machine translation

All Images News Videos Books More Settings Tools

About 477,000,000 results (0.74 seconds)

English – detected Chinese

machine

məˈSHēn

机

Jī

Машинный перевод: план доклада

1. Формулировка задачи
2. Базовые методы. Перевод по словарю, по правилам и по примерам
3. Статистический машинный перевод
4. Вспомним RNN
5. Sequence-to-Sequence. Beam search decoder.
6. Sequence-to-Sequence with Attention
7. BLEU и оценки качества перевода

Задача машинного перевода

Последовательности слов $x \in X^*, y \in Y^*$

$$f : x \to y$$

$$f(x) = argmax_y p(y|x)$$

- Я съел торт → I have eaten a cake

Перевод по словарю

Переводим каждое слово
по фиксированному словарю

Плюсы:

– Простая модель

Минусы:

– Слово имеет только один перевод
– Не учитывает особенности языка

Rule-based machine translation (RBMT)

Храним словарь и список правил

Переводим каждое слово
по фиксированному словарю,
после чего применяем
правила

Плюсы:

– Работает лучше

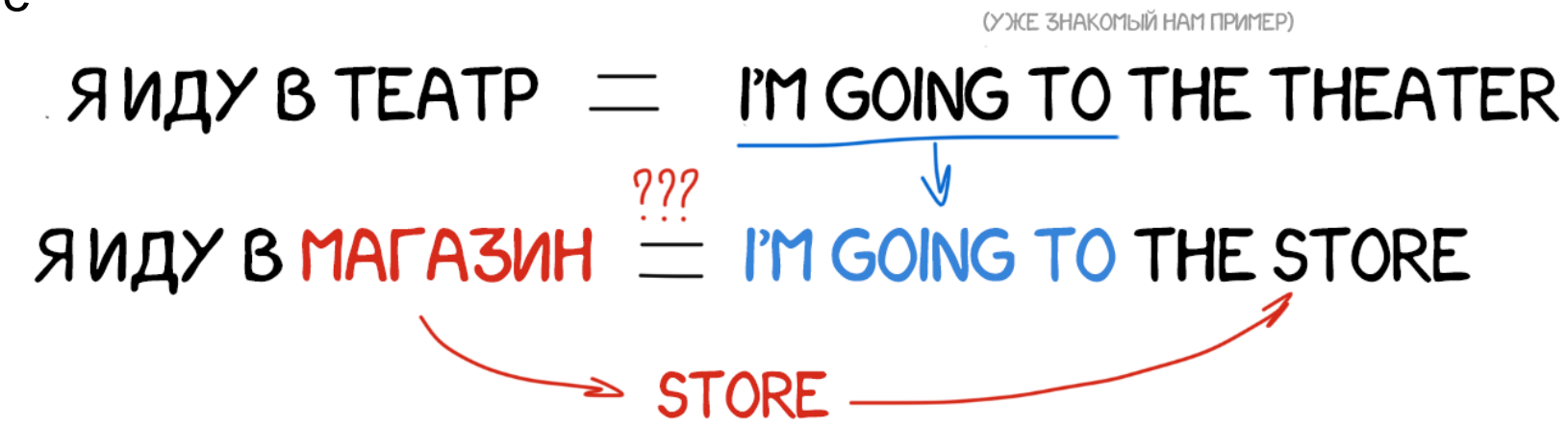
Минусы:

– Много ручного труда
– Всё правилами не покрыть

Example-based machine translation (EBMT)

Храним словарь и много примеров

Ищем похожее предложение
среди примеров
и подставляем туда
новые слова



Плюсы:

– Работает ещё лучше

Минусы:

– Нужно хранить очень много примеров

Statistical Machine Translation (SMT)

Решаем задачу:

$$argmax_y p(y|x)$$

По теореме Байеса,

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

$$argmax_y p(y|x) = argmax_y p(x|y)p(y)$$

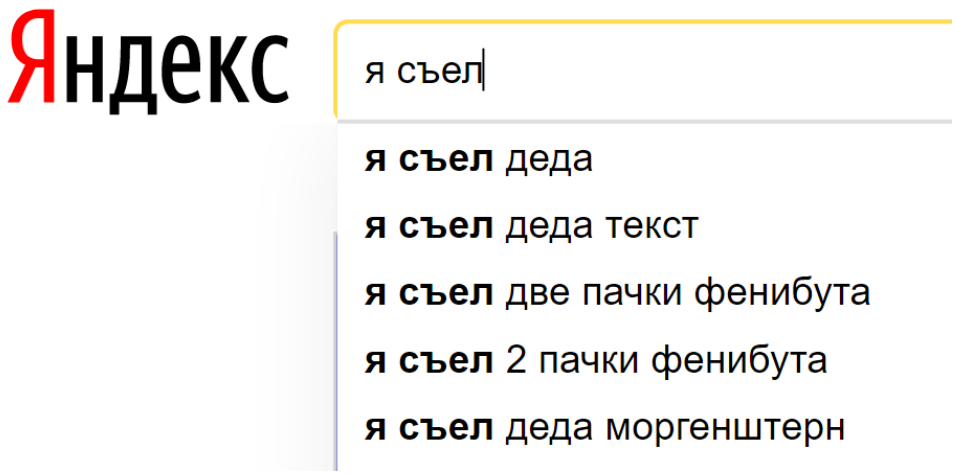
Translation Model
Насколько фразы соответствуют друг другу

Language Model
Несёт в себе структуру языка

SMT: Language Model

Надо искать $p(y)$

$$
\begin{aligned}
&= p(y_0)p(y_1 \ldots y_n|y_0) \\
&= p(y_0)p(y_1|y_0)p(y_2 \ldots y_n|y_0, y_1) \\
&= p(y_0)p(y_1|y_0)p(y_2|y_0, y_1) \ldots p(y_n|y_0, \ldots, y_{n-1})
\end{aligned}
$$



SMT: Language Model

Как искать $p(y)$

$$= p(y_0)p(y_1|y_0)p(y_2|y_1)p(y_3|y_2)\dots p(y_n|y_{n-1})$$

$$p(y_k|y_{k-1}) = \frac{\#\{y_{k-1}, y_k\}}{\#\{y_{k-1}\}}$$

SMT: Language Model

Как искать $p(y)$

$$= p(y_0)p(y_1|y_0)p(y_2|y_0, y_1)p(y_3|y_1, y_2) \dots p(y_n|y_{n-2}, y_{n-1})$$

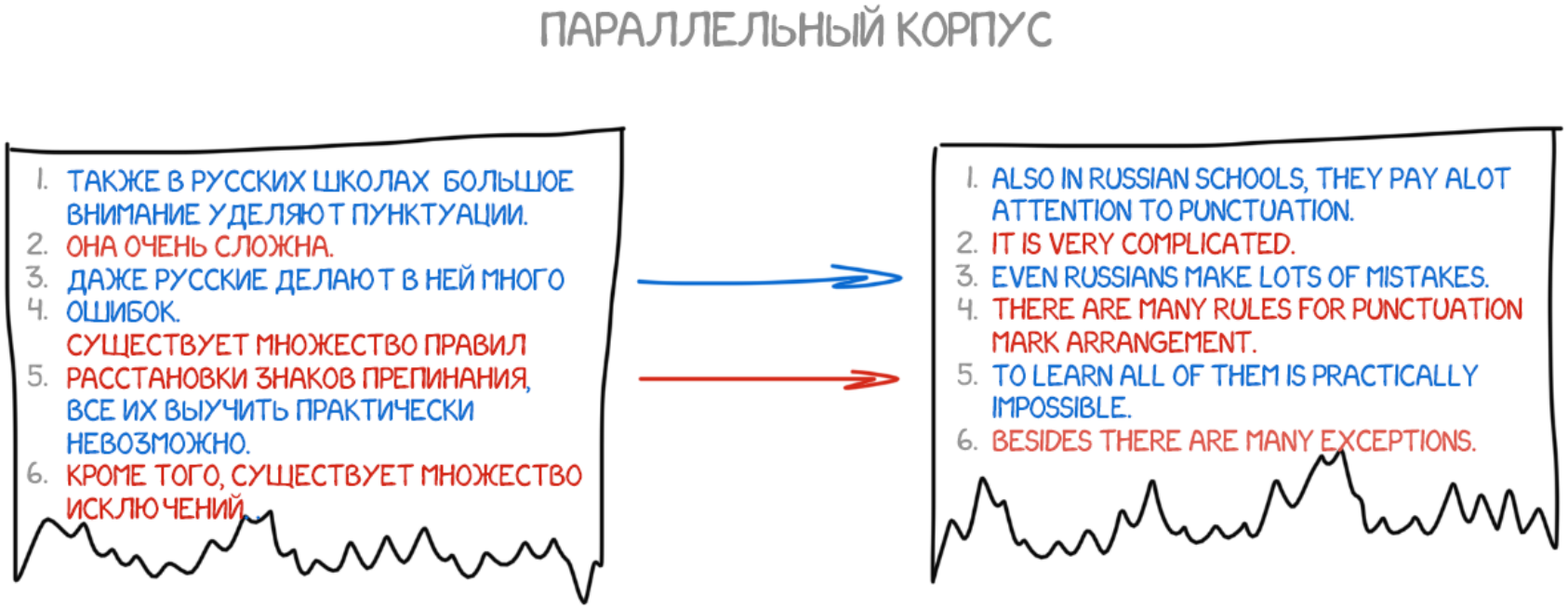
$$p(y_k|y_{k-1}, y_{k-2}) = \frac{\#\{y_{k-2}, y_{k-1}, y_k\}}{\#\{y_{k-2}, y_{k-1}\}}$$

