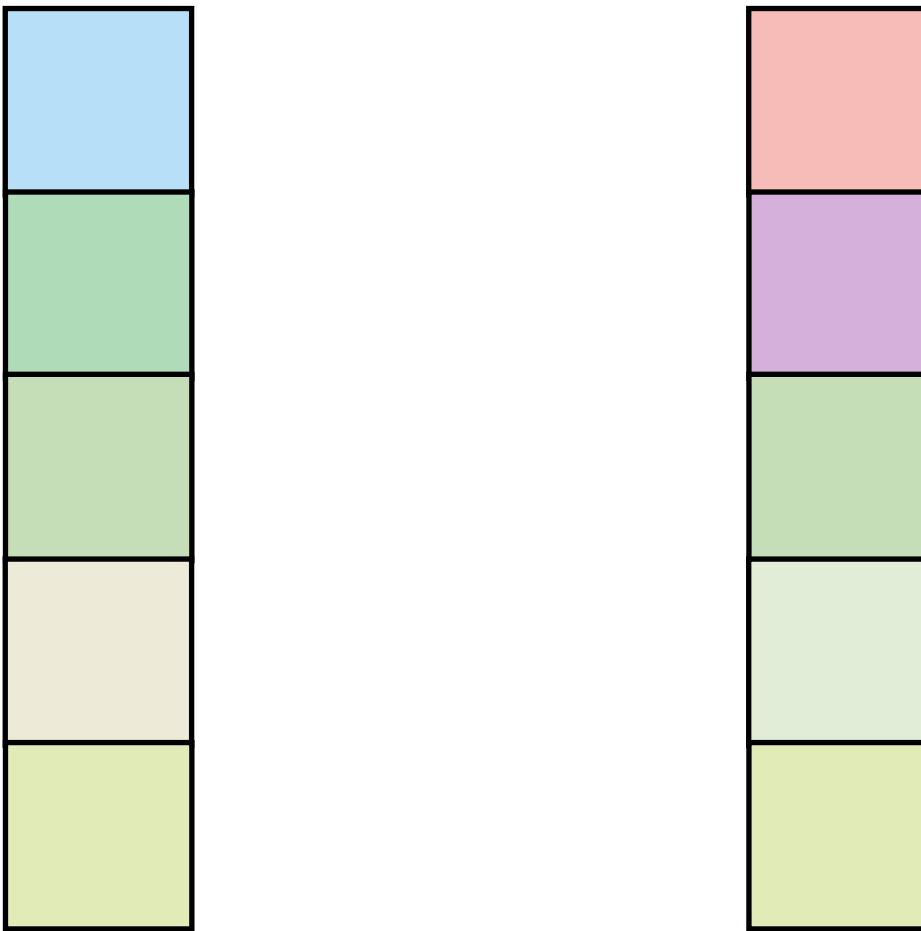
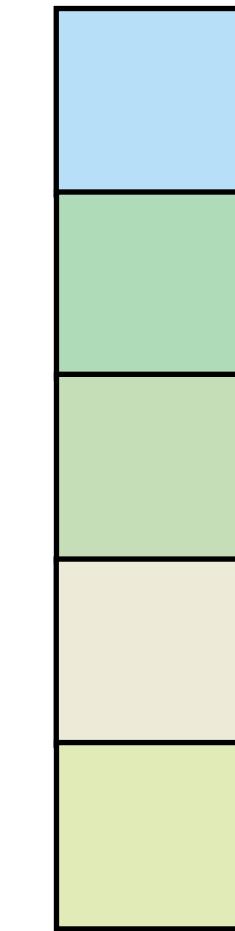
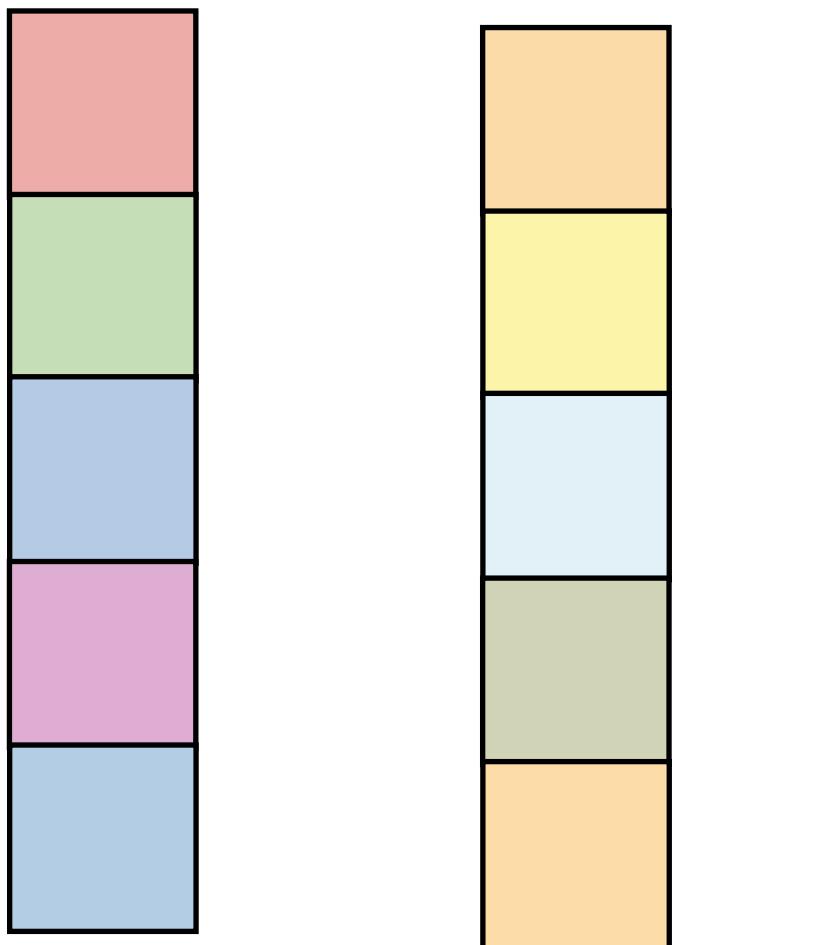


Простейшая нейронная сеть (вычислительный граф)!

Слой векторных представлений

Проход вперед:

В фильме замечательно играют актеры



число слов в словаре

Параметры:

размер представления

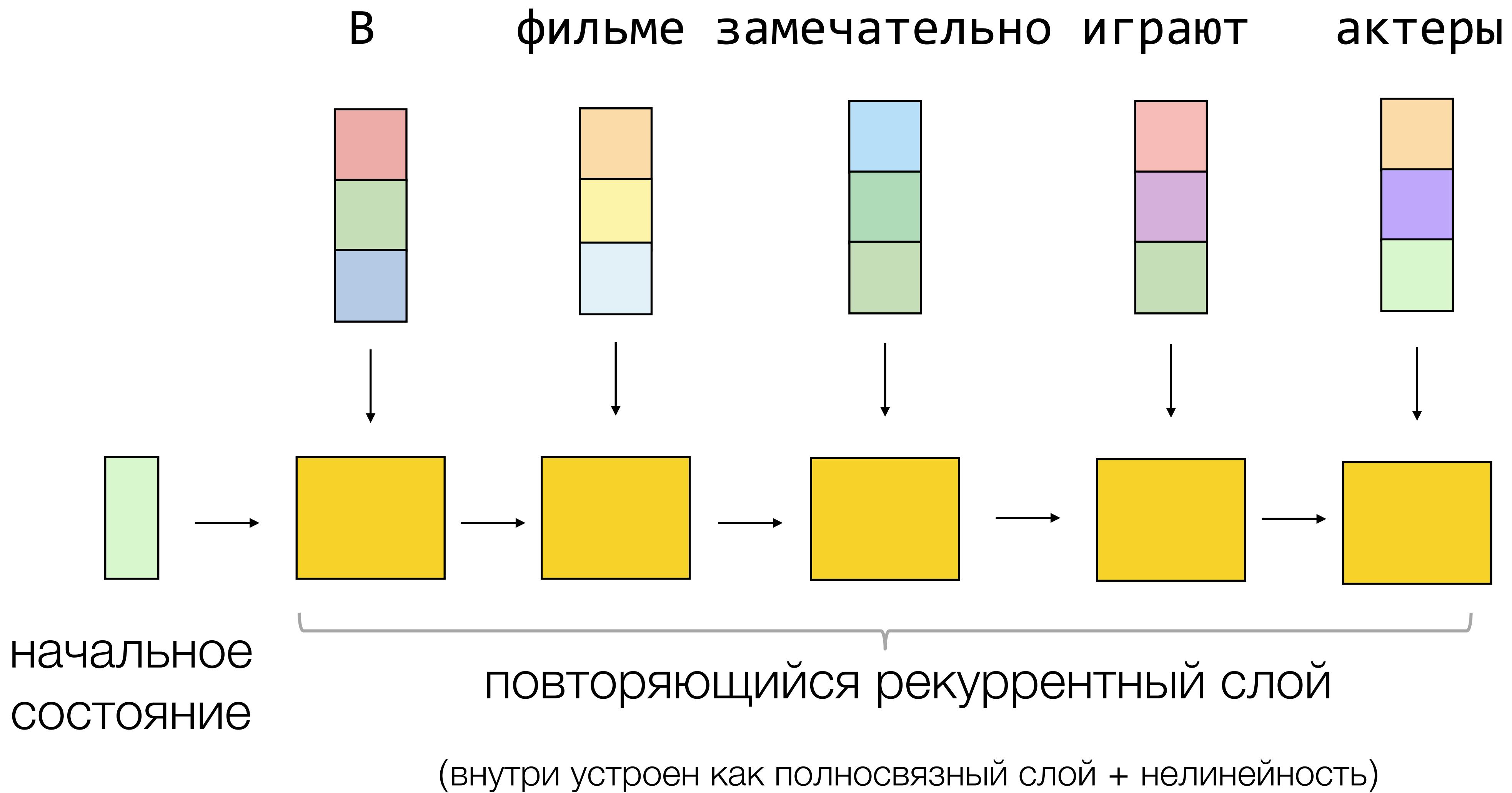


Таблица
векторных
представлений

План

- Задачи анализа текстов
- Векторные представления слов
- Архитектуры нейронных сетей
 - Рекуррентные нейронные сети
 - Трансформер
- Измерение качества в задачах генерации текста
- Перенос обучения в анализе текстов

Рекуррентная нейронная сеть

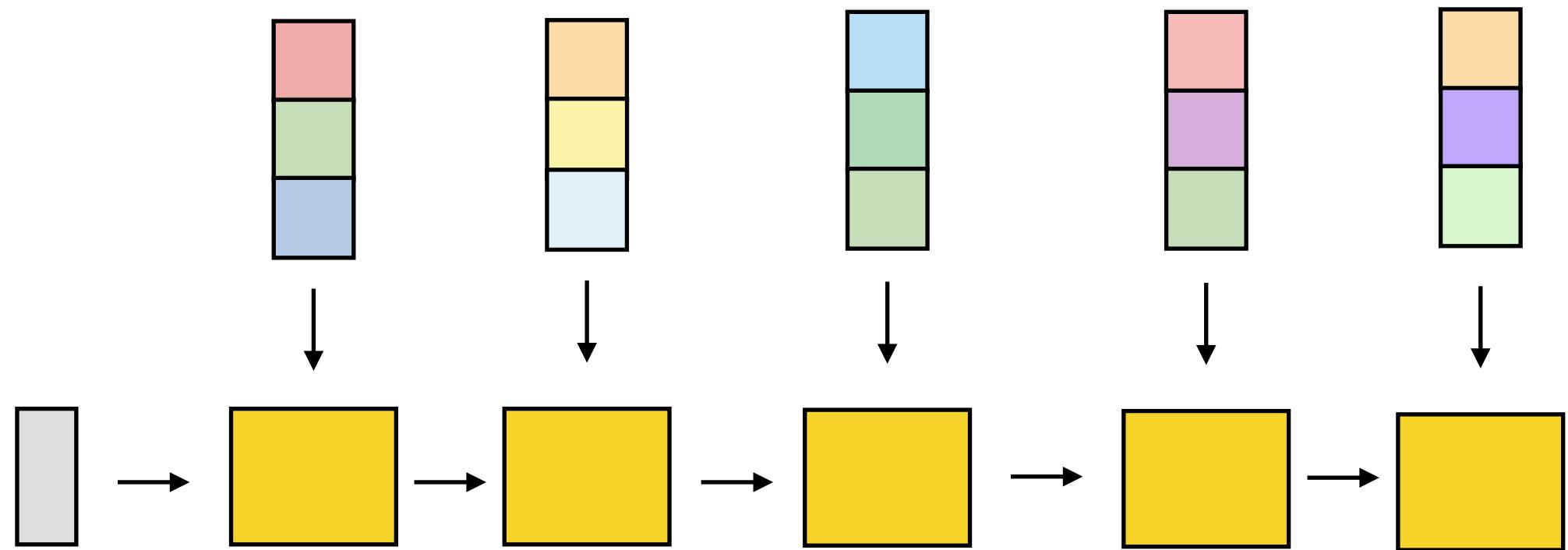


Рекуррентная нейронная сеть

$$h_0 = h_{init}$$

$$h_i = \sigma(W^h h_{i-1} + W^x x_i + b)$$

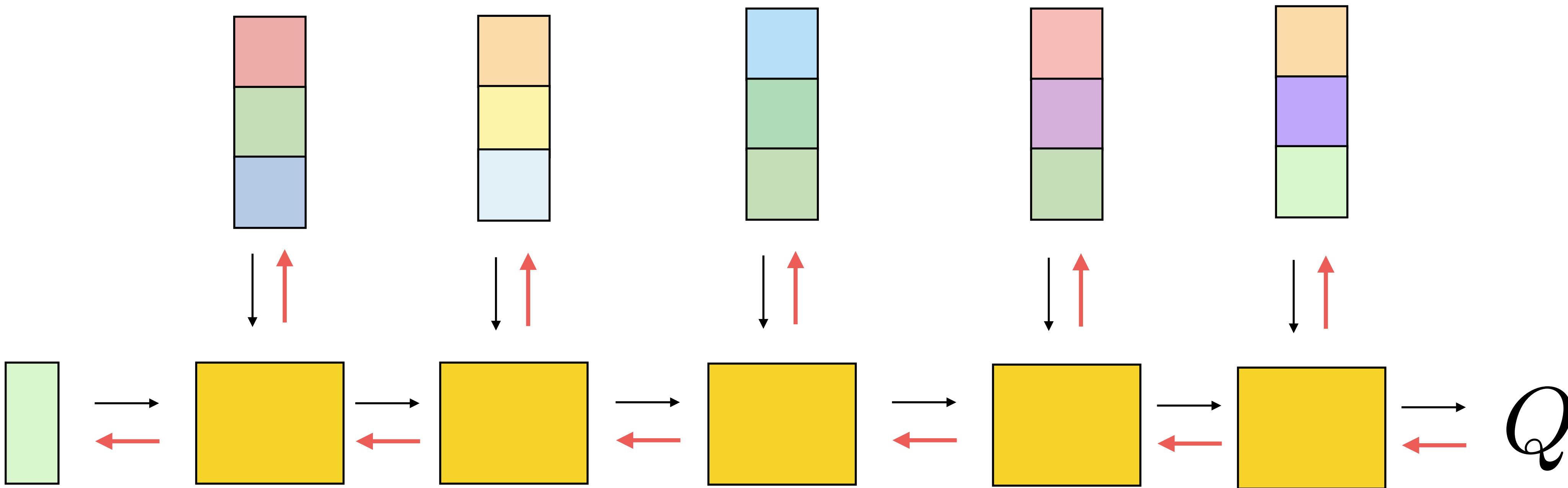
$$i = 1, \dots, L$$



LSTM / GRU: тоже переход от h_t к h_{t+1} , но более сложный механизм
Разработаны, чтобы избежать проблемы затухающих градиентов