Так же и с n-граммами – *phrase based*.

Language model учится только на корпусе данного языка.

SMT: Translation Model

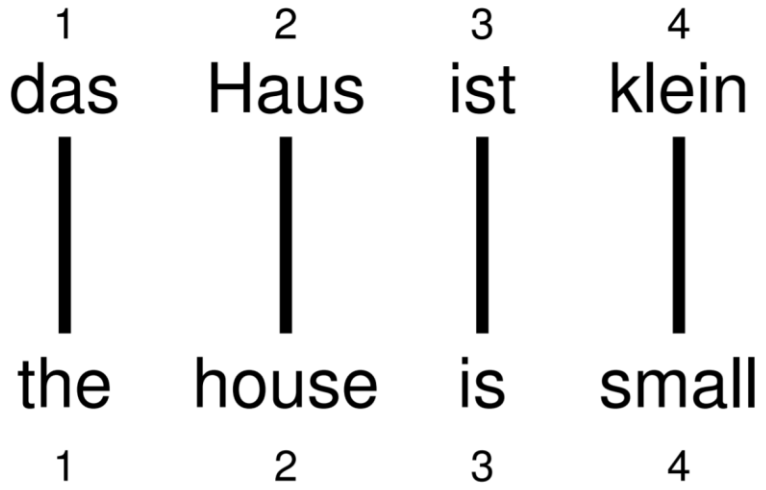
Надо искать $p(x|y)$



SMT: Translation Model

Как искать $p(x|y)$

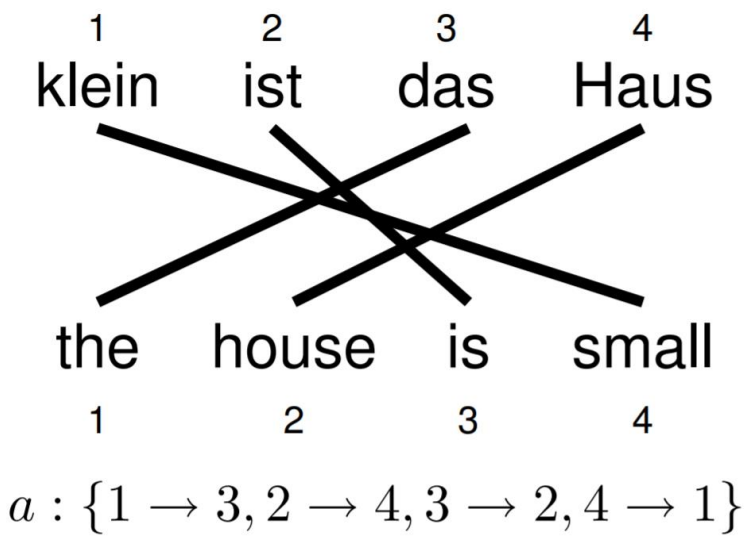
- Разбиваем предложения параллельного корпуса на слова и сопоставляем
- $p(x|y)$ = как часто слово x сопоставляется с y



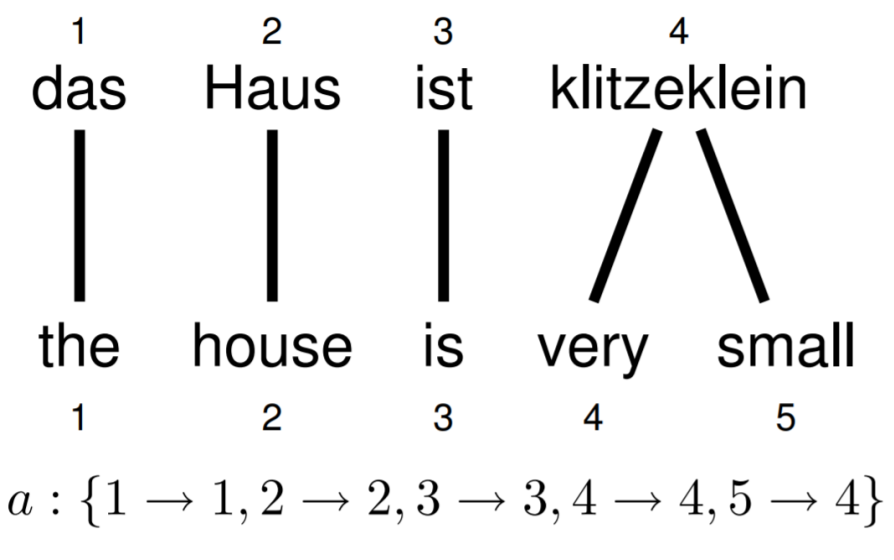
SMT: Translation Model

Как искать $p(x|y)$

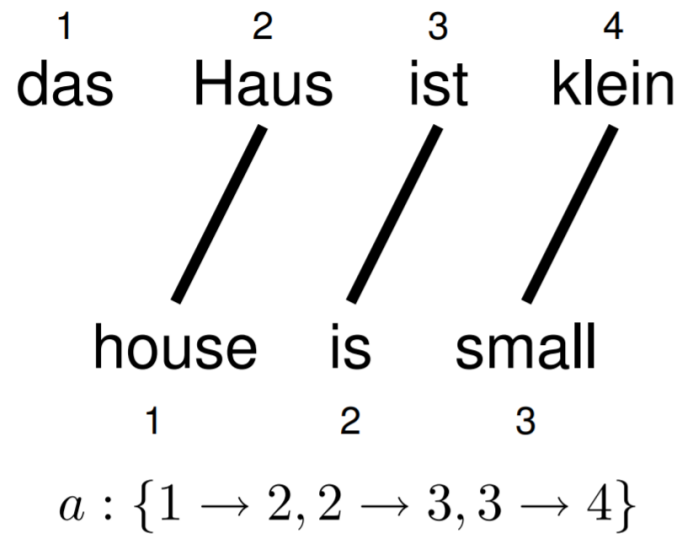
- Добавляем параметр α – *alignment* (выравнивание) $p(x, \alpha|y)$



$a : \{1 \rightarrow 3, 2 \rightarrow 4, 3 \rightarrow 2, 4 \rightarrow 1\}$



$a : \{1 \rightarrow 1, 2 \rightarrow 2, 3 \rightarrow 3, 4 \rightarrow 4, 5 \rightarrow 4\}$



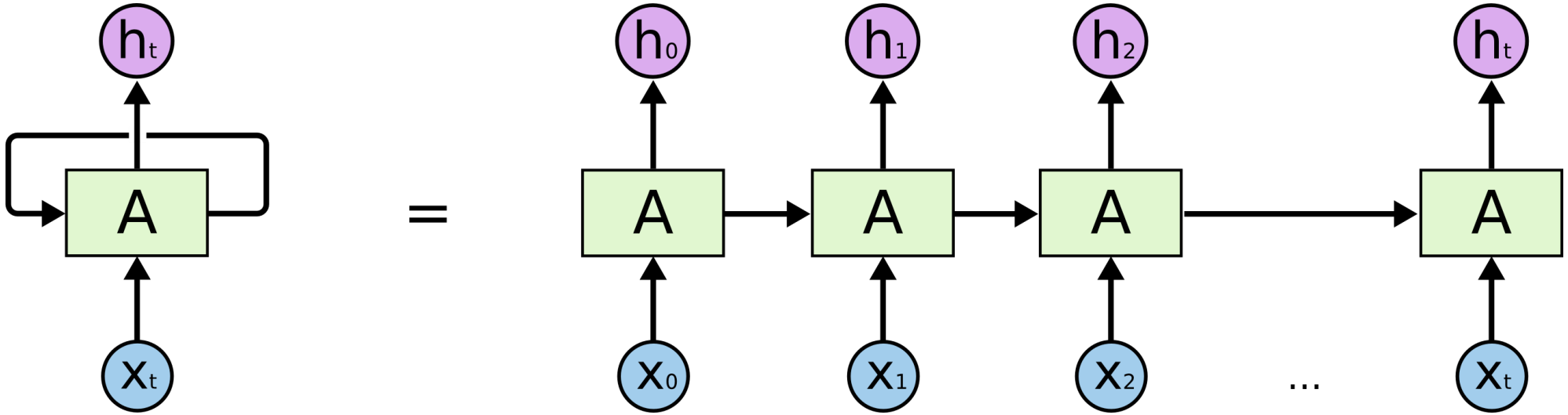
$a : \{1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 4\}$

SMT: overview

- Статистический машинный перевод широко применялся до 2015 года
- Меньше ручного труда, более универсальный
- **Огромное** количество деталей и вариаций
- Использует и хранит **огромные** объёмы данных

Recurrent Neural Network

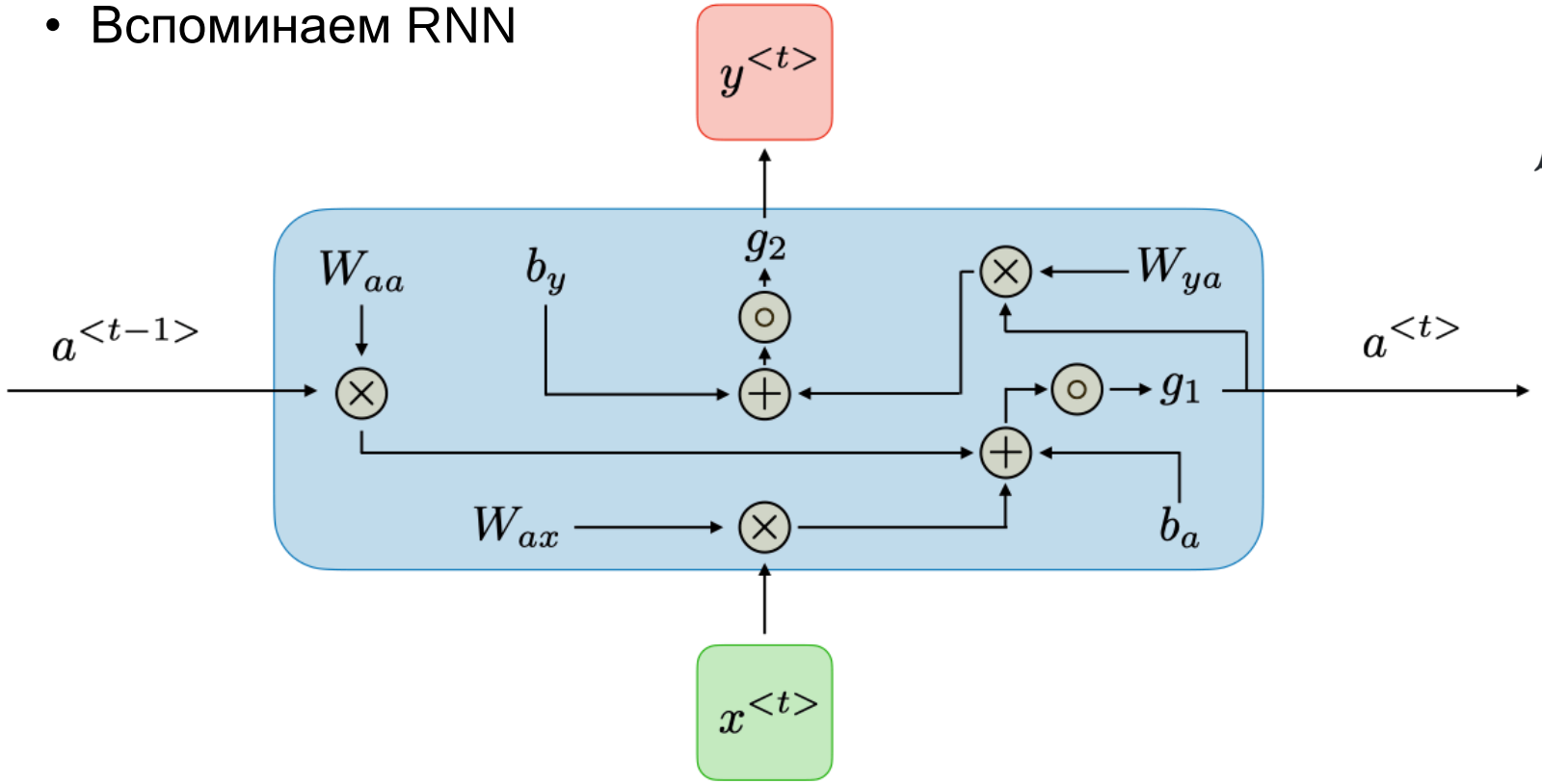
- Вспоминаем RNN



Recurrent Neural Network

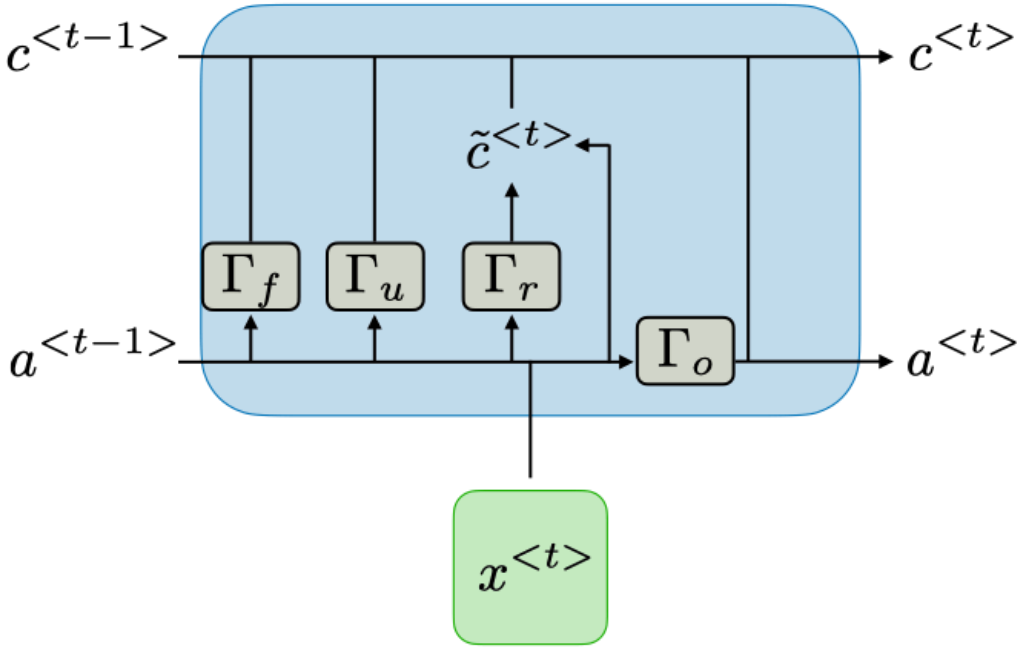
- Вспоминаем RNN

$$a^{<t>} = g_1(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a)$$
$$y^{<t>} = g_2(W_{ya}a^{<t>} + b_y)$$
$$\mathcal{L}(\hat{y}, y) = \sum_{t=1}^{T_y} \mathcal{L}(\hat{y}^{<t>}, y^{<t>})$$