Рекуррентная нейронная сеть: проход назад

в фильме замечательно играют актеры

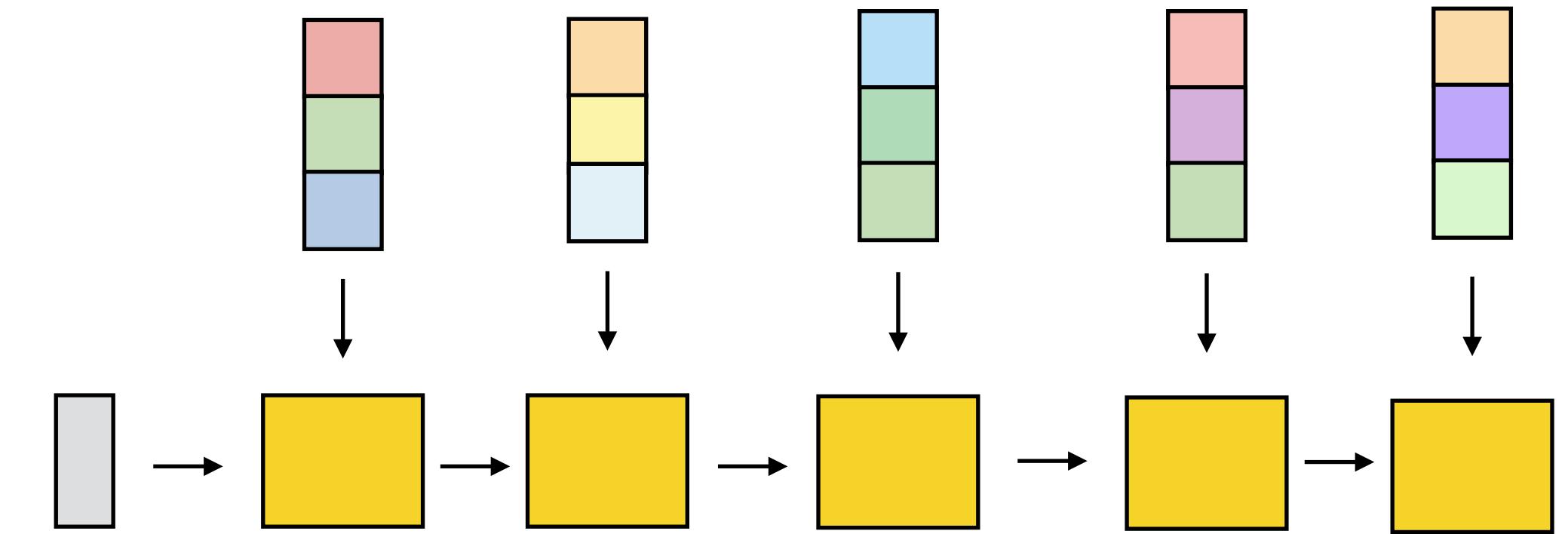


Градиенты по параметрам слоя
и по времени

Рекуррентные нейронные сети

Плюс:

- обрабатывают последовательности произвольной длины (хотя на практике обычно все равно ограничивают)

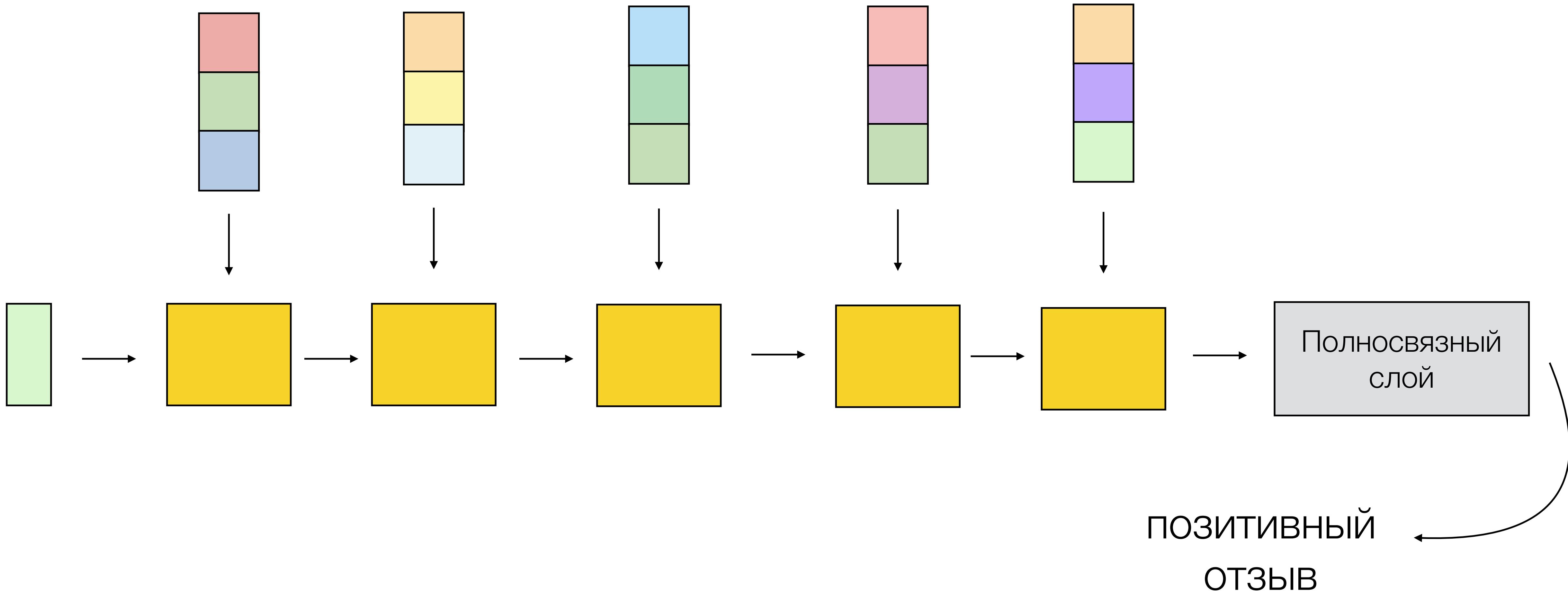


Минус:

- проблема затухающих и взрывающихся градиентов (сложно находить “длинные” зависимости)

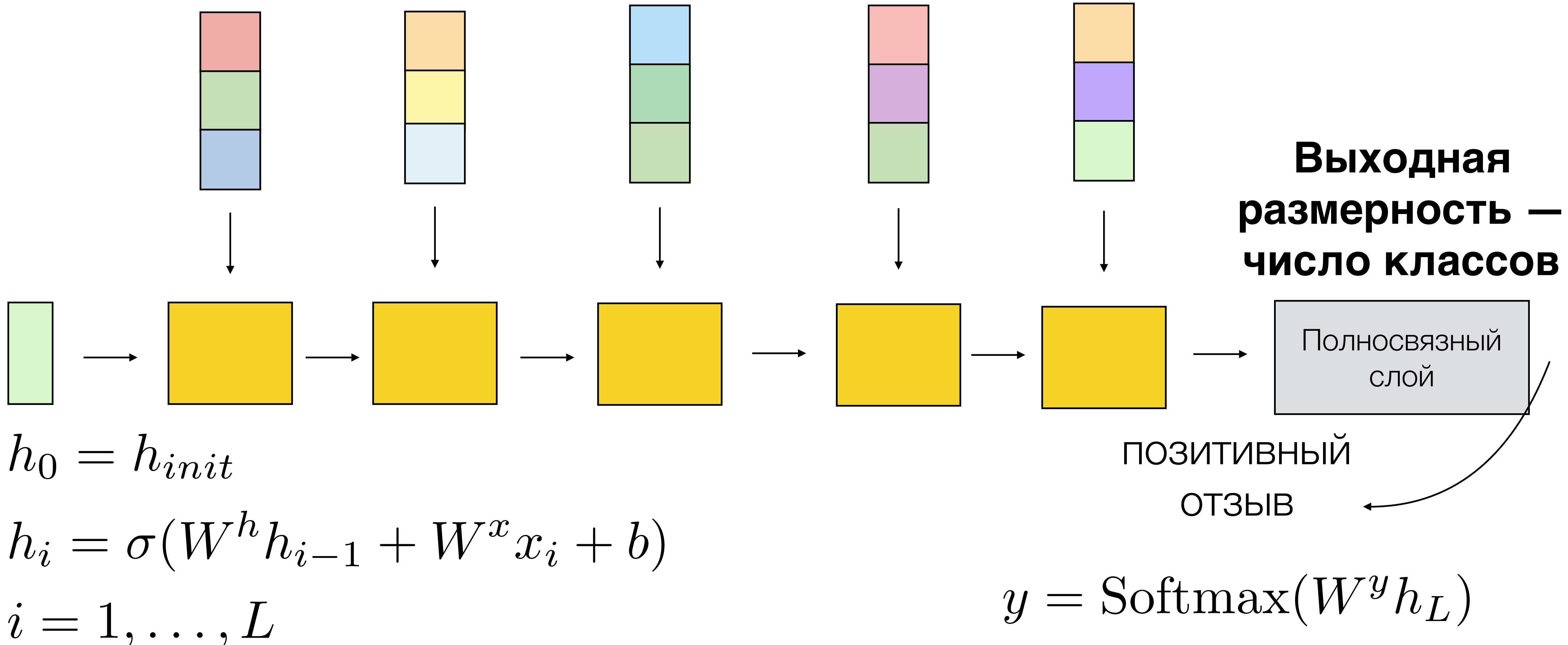
Рекуррентная нейросеть в задаче классификации

В фильме замечательно играют актеры



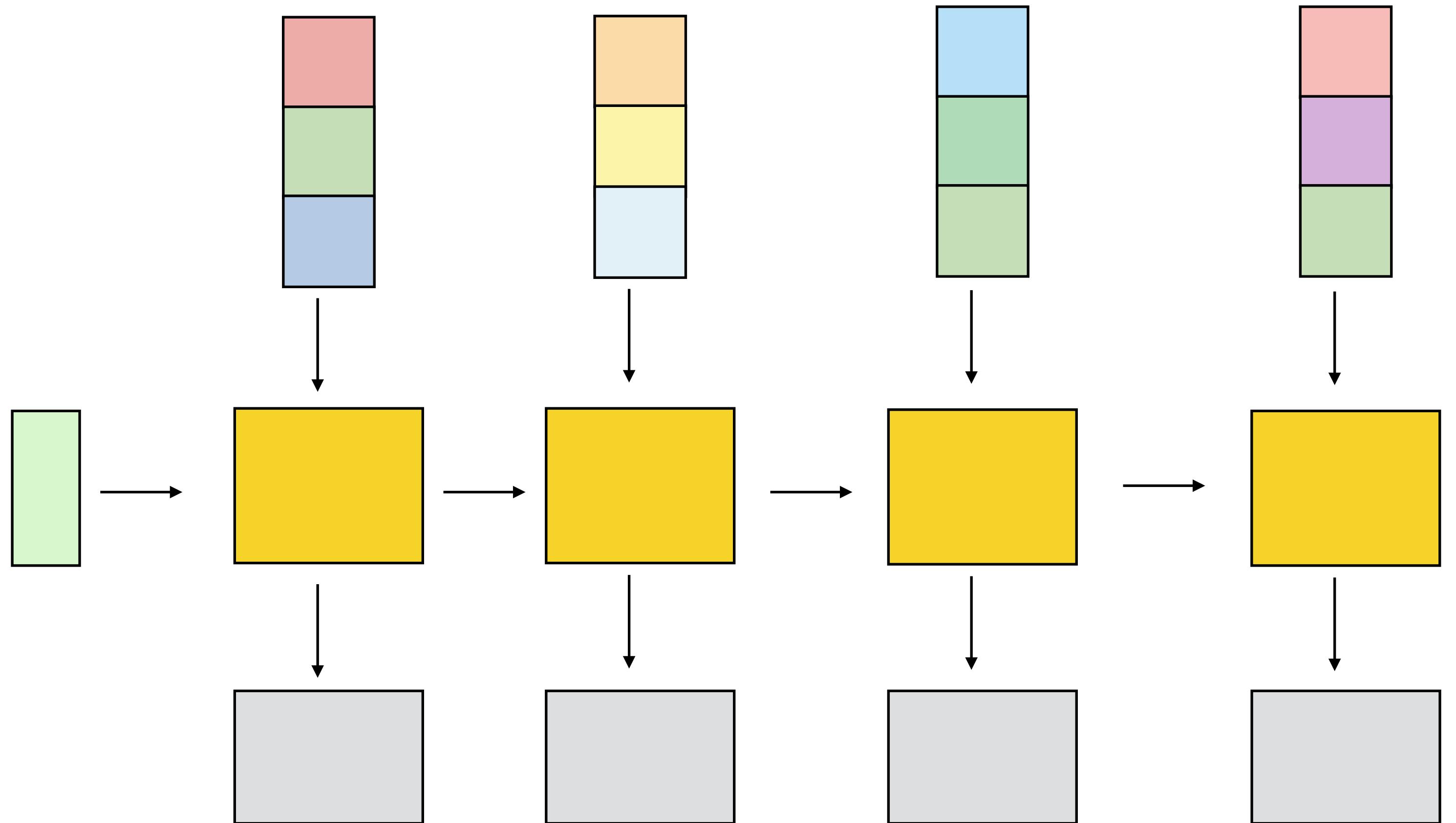
Рекуррентная нейросеть в задаче классификации

В фильме замечательно играют актеры



Рекуррентная нейросеть в задаче генерации

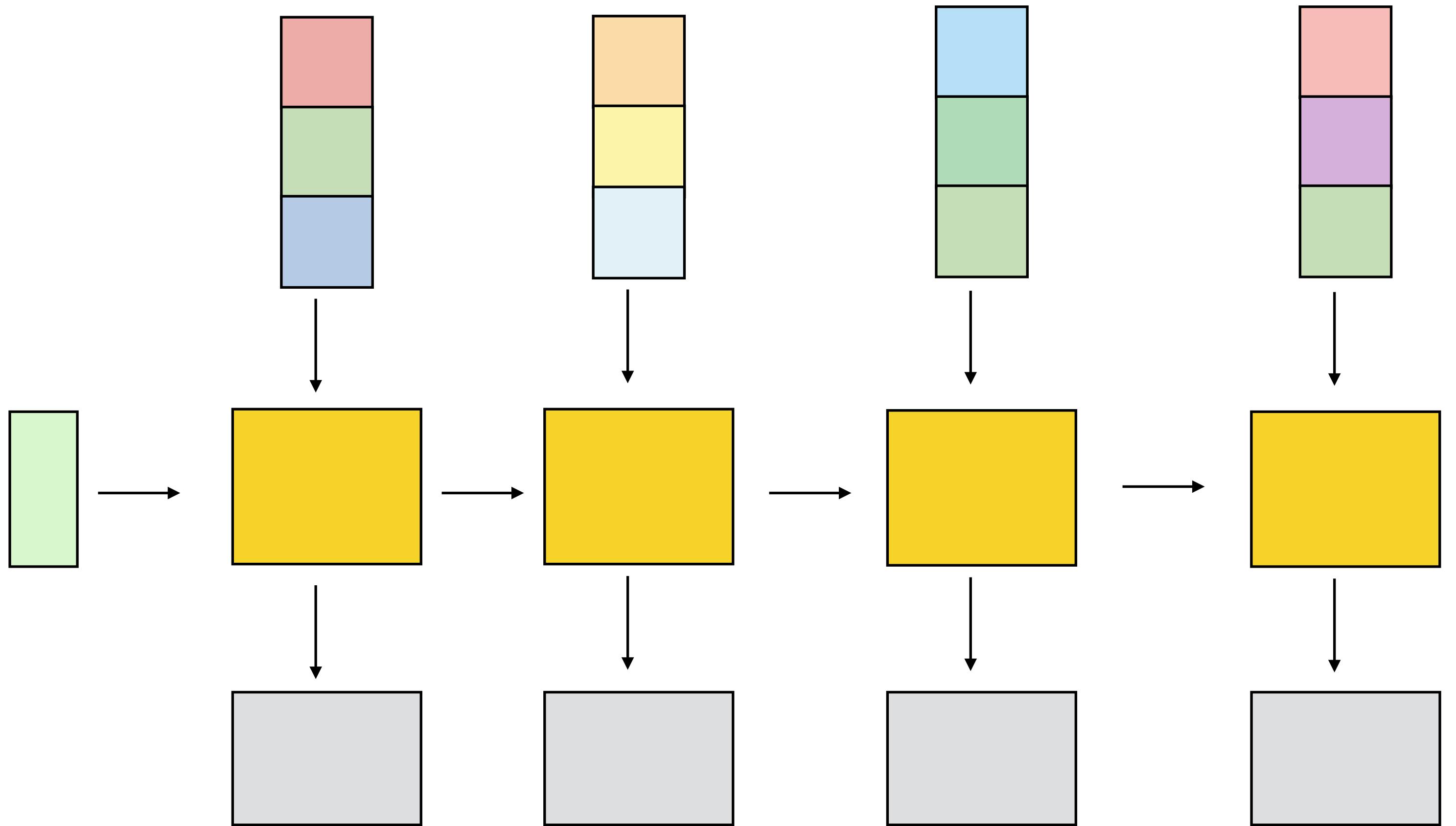
В фильме замечательно играют ?