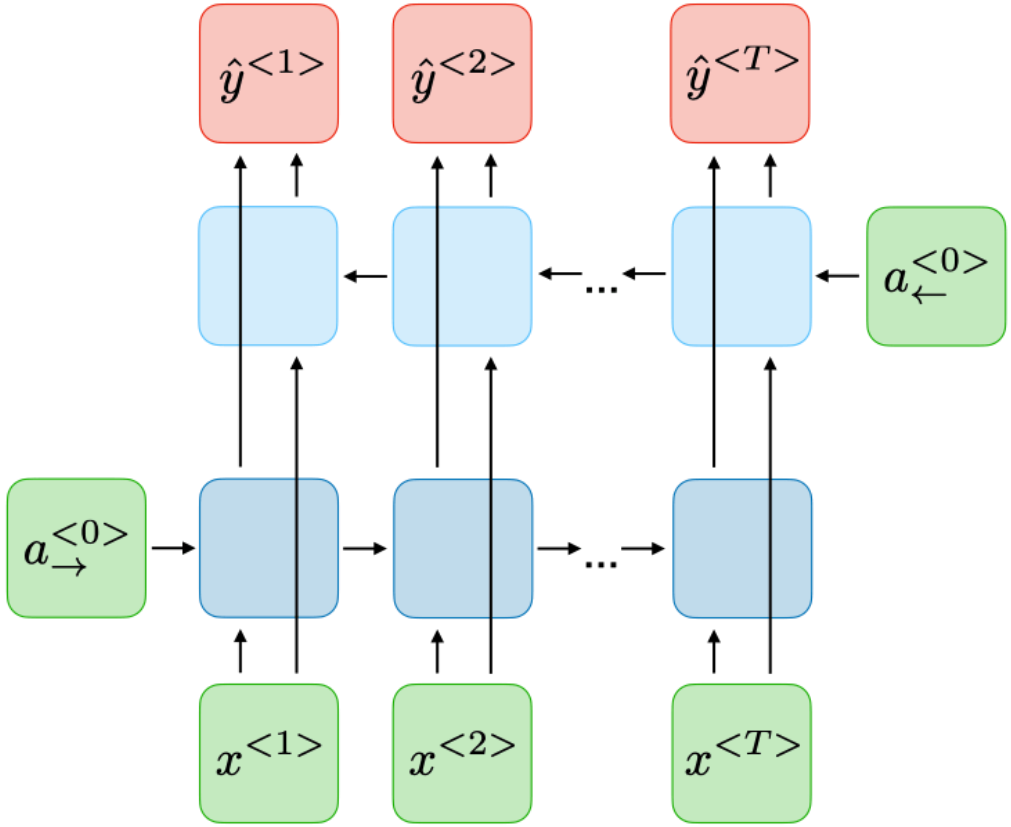


Recurrent Neural Network

Long Short Term Memory (LSTM)