фильме замечательно играют **актеры**

Рекуррентная нейросеть в задаче генерации

В фильме замечательно играют



$$h_0 = h_{init}$$

$$h_i = \sigma(W^h h_{i-1} + W^x x_i + b)$$

$$i = 1, \dots, L$$

$$y_i = \text{Softmax}(W^y h_i)$$

**Выходная
размерность —
размер словаря**

фильме замечательно играют

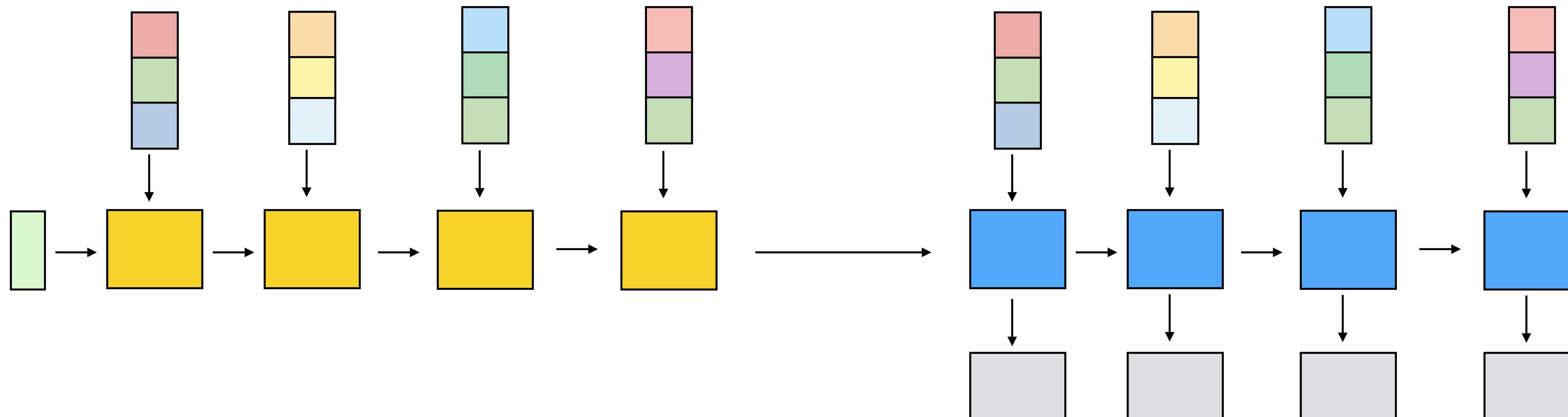
актеры

Рекуррентная нейросеть в задаче перевода

Кодировщик
(encoder)

Декодировщик
(decoder)

London is the capital... <начало> Лондон – столица



Лондон – столица Великобритании

Примеры: генерация текста

Мы поехали в Италию. Мы жили в центре Рима. Однажды я пошла в церковь Святого Петра и... и вдруг поняла, что все это раньше видела во сне.

Сегодня на занятии мы будем изучать рекуррентные нейронные сети. Я расскажу о том, как мы строим нейронные цепи, используя две стойки памяти, связанные друг с другом и являющиеся двумя параллельно движущимися контур|

Мороз и солнце, день чудесный. Приносит мне лучик света. Но не как Мария, а как бы через уста.|

Примеры: машинный перевод

The screenshot shows a machine translation interface with two panels. The left panel has "ENGLISH" selected as the source language and "RUSSIAN" as the target language. The right panel has "RUSSIAN" selected as the source language and "ENGLISH" as the target language. A central double-headed arrow icon indicates the bidirectional nature of the comparison.

English Panel:

"What to do about mass unemployment? This is going to be a massive social challenge. There will be fewer and fewer jobs that a robot cannot do better [than a human]. These are not things that I wish will happen. These are simply things that I think probably will happen." –Elon Musk

Russian Panel:

«Что делать с массовой безработицей? Это будет серьезной социальной проблемой. Будет все меньше и меньше работ, которые робот не сможет выполнить лучше, чем человек. Это не то, что я желаю, чтобы это произошло. Это просто вещи, которые, я думаю, вероятно, произойдут ». - Элон Маск

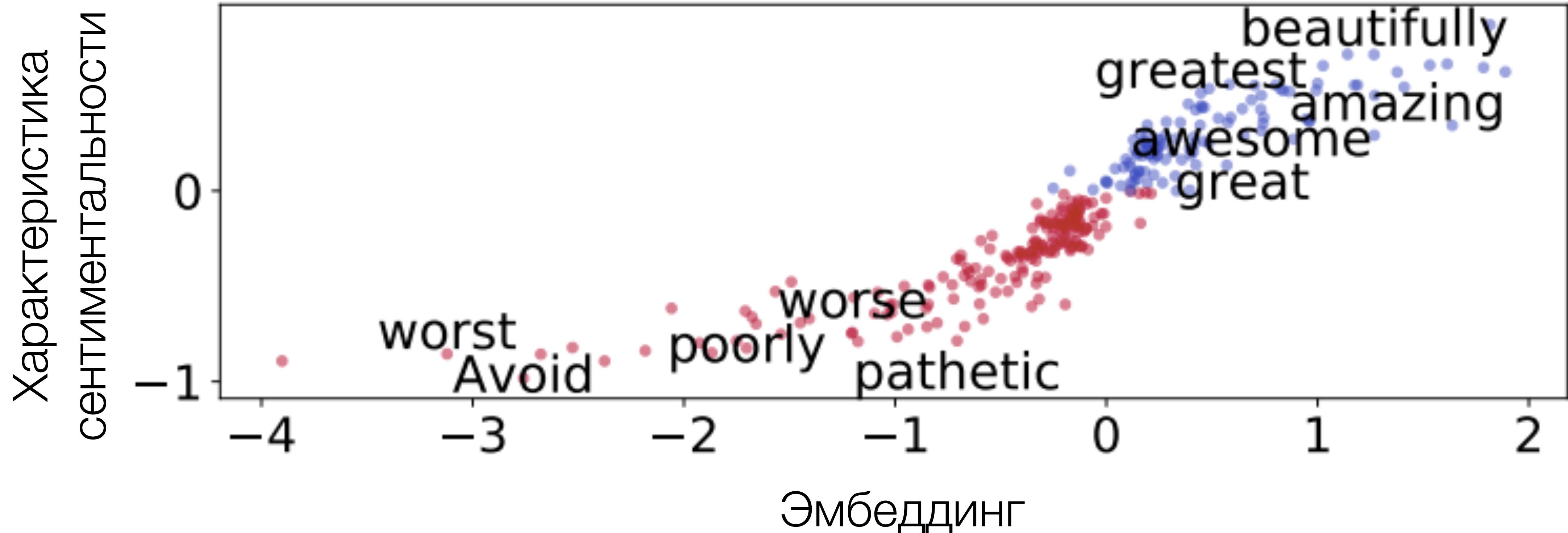
Below the Russian text, there is a transcription in Russian: «Chto delat' s massovoy bezrabititsey? Eto budet ser'yeznay sotsial'noy problemoy. Budet vse men'she i men'she rabot, kotoryye robot ne smozhet vypolnit' luchshe, chem chelovek. Eto ne to, chto ya zhelayu, chtoby eto proizoshlo. Eto prosto veshchi, kotoryye, ya dumayu,

[Show more](#)

At the bottom of each panel are various interaction icons: microphone, speaker, text input, and sharing options.

Но это уже Трансформер!

Визуализация векторных представлений в задаче сентимент-анализа

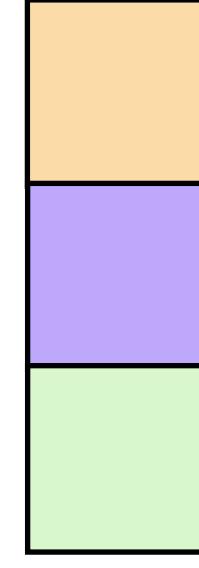
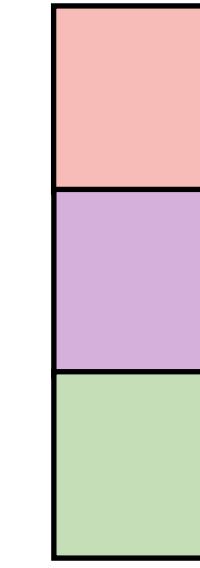
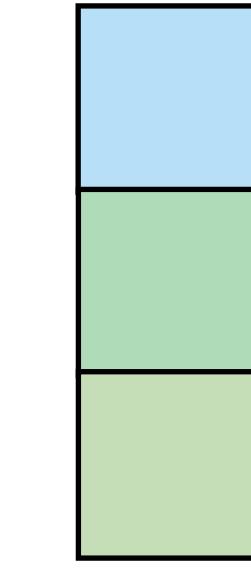
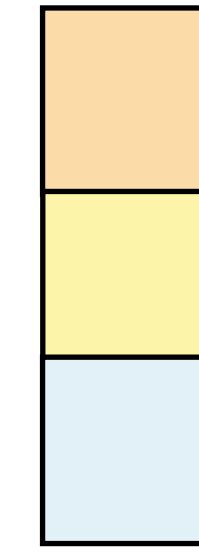


План

- Задачи анализа текстов
- Векторные представления слов
- Архитектуры нейронных сетей
 - Рекуррентные нейронные сети
 - Трансформер
- Измерение качества в задачах генерации текста
- Перенос обучения в анализе текстов

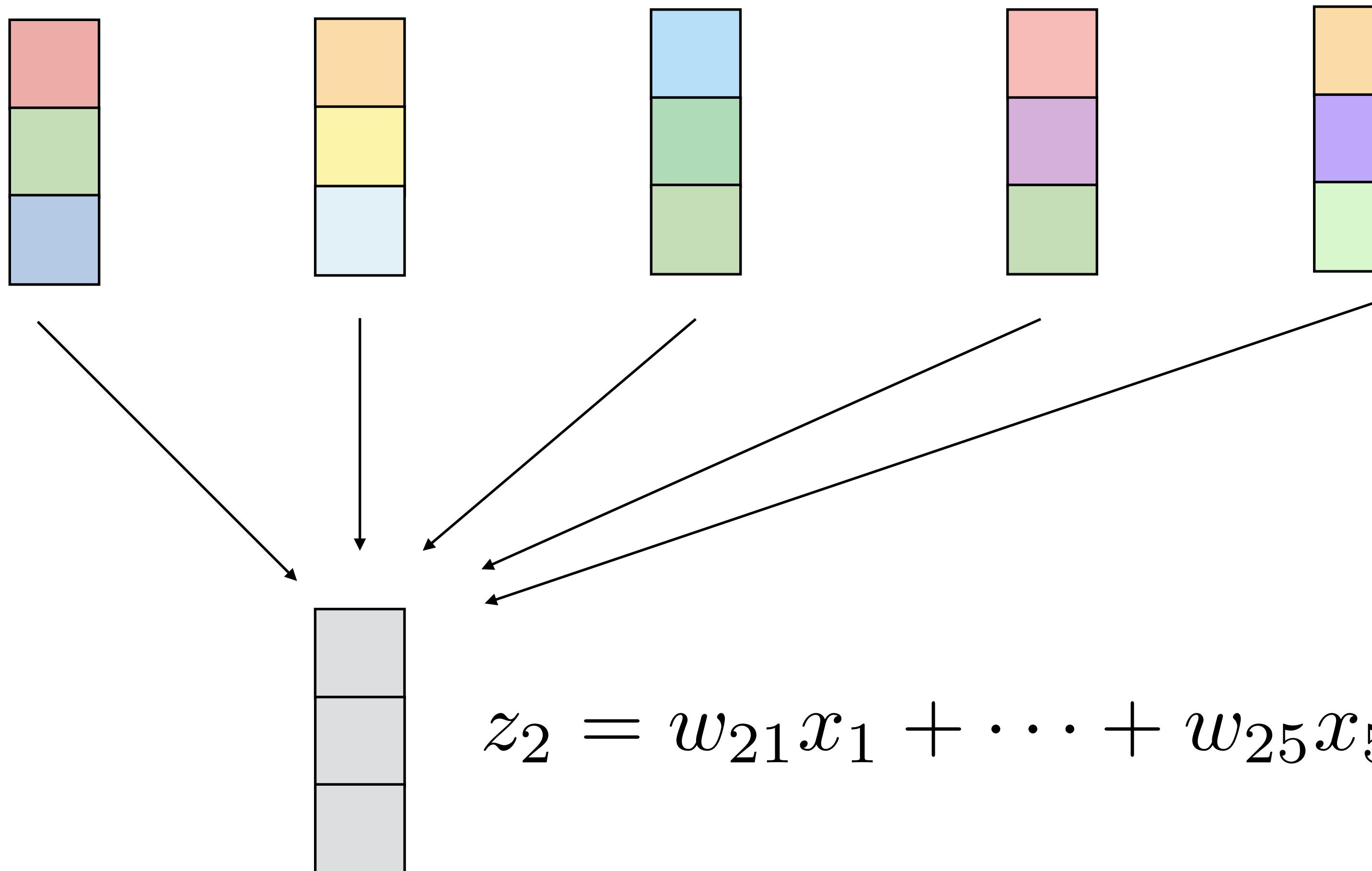
Трансформер: механизм self-attention

В фильме замечательно играют актеры



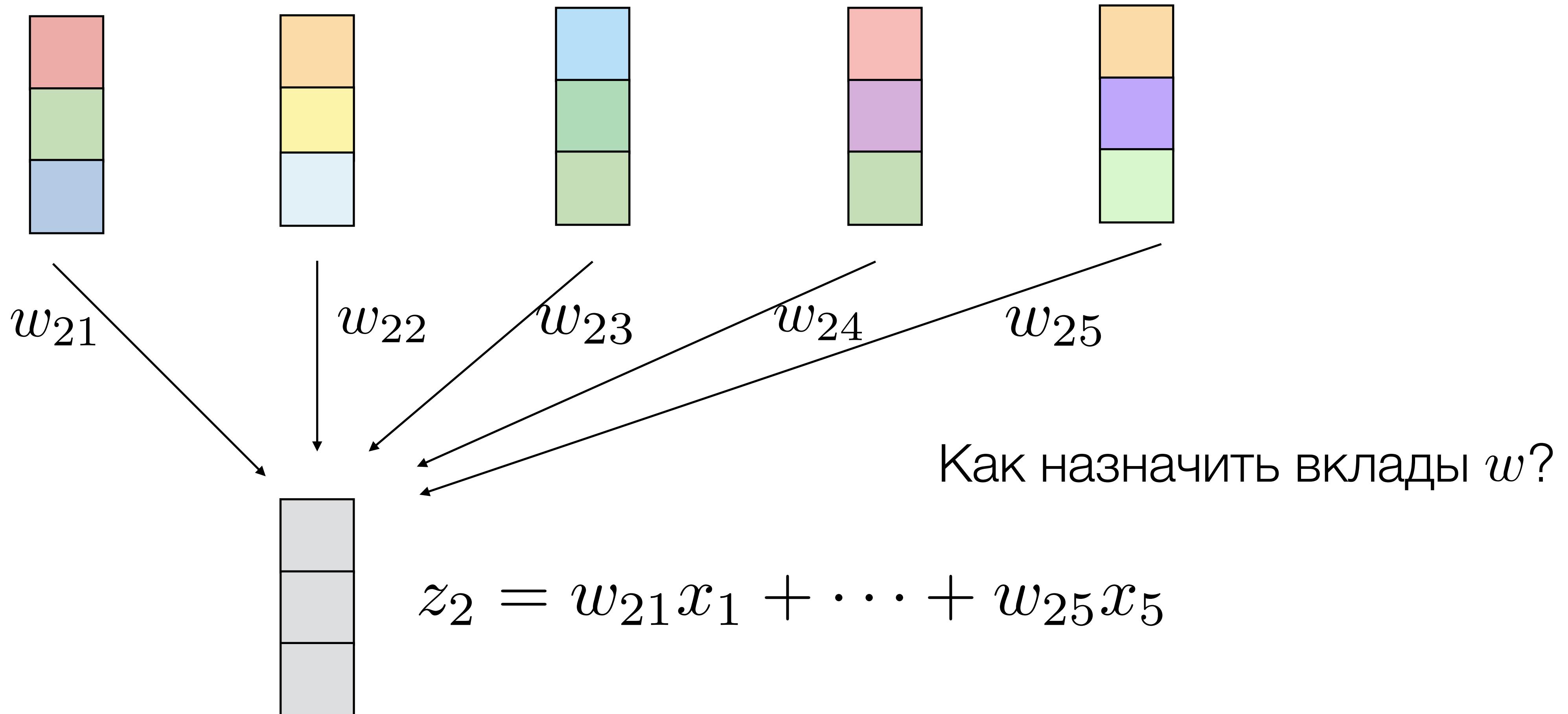
Трансформер: механизм self-attention

В фильме замечательно играют актеры



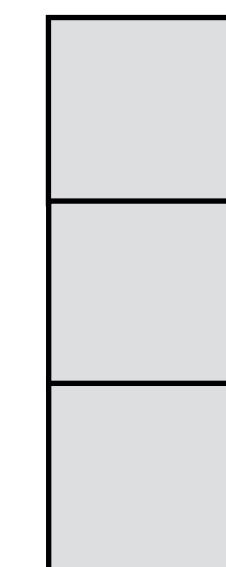
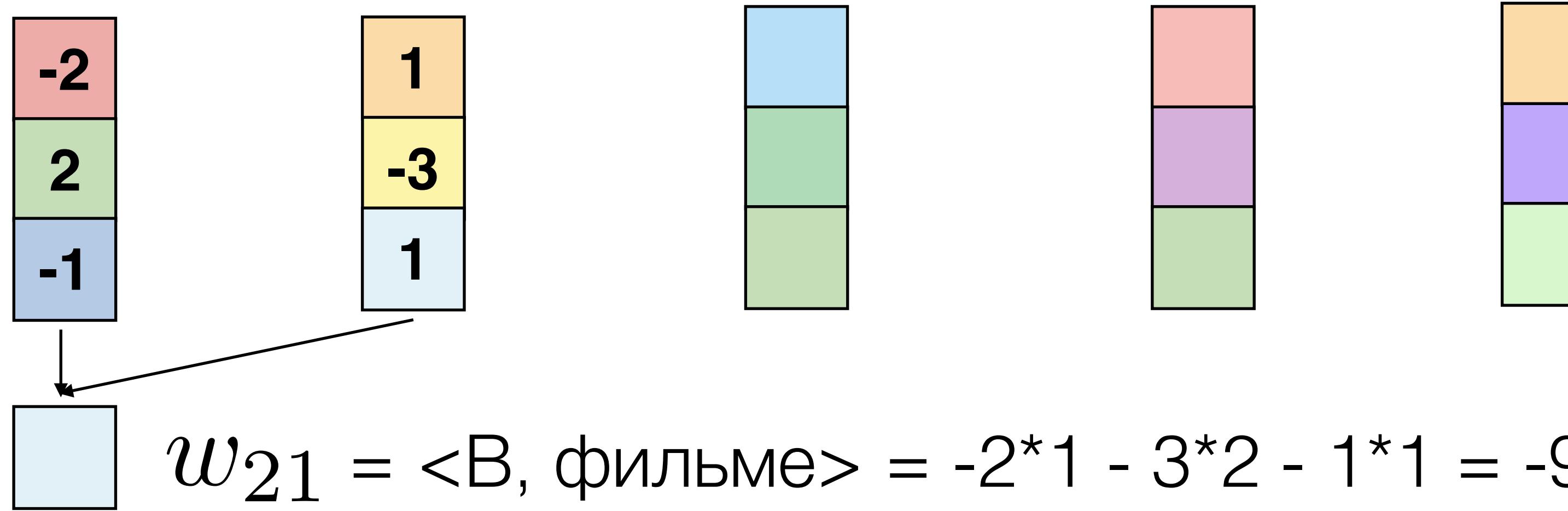
Трансформер: механизм self-attention

В фильме замечательно играют актеры



Трансформер: механизм self-attention

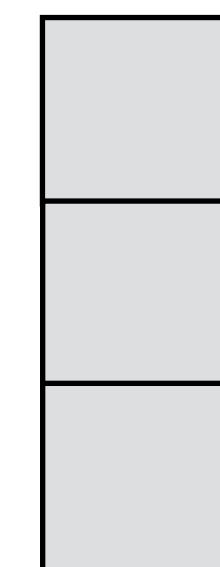
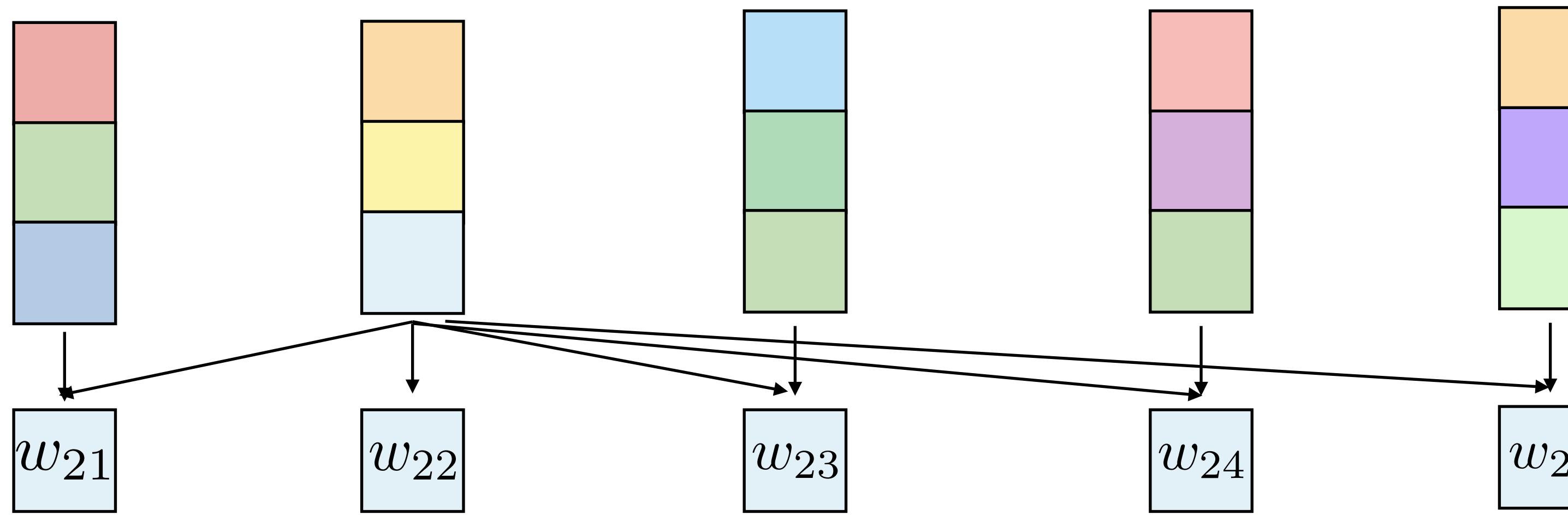
В фильме замечательно играют актеры



$$z_2 = w_{21}x_1 + \dots + w_{25}x_5$$

Трансформер: механизм self-attention

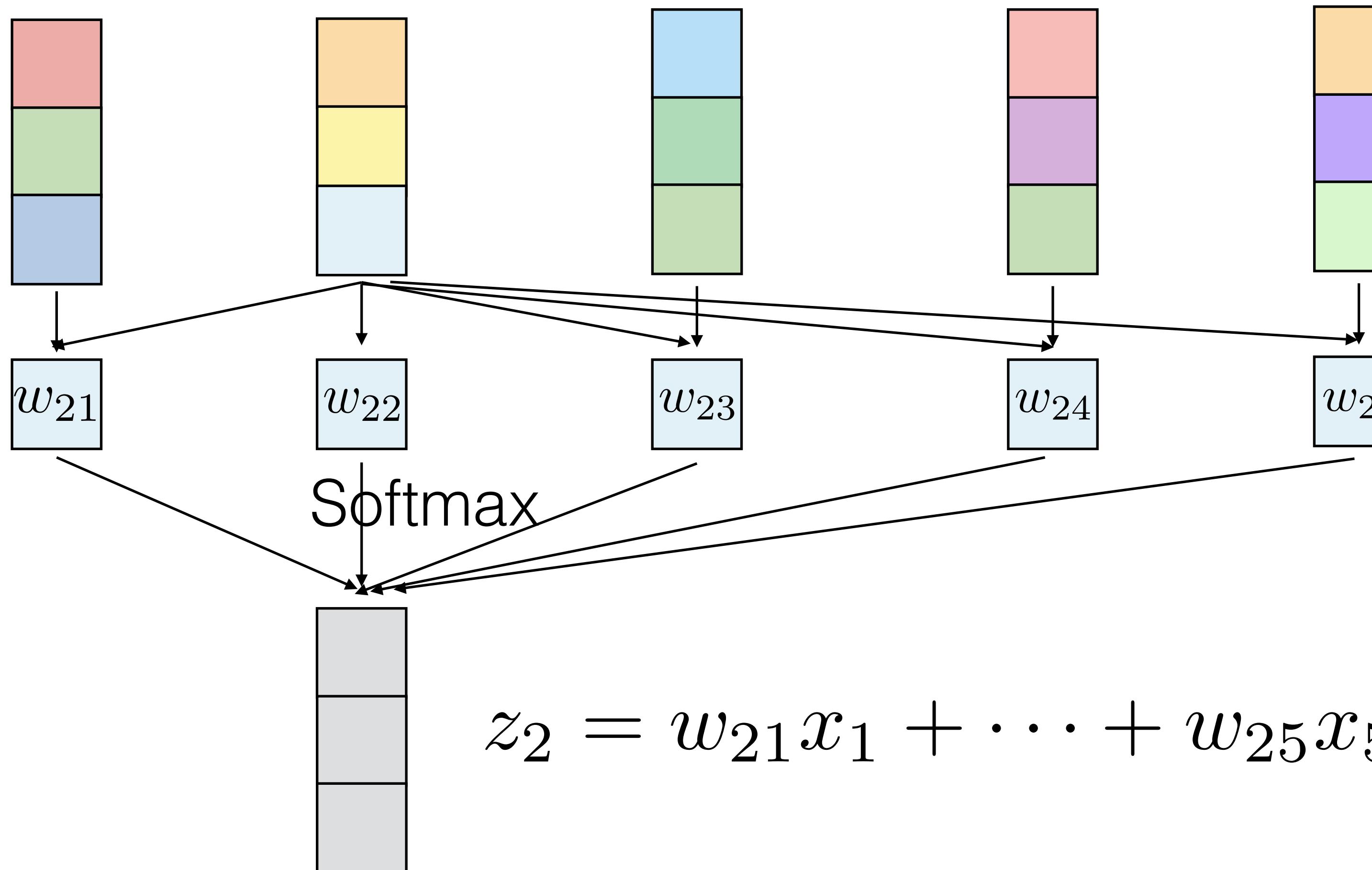
В фильме замечательно играют актеры



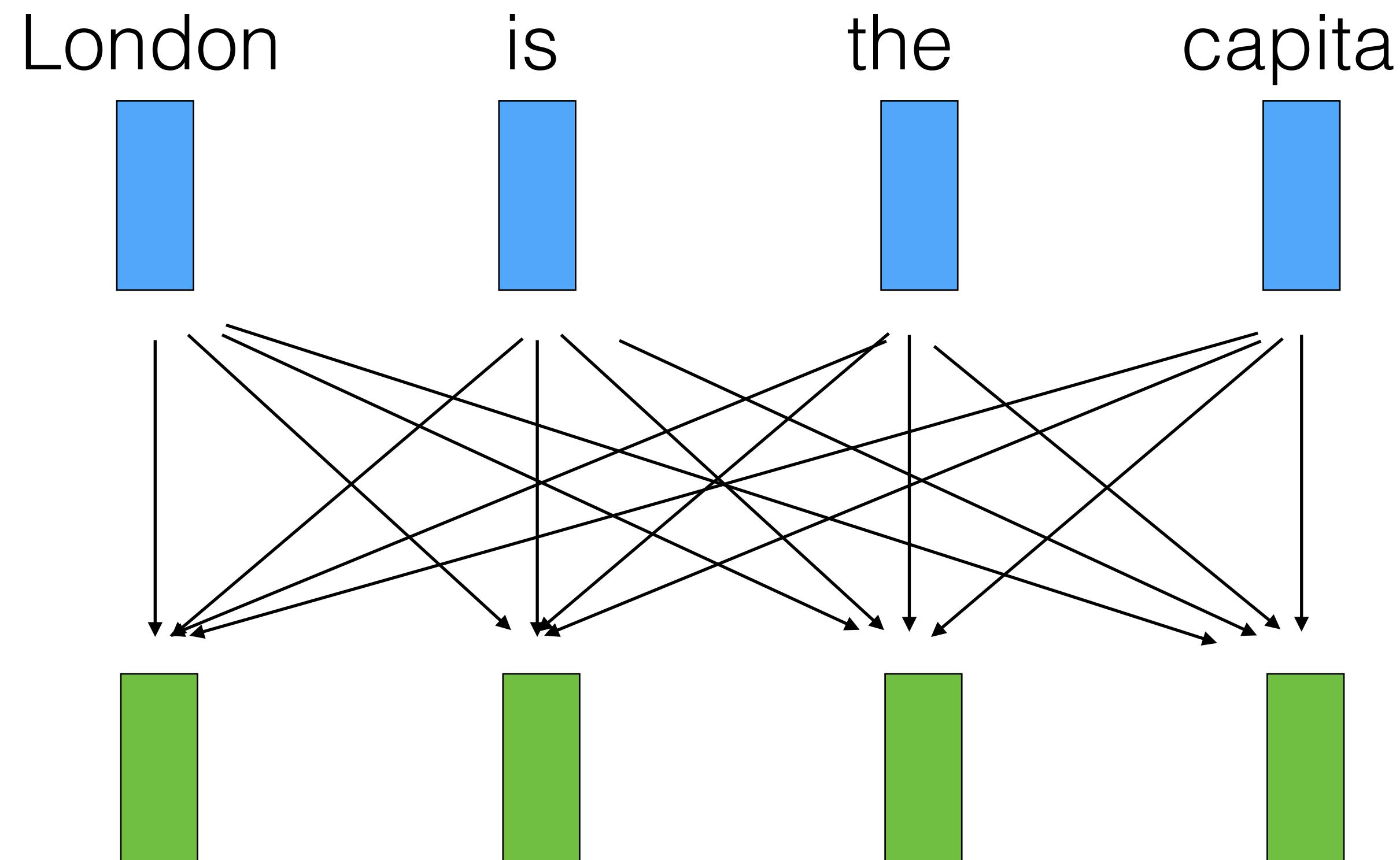
$$z_2 = w_{21}x_1 + \cdots + w_{25}x_5$$