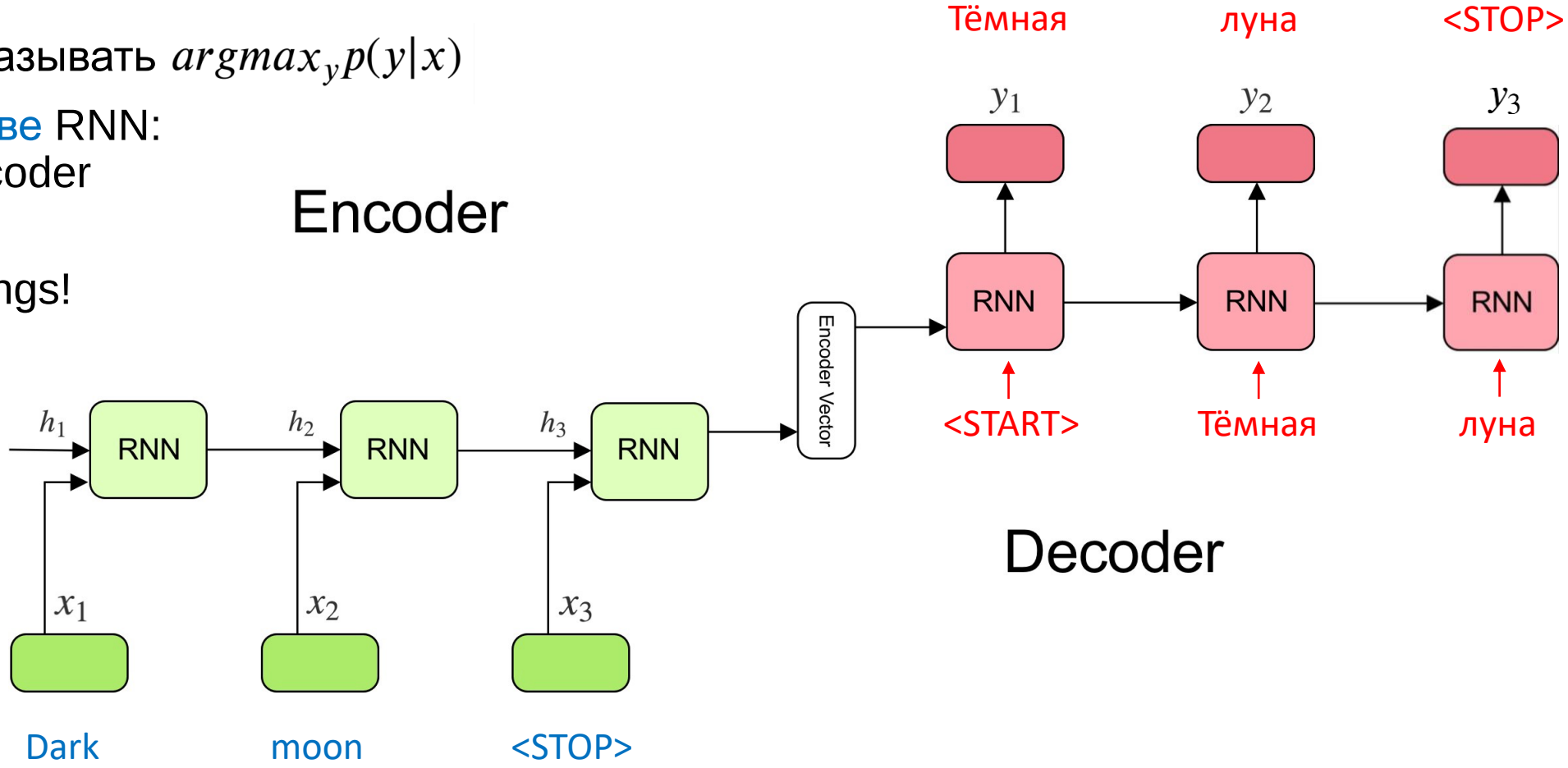


Bidirectional RNN

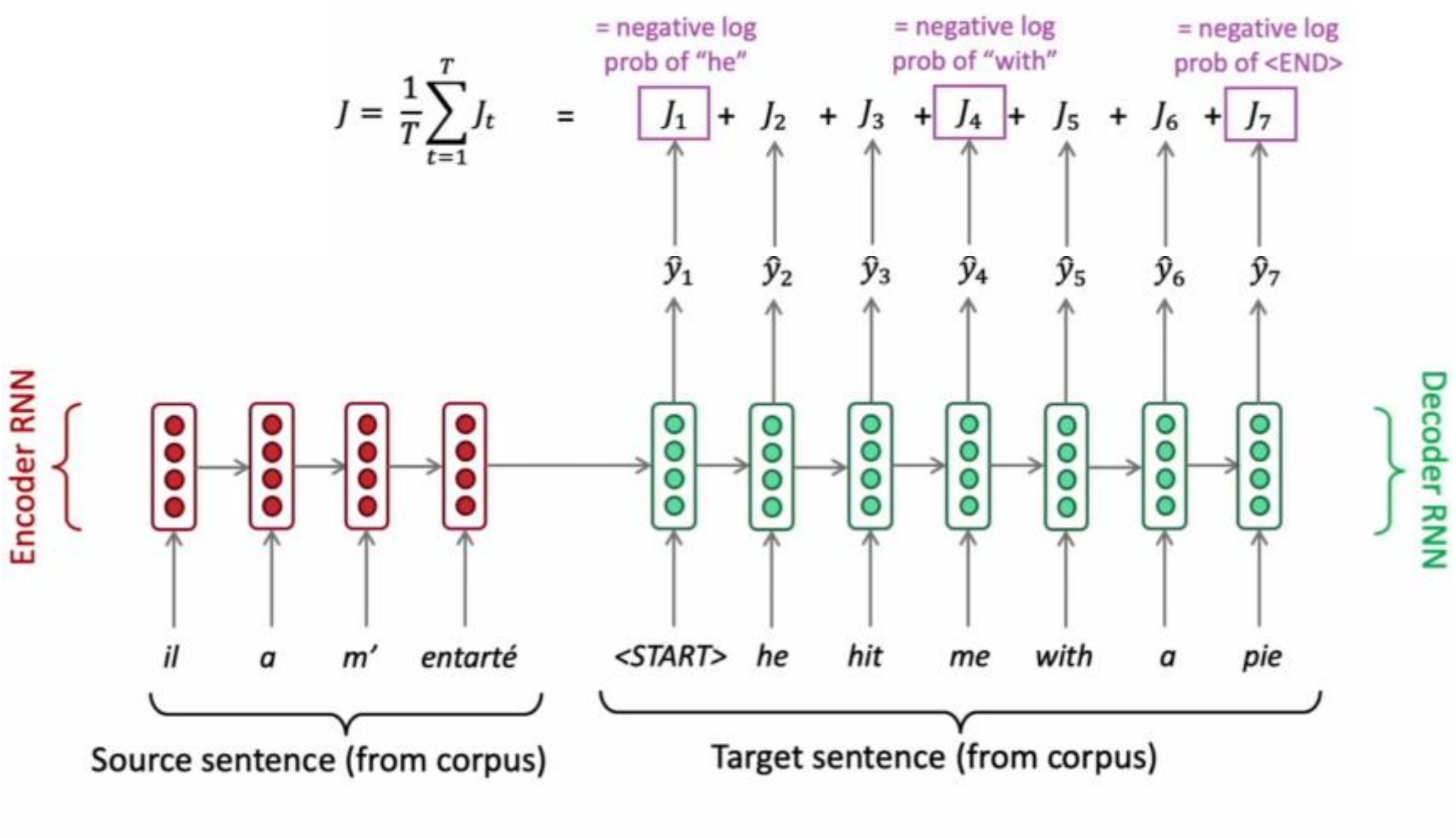


Sequence-to-Sequence (seq2seq)

- Хотим предсказывать $argmax_y p(y|x)$
- Используем две RNN: Encoder и Decoder
- На вход word embeddings!

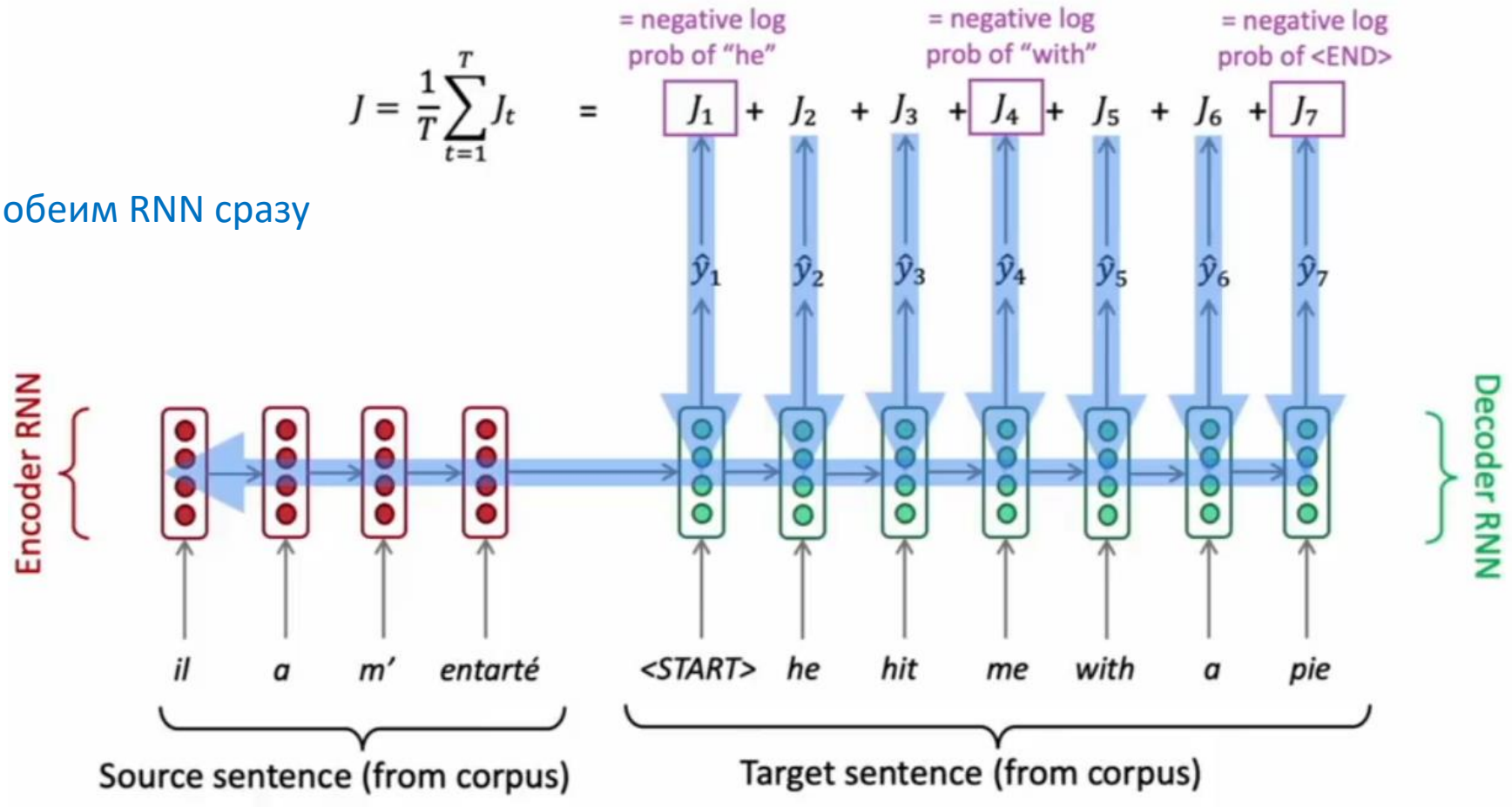


Sequence-to-Sequence: обучение



Sequence-to-Sequence: обучение

Backprop по обеим RNN сразу



Sequence-to-Sequence: greedy decoding

- Decoder каждый раз выдаёт *argmax* вероятностного распределения на словаре (Note: здесь условные вероятности по x, нотация опущена)