Трансформер: механизм self-attention

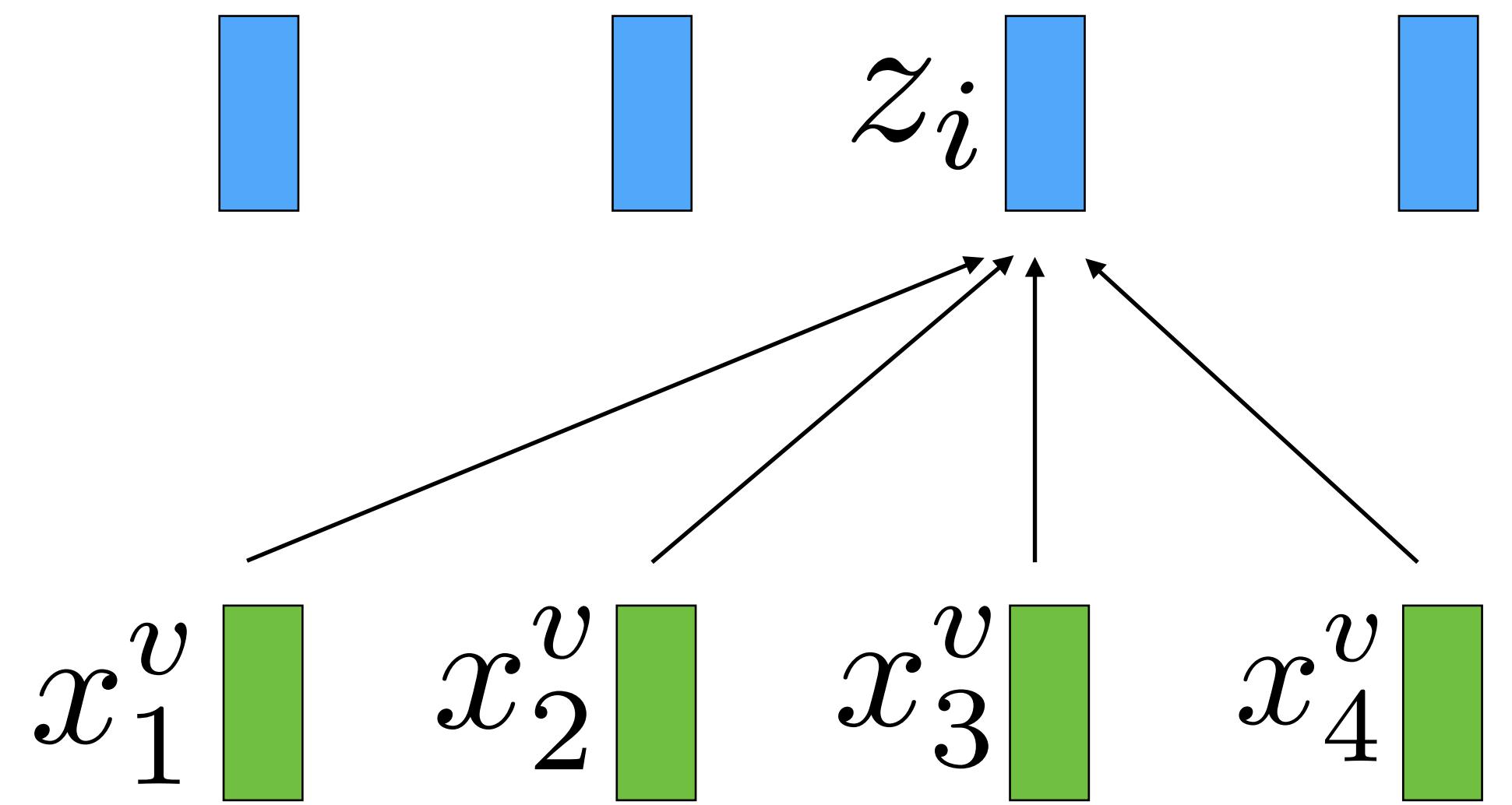
В фильме замечательно играют актеры



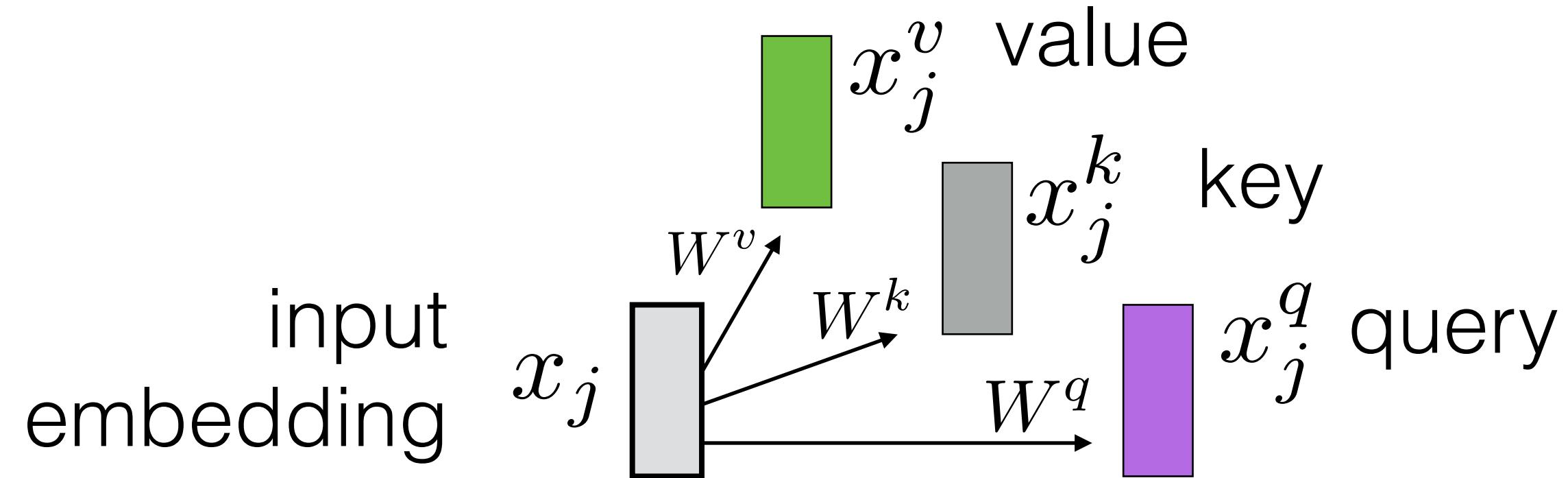
Трансформер: механизм self-attention



Self-attention mechanism



Some random short sequence

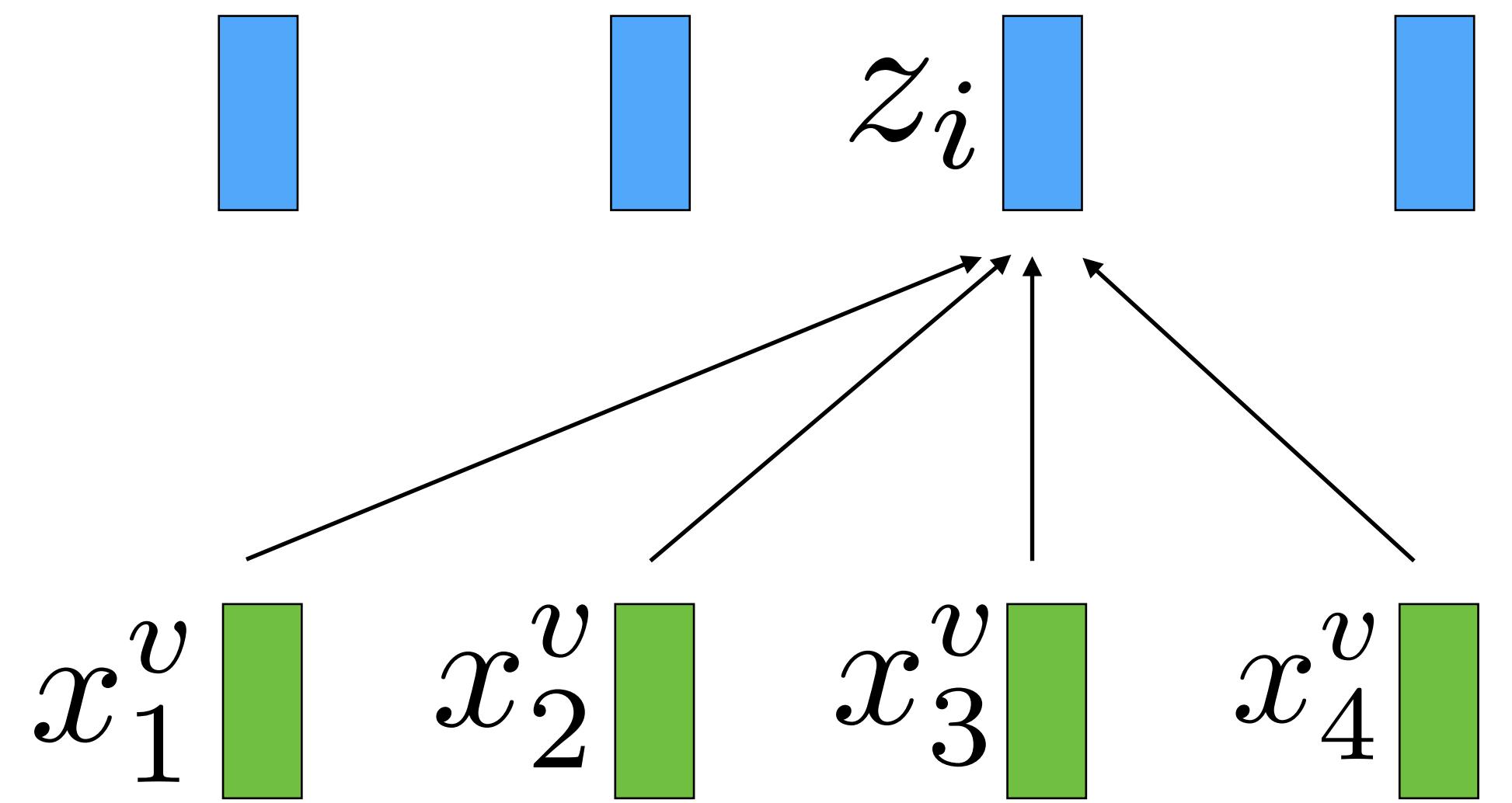


$$z_i = \sum_j \tilde{\alpha}_{ij} x_j^v$$

$$\tilde{\alpha}_{ij} = \frac{\exp(a_{ij})}{\sum_j \exp(a_{ij})}$$

$$a_{ij} = \frac{x_i^q x_j^k {}^T}{\sqrt{d_z}}$$

Self-attention mechanism



Some random short sequence

+multiple heads

+skip-connection,
layer normalization,
fully-connected layer

+ 6 layers

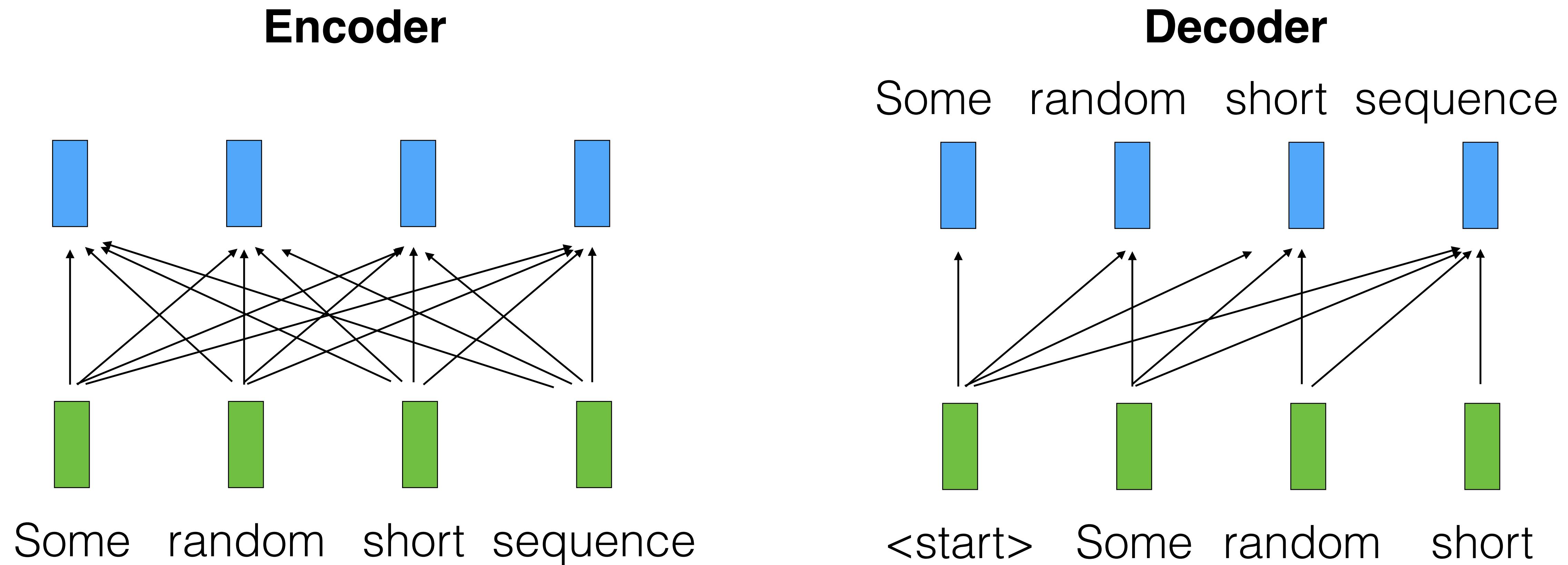
+encoder /
decoder

$$z_i = \sum_j \tilde{\alpha}_{ij} x_j^v$$

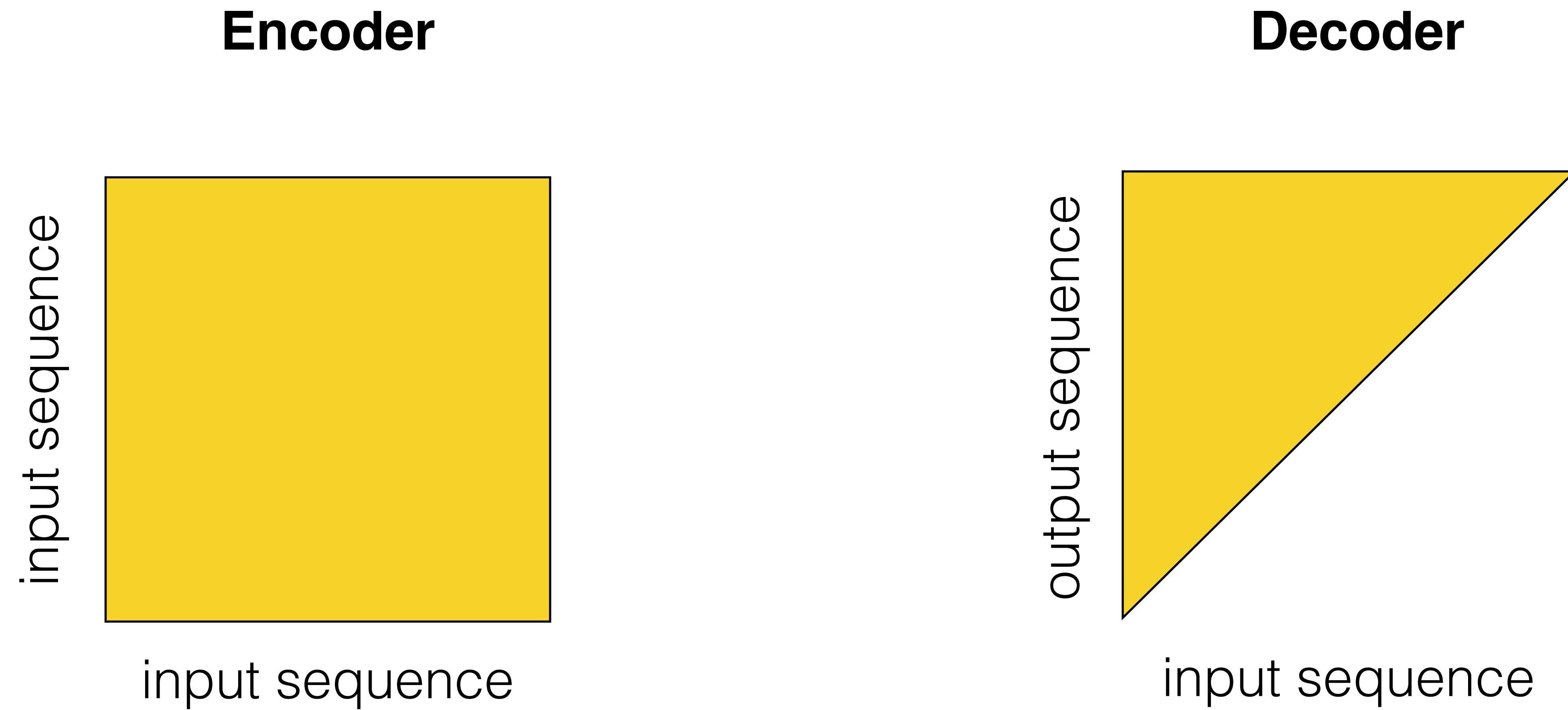
$$\tilde{\alpha}_{ij} = \frac{\exp(a_{ij})}{\sum_j \exp(a_{ij})}$$