$$p(y) = p(y_0)p(y_1|y_0)p(y_2|y_0, y_1)\dots p(y_n|y_0, \dots, y_{n-1})$$

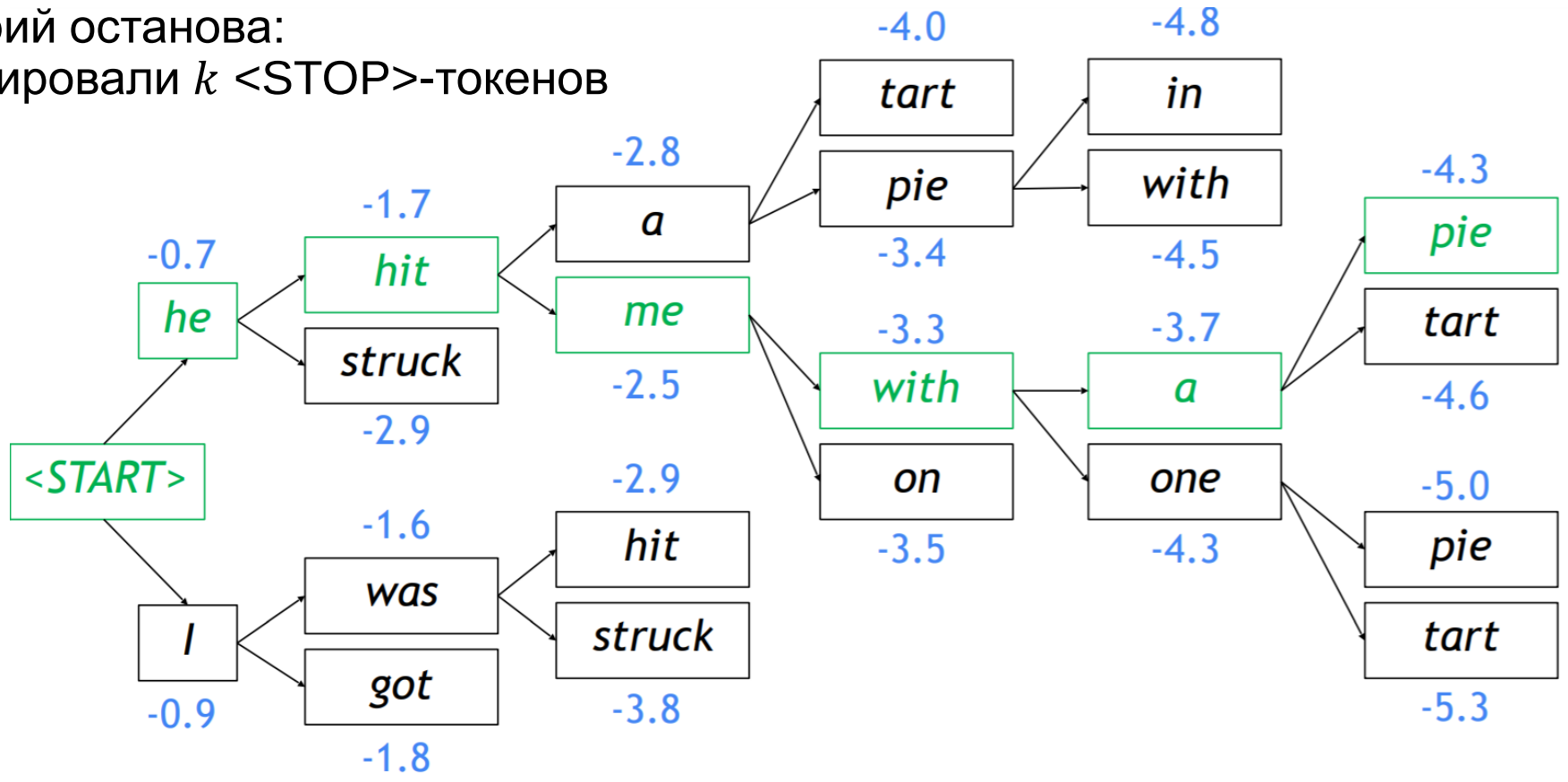
- Проблема: Иногда может быть, что на первом шаге Decoder не угадывает. Тогда всё предсказание рушится

$$\arg\max_y \prod_{t=1}^{n} p(y_t|y_{<t}, x) \neq \prod_{t=1}^{n} \arg\max_{y_t} p(y_t|y_{<t}, x)$$

- Решение: beam search

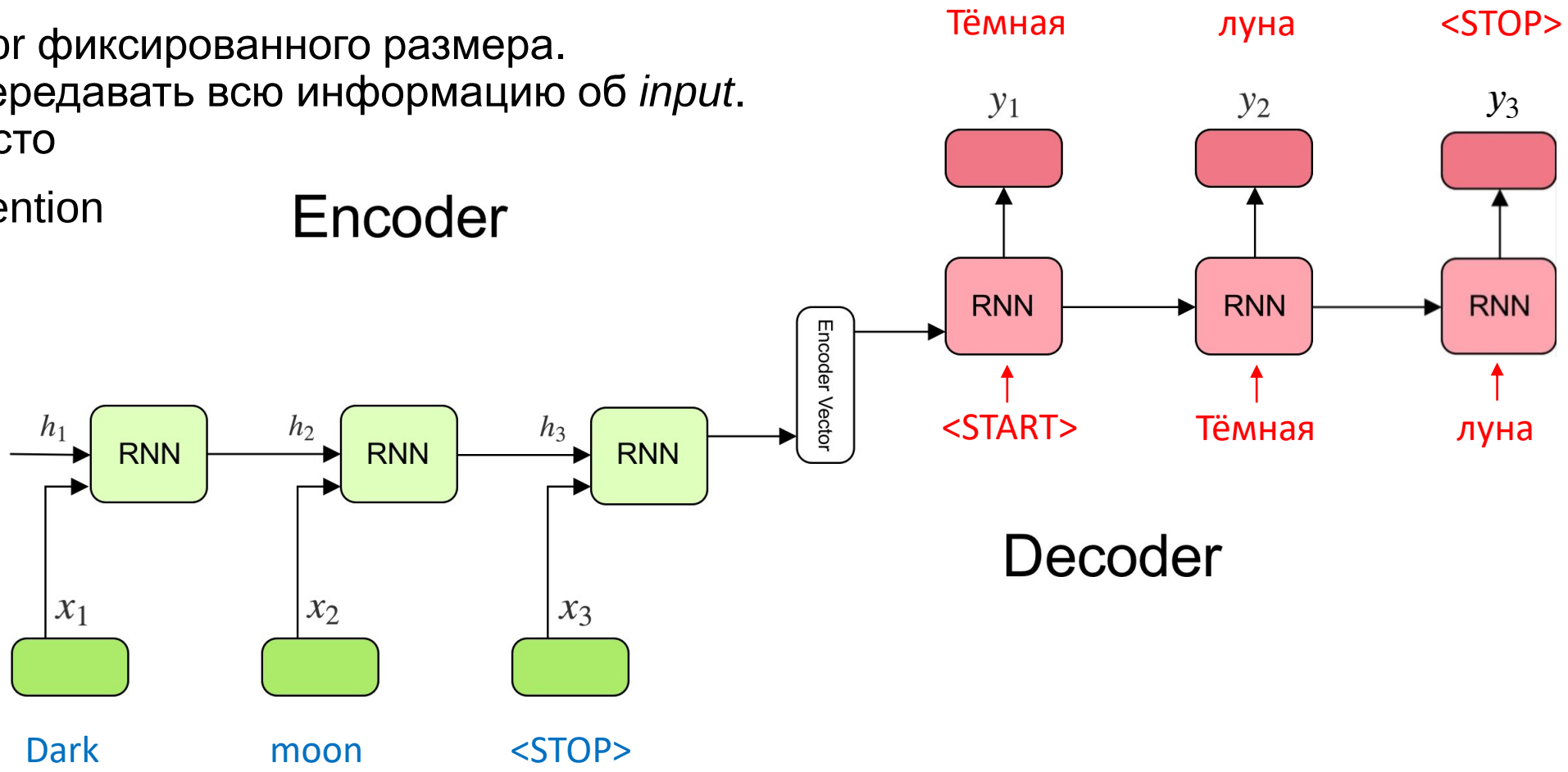
Sequence-to-Sequence: beam search

- Поддерживаем k самых вероятных последовательностей. Пример для $k = 2$:
- Критерий останова:
  сгенерировали k <STOP>-токенов