$$a_{ij} = \frac{x_i^q x_j^k {}^T}{\sqrt{d_z}}$$

Encoder and Decoder in Transformer



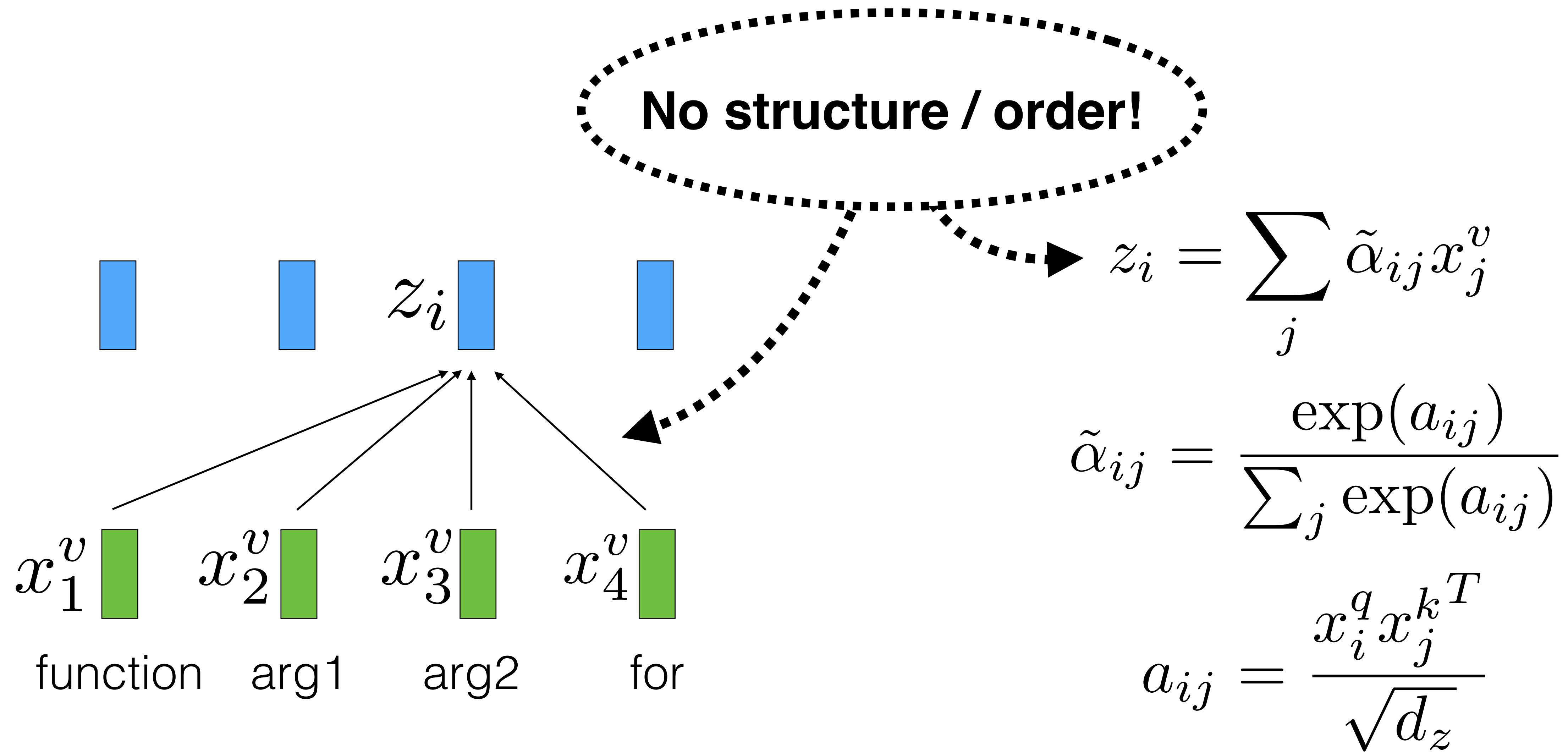
Encoder and Decoder: attention masks



Transformer: GIF

<https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>

Self-attention: “bag” of elements



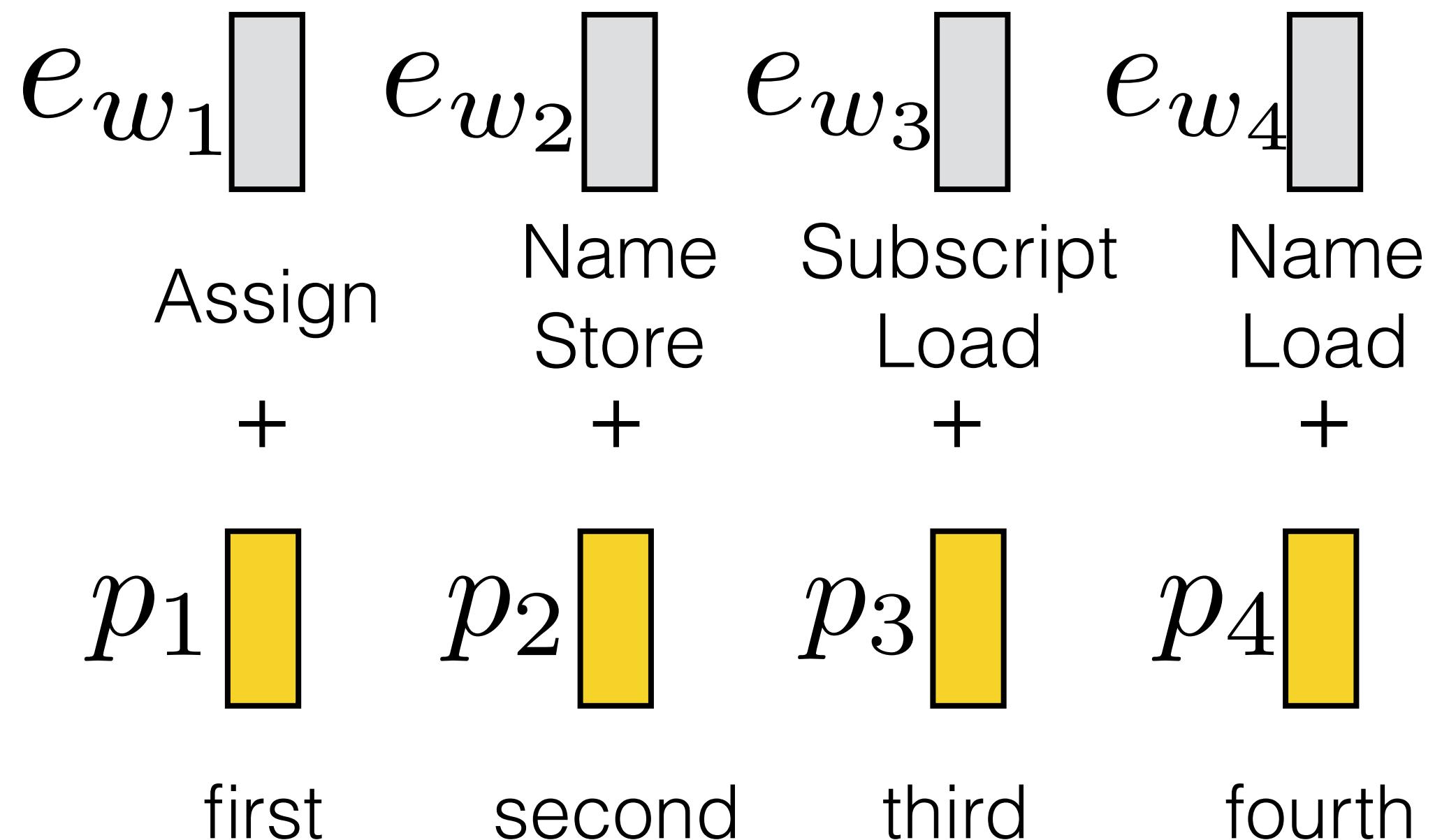
Positional embeddings

Input layer:

$$x_j = e_{w_j} + p_j$$

Positional embeddings:

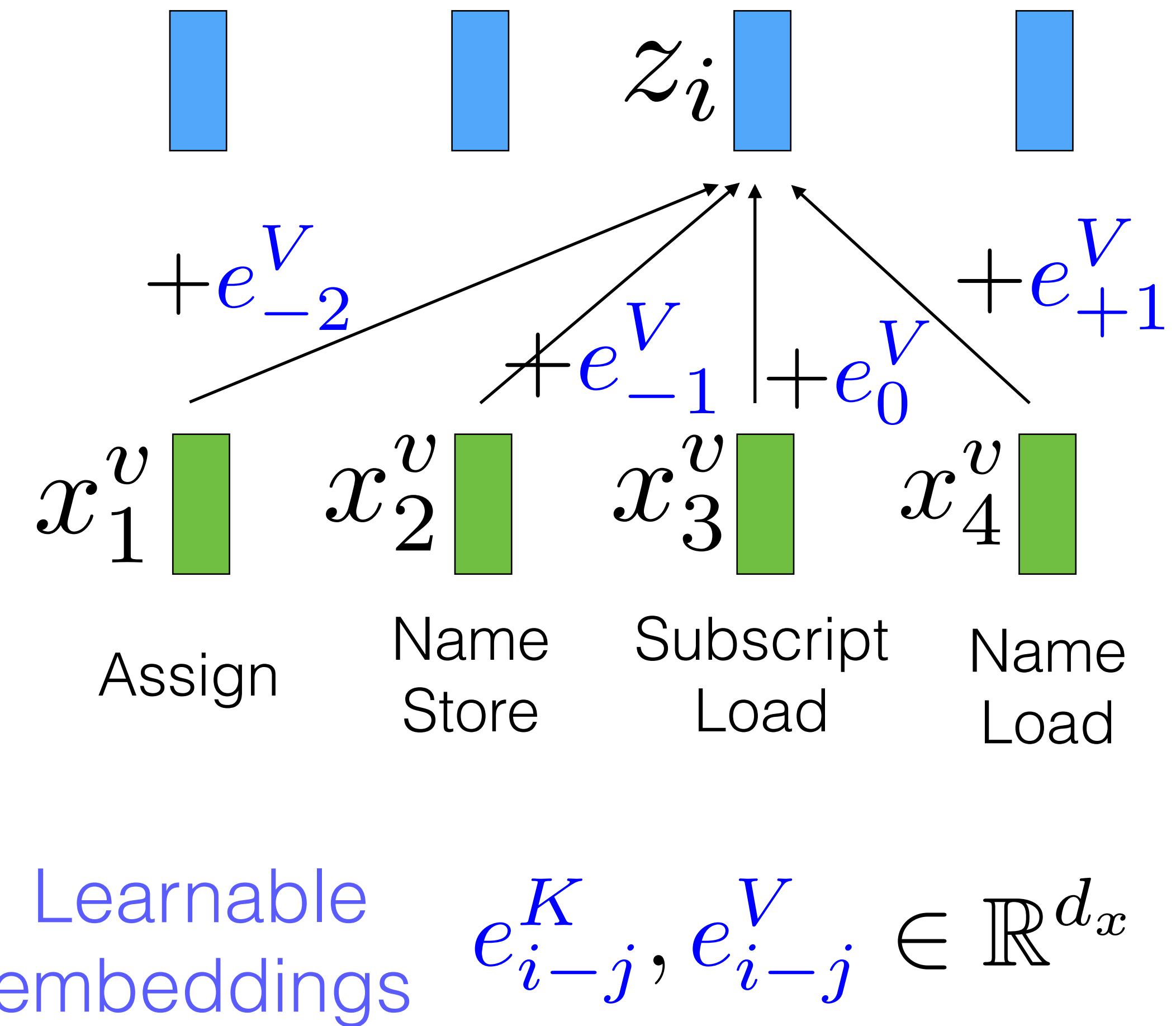
$$p_j = e_j \text{ — learnable embeddings}$$



Parameters: learnable embeddings

Hyperparameters: none

Relative attention

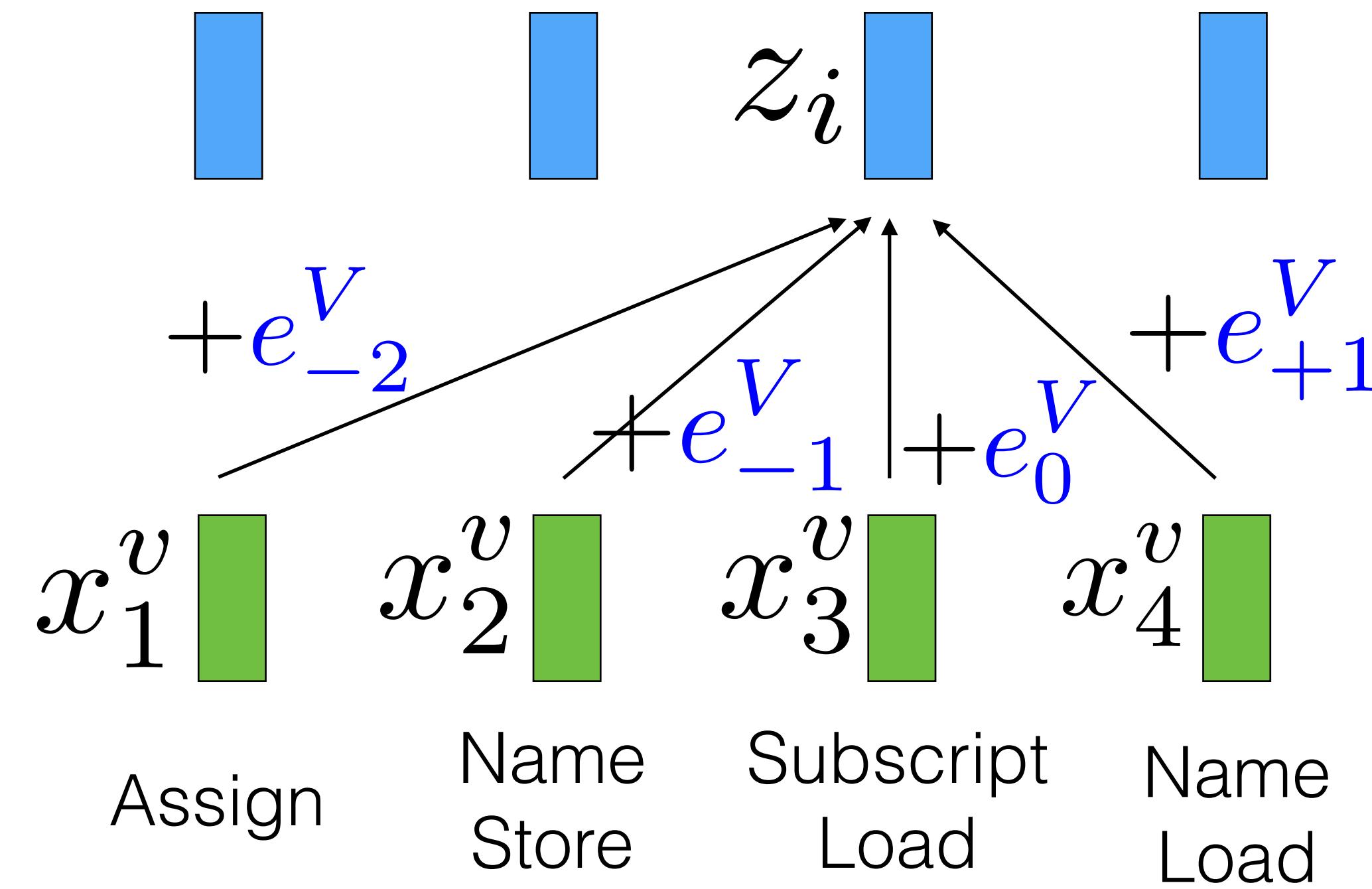


$$z_i = \sum_j \tilde{\alpha}_{ij} (x_j^v + e_{i-j}^V)$$

$$\tilde{\alpha}_{ij} = \frac{\exp(a_{ij})}{\sum_j \exp(a_{ij})}$$

$$a_{ij} = \frac{x_i^q (x_j^k + e_{i-j}^K)^T}{\sqrt{d_z}}$$

Relative attention



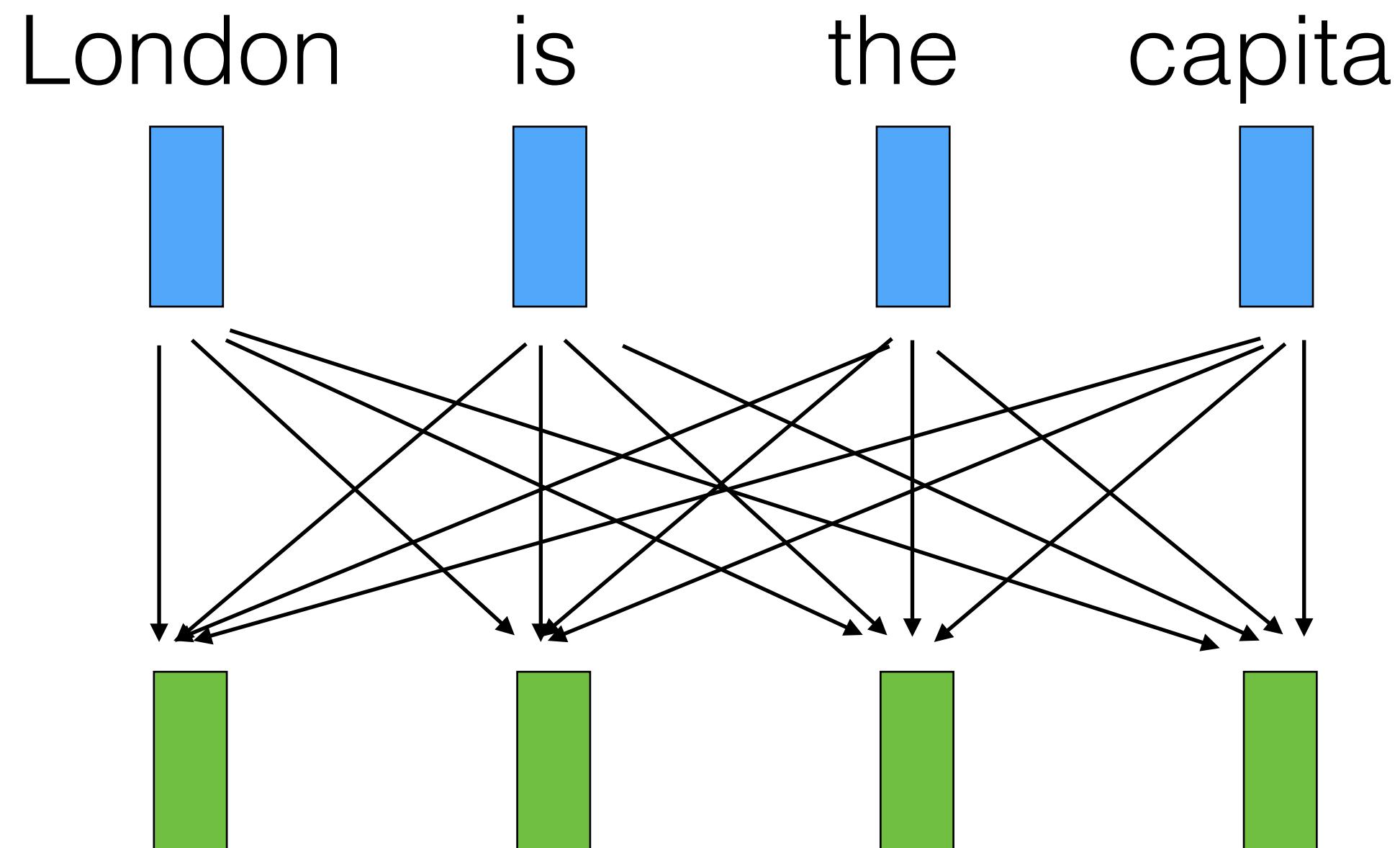
Learnable
embeddings $e_{i-j}^K, e_{i-j}^V \in \mathbb{R}^{d_x}$

Parameters:
embeddings of “relations”

Hyperparameters:
maximum distance between tokens

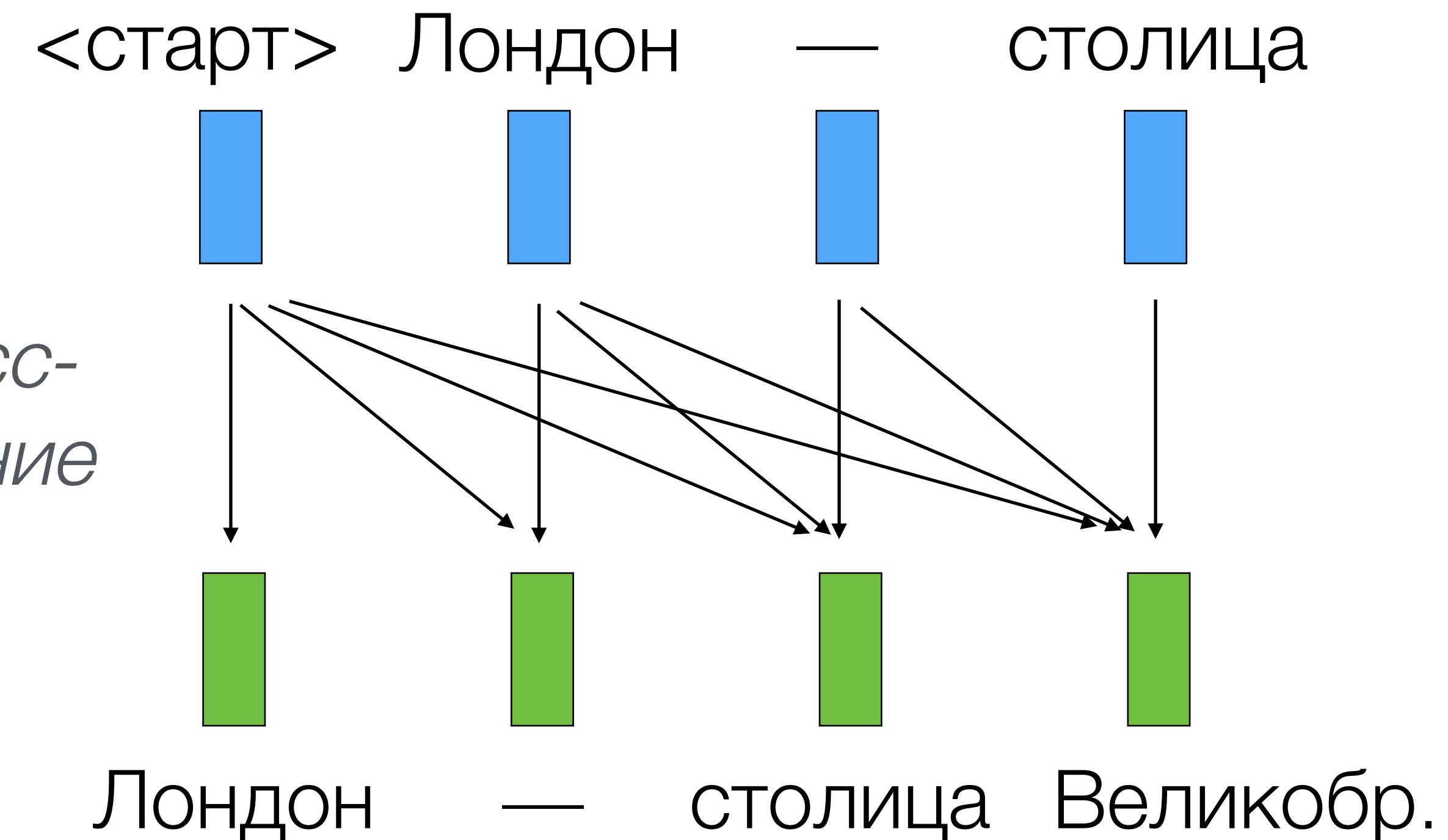
Трансформер в задаче перевода

Кодировщик (encoder)



+ кросс-
внимание

Декодирощик (decoder)

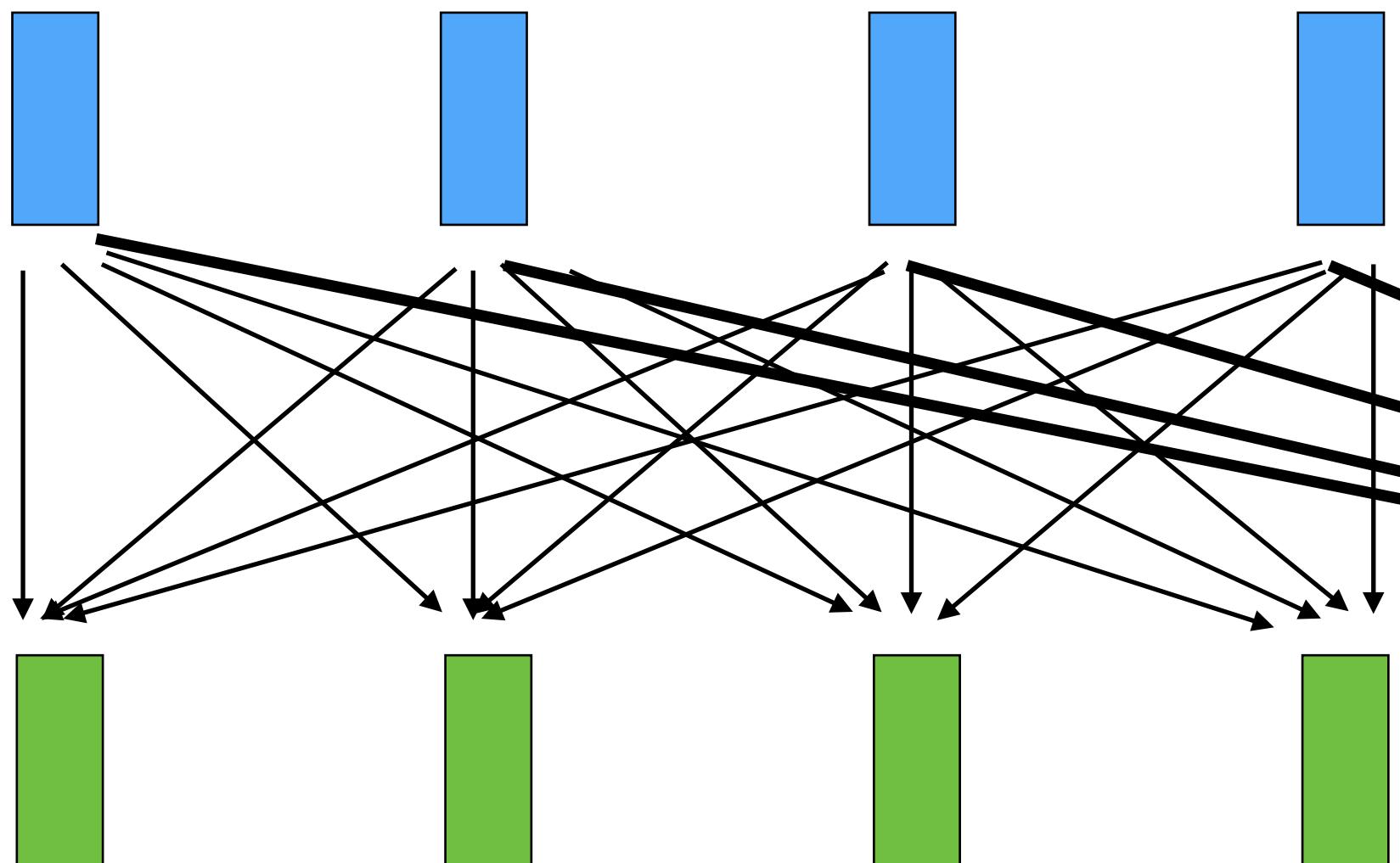


Обычно 6 слоев!

Трансформер в задаче перевода

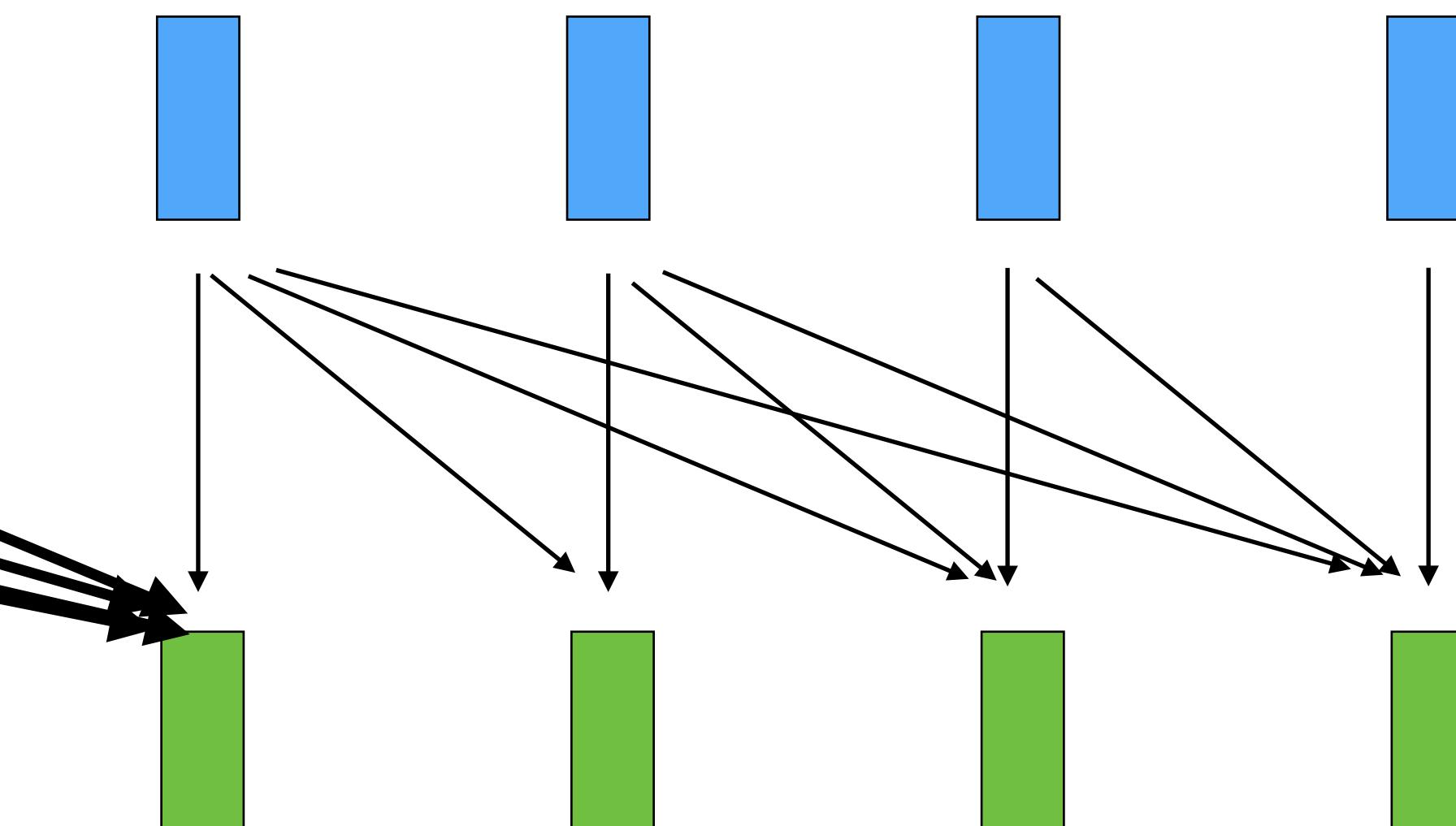
Кодировщик (encoder)

London is the capital



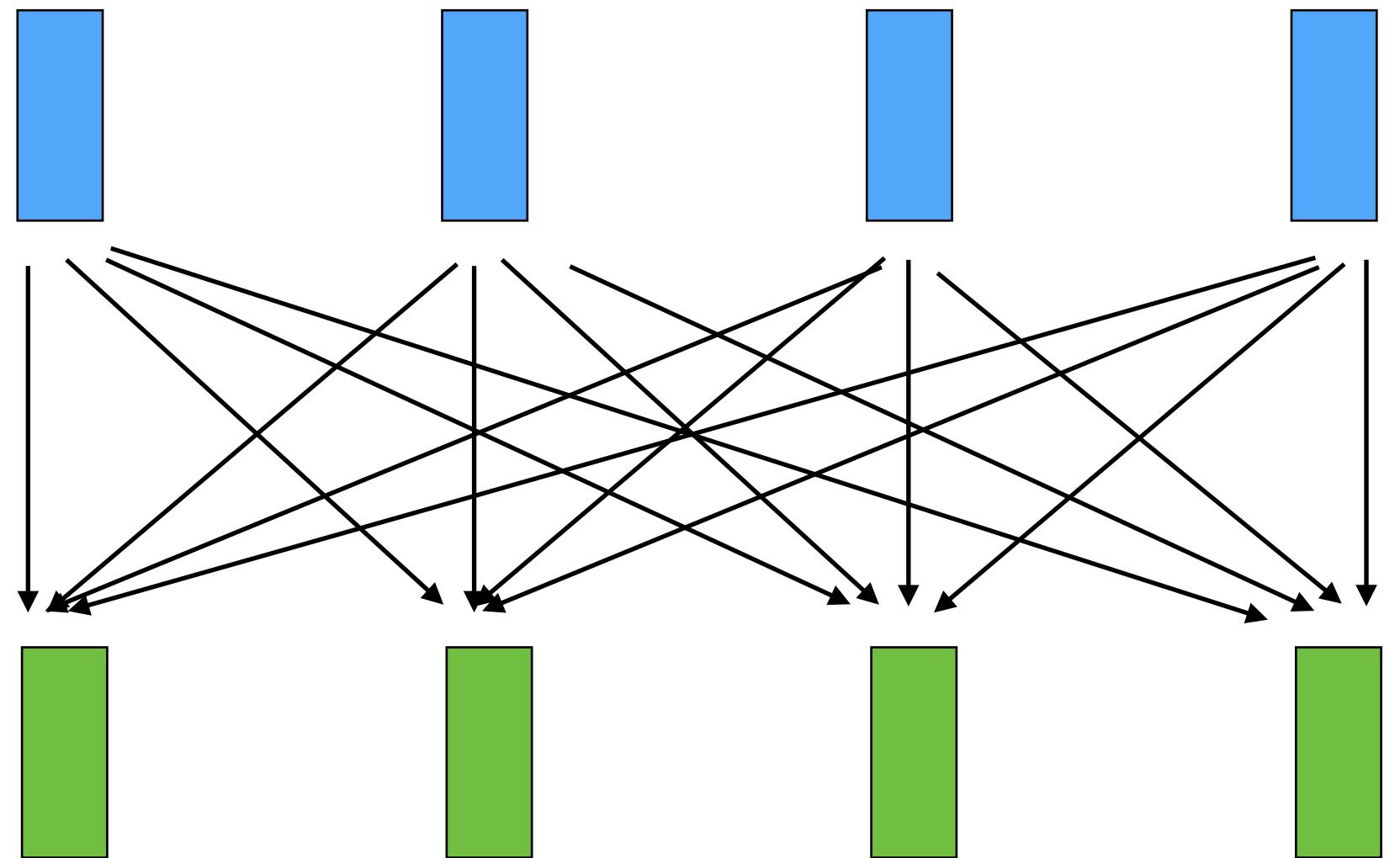
Декодирощик (decoder)

<старт> Лондон — столица



Обычно 6 слоев!