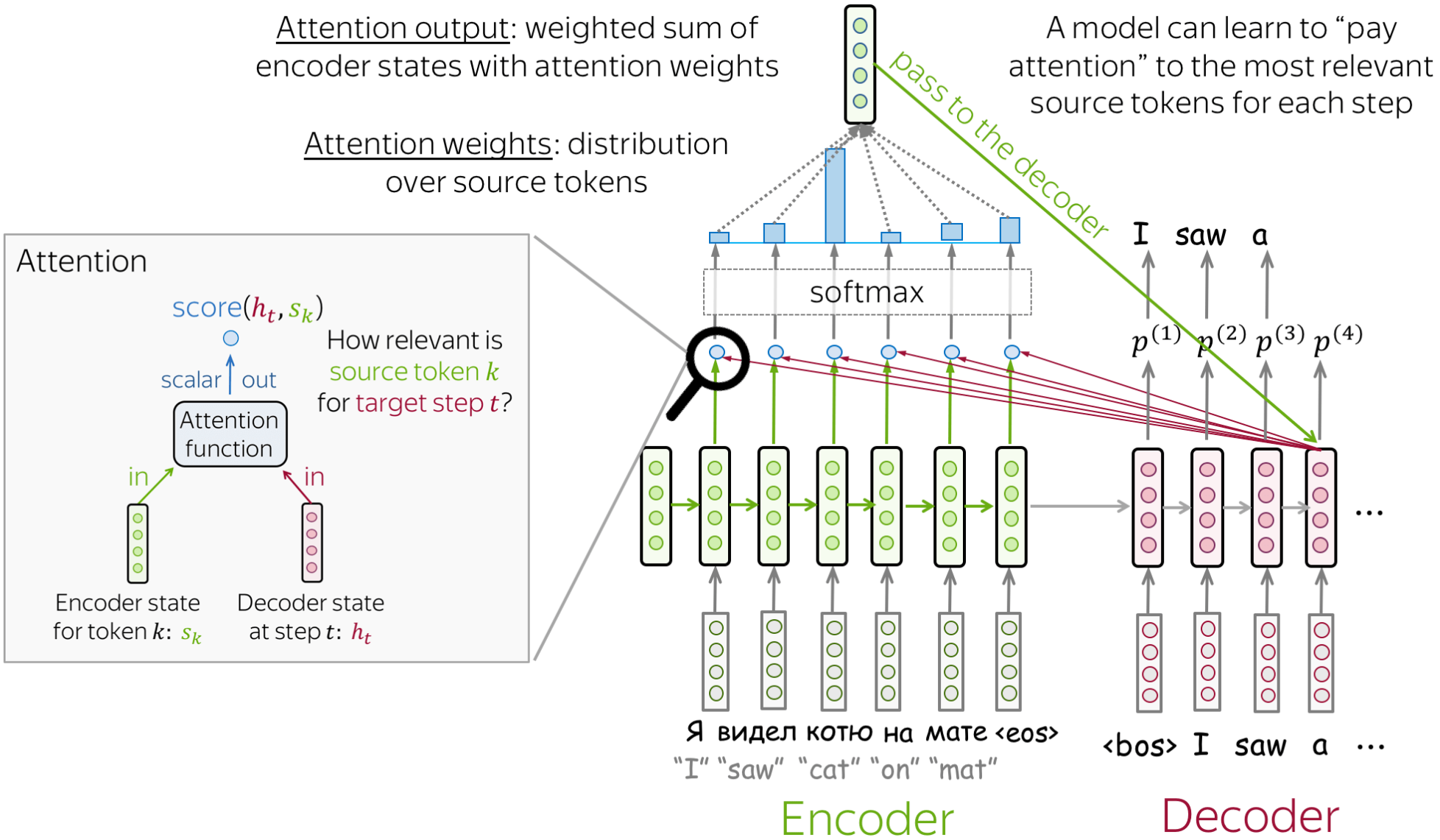


Sequence-to-Sequence: bottleneck problem

- Проблема:
  Encoder Vector фиксированного размера.
  Он должен передавать всю информацию об *input*.
  Это узкое место
- Решение: Attention



Sequence-to-Sequence with Attention



Sequence-to-Sequence: слой Attention

Для hidden state t-й итерации декодера h_t и каждого вектора s_k из *Encoder states*:

1. Вычисляем **attention score** $e_k = score(h_t, s_k)$
   (например, скалярное произведение $< h_t, s_k >$)
2. Вычисляем **attention distribution** $(a_0, \dots a_T) = softmax(e_0, \dots, e_T)$
3. Взвешиваем все s_k через a_k: $o_t = \sum_{i=0}^{T} s_i a_i$ – **attention output**
4. Конкатенируем attention output к h_t в декодере

Слой Attention фактически указывает *контекст*: чем больше слово из энкодера влияет на текущий hidden state декодера, тем сильнее оно участвует в предсказании.

Метрика BLEU

1. BiLingual Evaluation Understudy
2. Оценивает схожесть переводов
3. MT – перевод модели
   ref – истинный перевод

$$p_n = \frac{\#n\text{-грамм в MT и ref}}{\#n\text{-грамм в MT}}$$

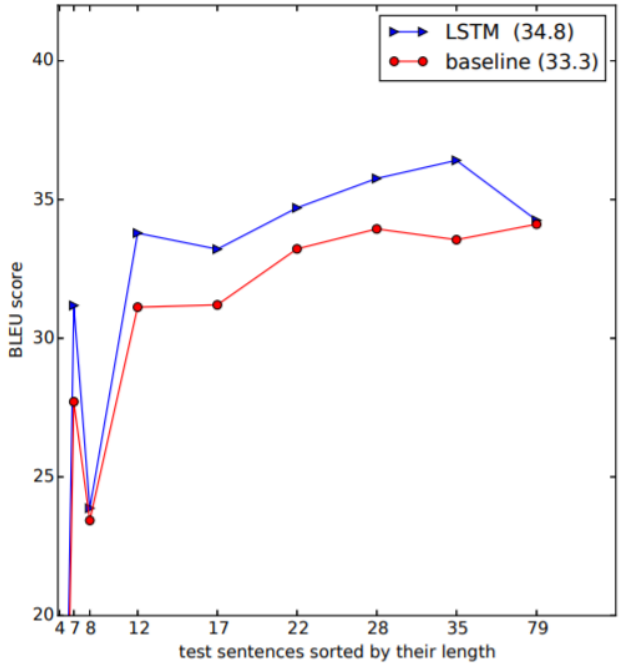
$$\beta = e^{\min(0,1-\frac{\text{len}_{\text{ref}}}{\text{len}_{\text{MT}}})}$$