Визуализация внимания

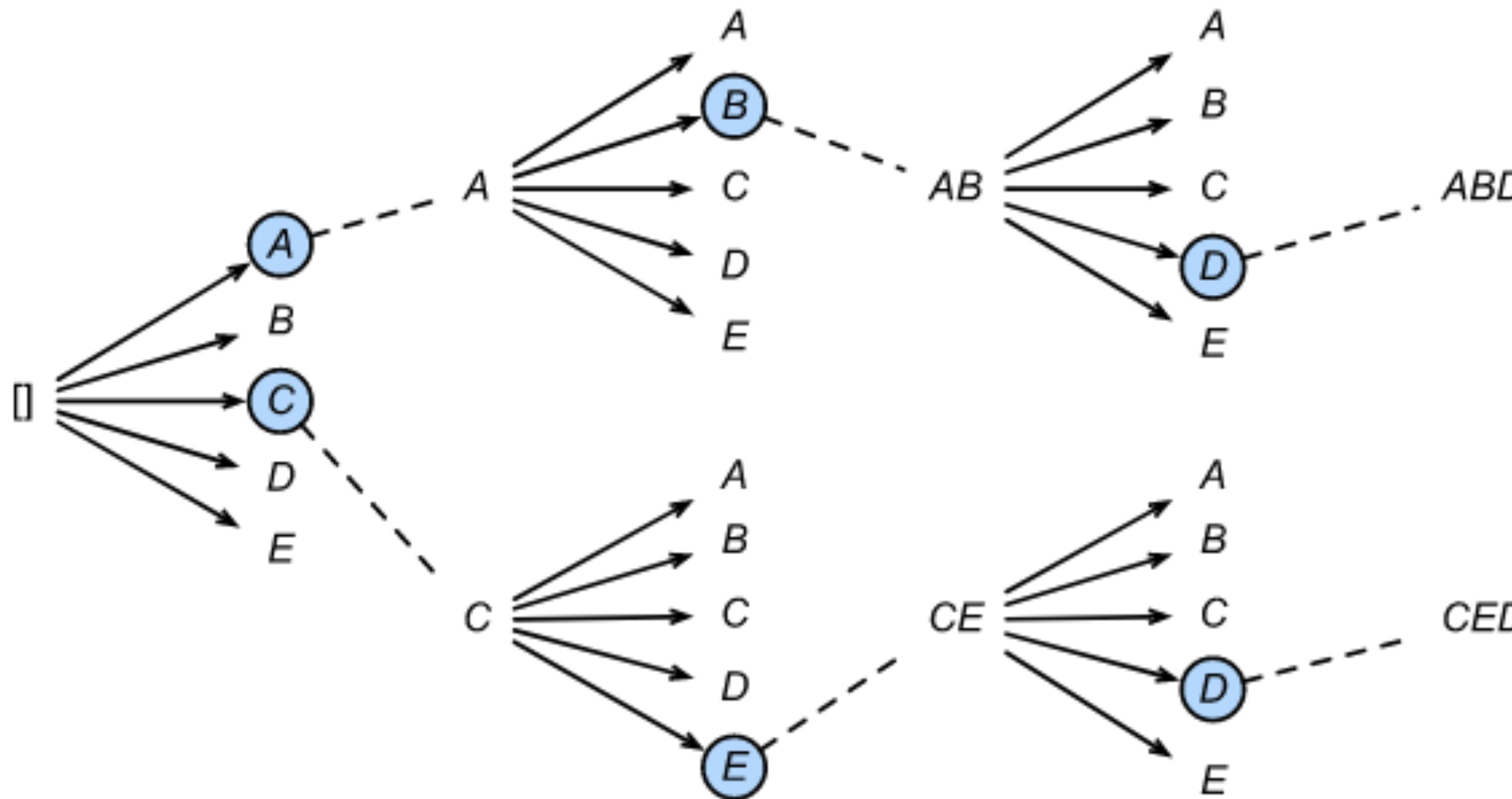


Смешная
обезьяна
съела
наш
банан
ведь
она
была
очень
голодная

Смешная
обезьяна
съела
наш
банан
весь
она
была
очень
голодная

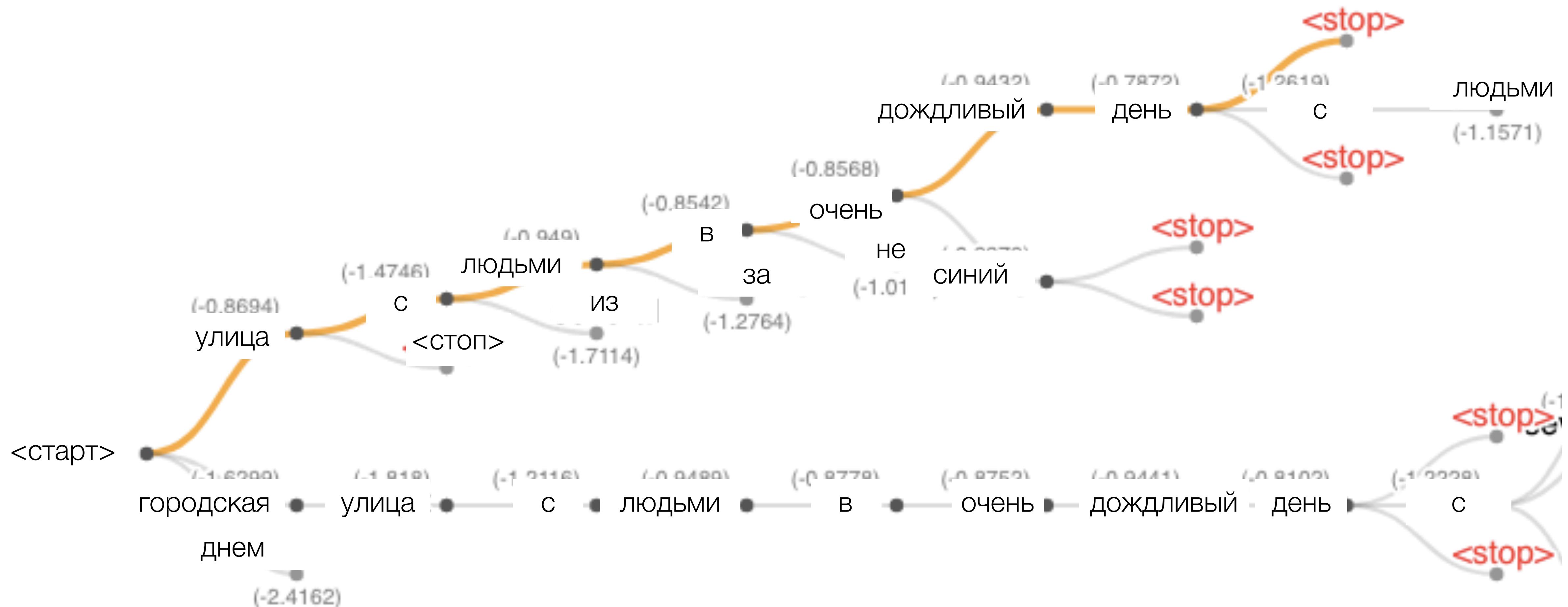
Еще немного про генерацию текстов

Лучевой поиск (beam search) для словаря $\{A, B, C, D, E\}$:



Без beam search: на каждом шаге выбирать один вариант

Еще немного про генерацию текстов



План

- Задачи анализа текстов
- Векторные представления слов
- Архитектуры нейронных сетей
 - Рекуррентные нейронные сети
 - Трансформер
- Измерение качества в задачах генерации текста
- Перенос обучения в анализе текстов

Задачи анализа текстов

Дискриминативные

Текст → ?

Классификация текстов

А. С. Пушкин → **литература**
писал стихи

Предсказание заработной платы по тексту объявления

Стажер в отдел
продаж, опыт не нужен → **25**

Генеративные

? → Текст

Автодополнение

Мы поехали в Италию... →
в La Scuole. Там все то же самое.

Генерация подписи к изображению



→ **группа людей
на пляже**

Измерение качества при генерации текстов



Девушка с чашкой кофе

Девушка держит чашку кофе

Девушка держит в руке чашку кофе

Девушка сидит за столом с чашкой кофе

Измерение качества при генерации текстов

We have a ski trip tomorrow

EN → . RU

У нас завтра лыжная прогулка

У нас на завтра запланирована поездка на лыжах

Завтра мы поедем кататься на лыжах

Завтра — наш день катания на лыжах!

Измерение качества при генерации текстов

В выборке: Девушка с чашкой кофе

биграммы:

девушка с
с чашкой
чашкой кофе

Предсказание: Девушка сидит за
столом с чашкой кофе

биграммы:

девушка сидит
сидит за
за столом
столом с
с чашкой
чашкой кофе

BLEU: (немного модифицированная) точность (precision) по n-граммам

✓ Обычно 1—4-граммы

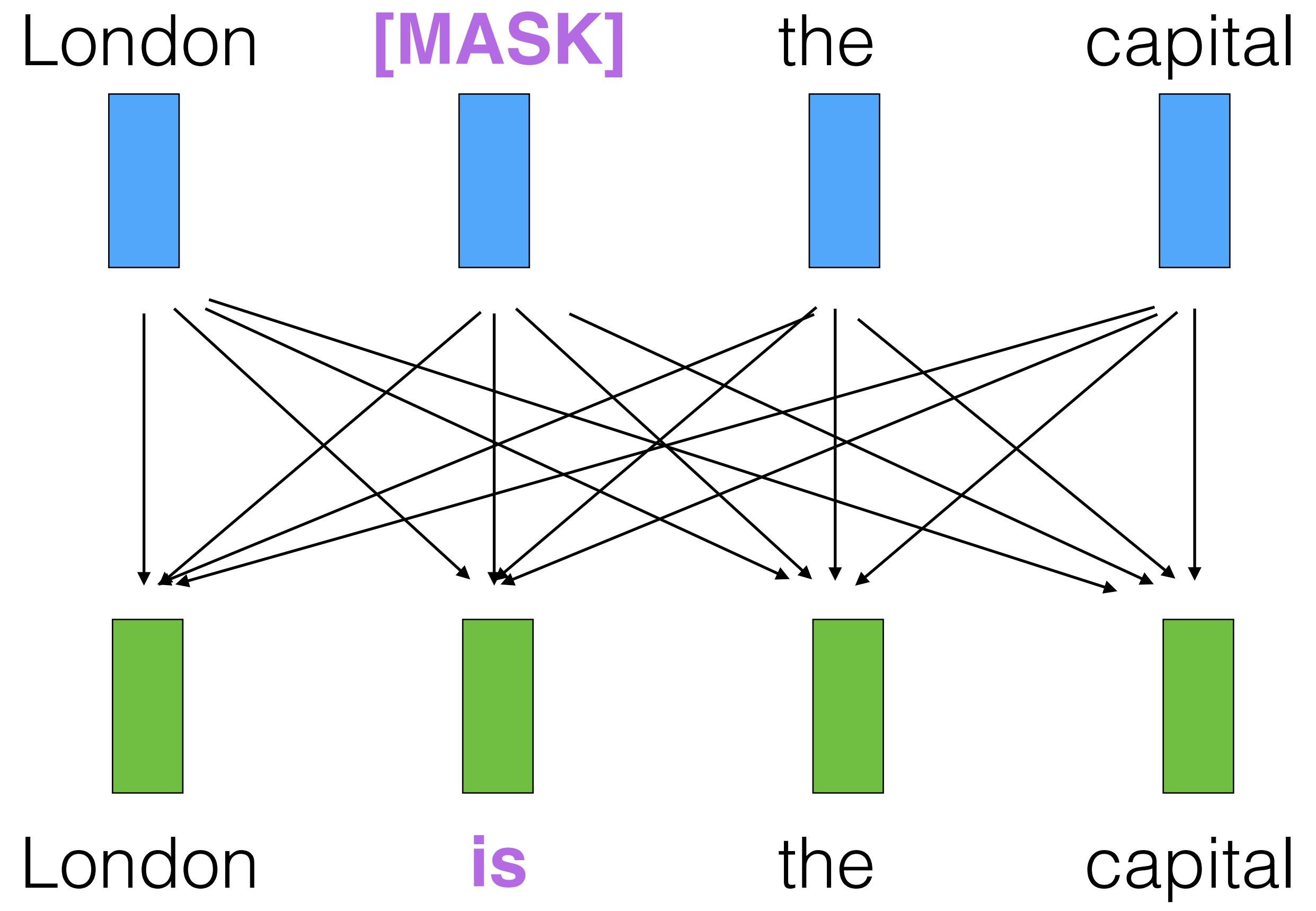
✓ Штраф за длинные / короткие предсказания

План

- Задачи анализа текстов
- Векторные представления слов
- Архитектуры нейронных сетей
 - Рекуррентные нейронные сети
 - Трансформер
- Измерение качества в задачах генерации текста
- Перенос обучения в анализе текстов

BERT

- Обучаем Трансформер на большом корпусе данных (Википедия)
- Теперь Трансформер умеет “понимать” язык
- При решении “маленькой” задачи (предсказание зарплаты по объявлению) *дообучаем* Трансформер на маленьком наборе данных

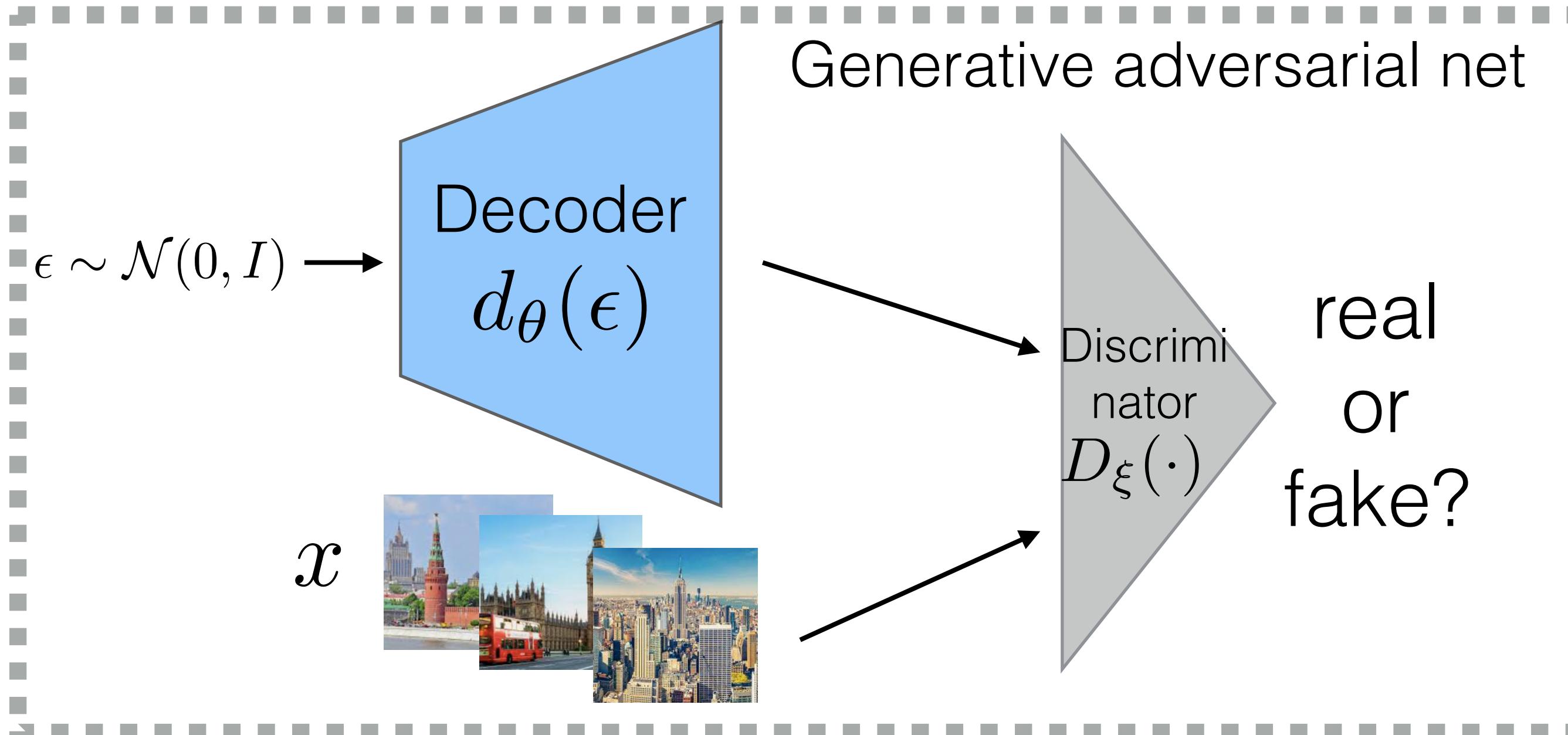


Можно один BERT на несколько языков!

ИТОГИ

- Задачи анализа текстов: дискrimинативные (понимать текст, решаются хорошо)
генеративные (создавать текст, решаются чуть хуже)
- Векторные представления слов: используются во всех моделях
- Архитектуры нейронных сетей
 - Рекуррентные нейронные сети
 - Трансформер — сейчас все используют его!
- Измерение качества в задачах генерации текста:
BLEU хорошо коррелирует с экспертными оценками
- Перенос обучения в анализе текстов: BERT - используется повсеместно!

Generative adversarial networks (GANs)



$$\begin{aligned} & \frac{1}{N} \sum_{i=1}^N \log D_\xi(x_i) + \\ & + \mathbb{E}_{\epsilon \sim \mathcal{N}(0, I)} \log(1 - D_\xi(d_\theta(\epsilon))) \rightarrow \max_{\xi} \min_{\theta} \end{aligned}$$

GANs: examples



<https://petapixel.com/2017/11/07/ai-creates-photo-realistic-faces-people-dont-exist/>

GANs: examples