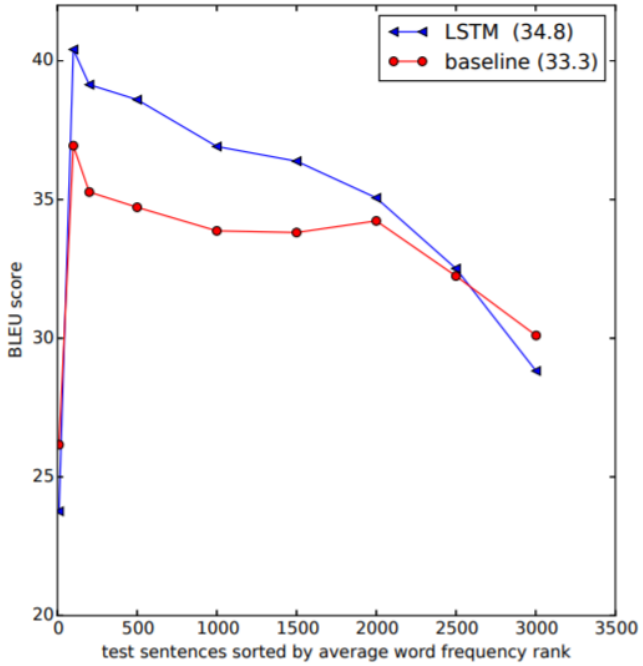
$$w_n = 1/2^n$$

$$\text{BLEU} = \beta \prod_{i=1}^{k} p_n^{w_n}$$

BLEU для seq2seq

- Baseline – лучшая SMT-модель
- Результаты 2014 года

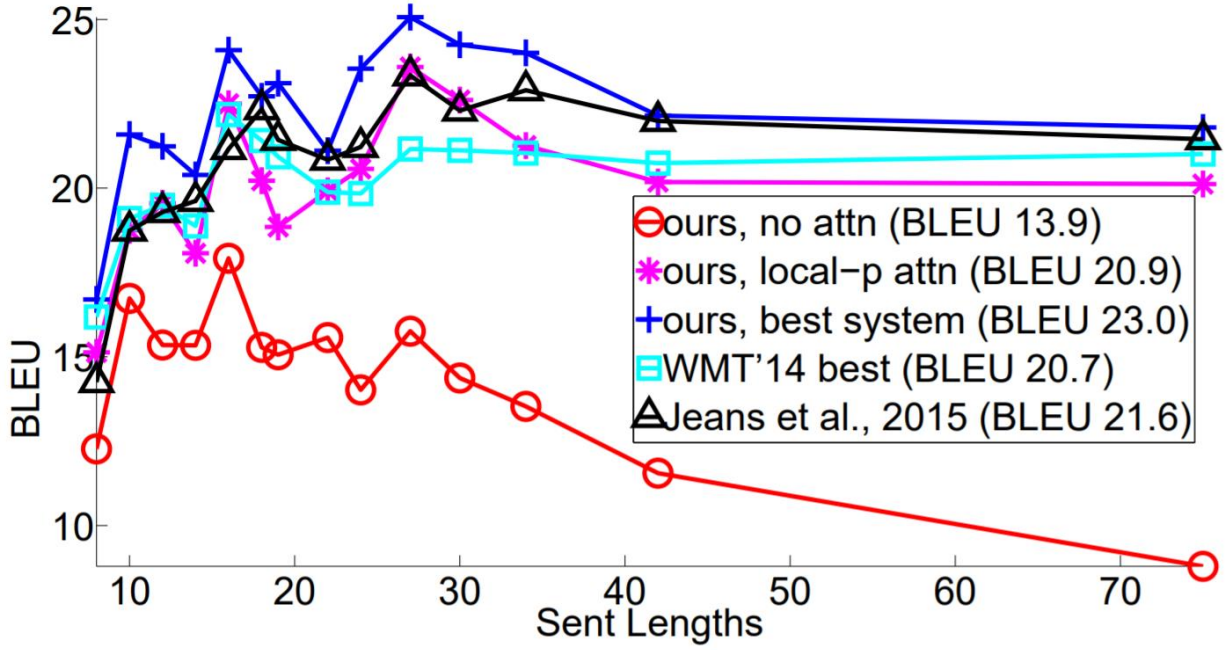




Method	test BLEU score (ntst14)
Bahdanau et al. [2]	28.45
Baseline System [29]	33.30
Single forward LSTM, beam size 12	26.17
Single reversed LSTM, beam size 12	30.59
Ensemble of 5 reversed LSTMs, beam size 1	33.00
Ensemble of 2 reversed LSTMs, beam size 12	33.27
Ensemble of 5 reversed LSTMs, beam size 2	34.50
Ensemble of 5 reversed LSTMs, beam size 12	**34.81**

BLEU для seq2seq

- WMT’14 – лучшая SMT-модель
- Результаты 2015 года



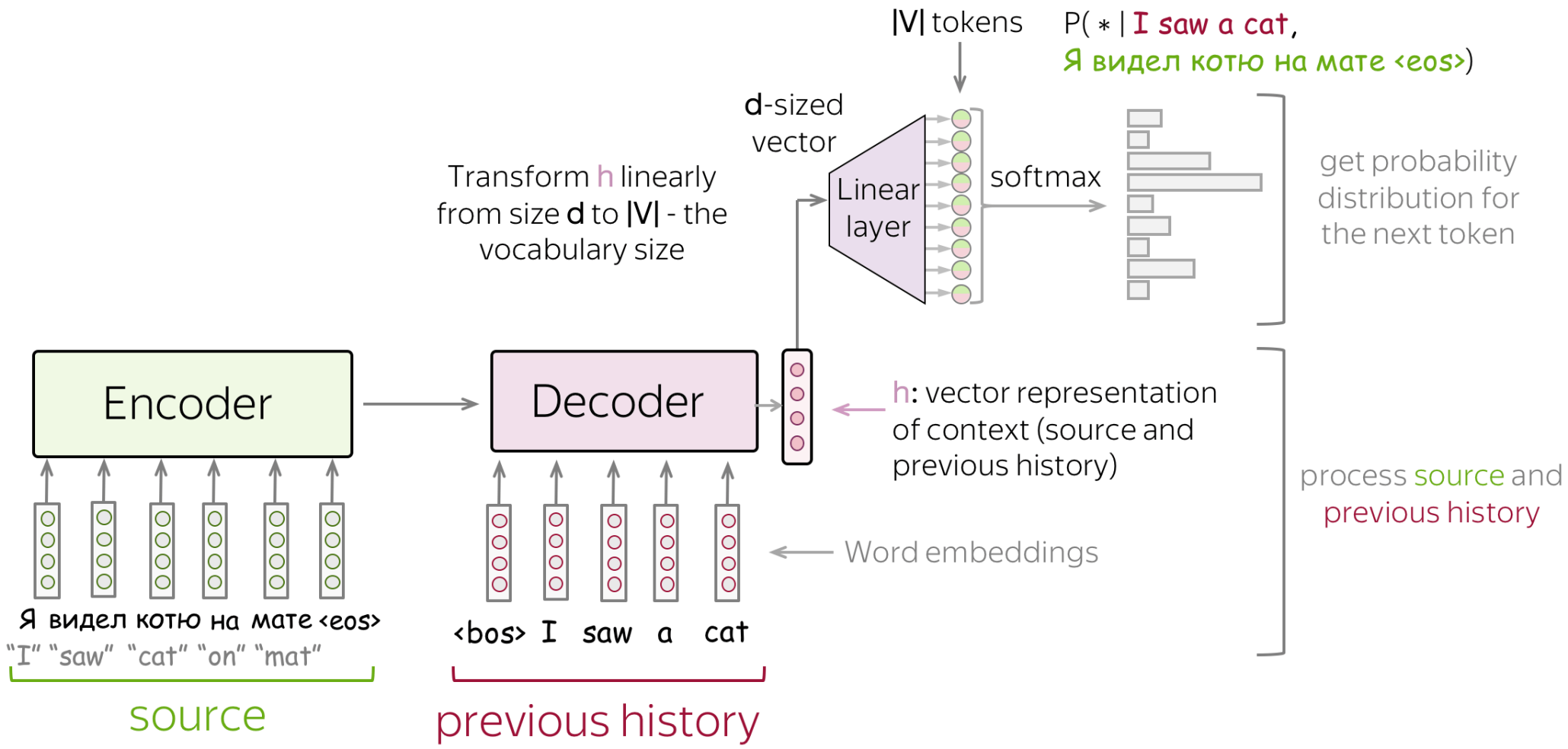
Заключение

- Статистический машинный перевод решает декомпозированную задачу и показывает хороший результат
- Seq2seq генерирует текст напрямую и показывает результат лучше
- СМТ учит две задачи, seq2seq – одну
- Лучше комбинировать

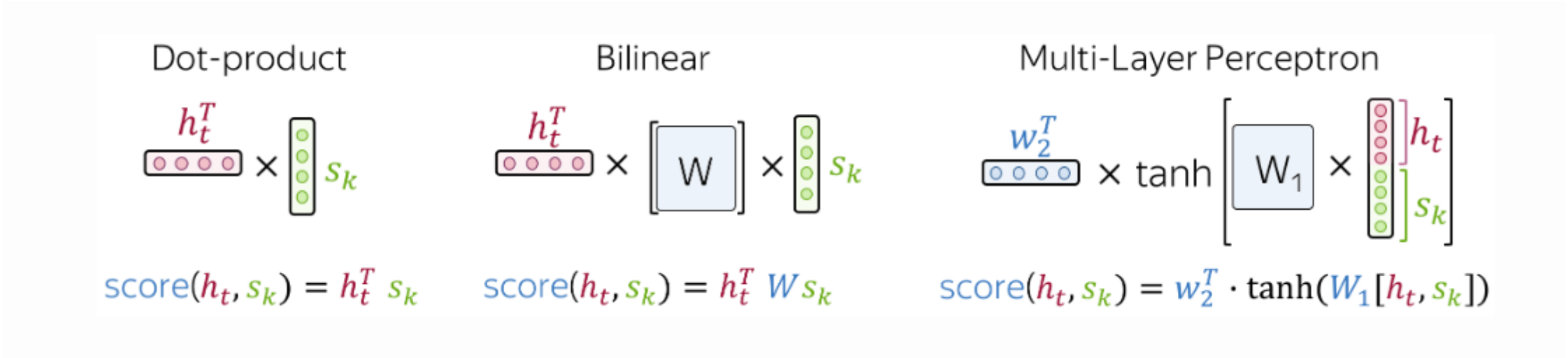
Источники

- Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation – Cho et al.
- Neural Machine Translation by Jointly Learning to Align and – Bahdanau et al.
- Statistical Machine Translation -- Philipp Koehn
- Natural Language Processing with Deep Learning -- Abigail See, Matthew Lamm
- Effective Approaches to Attention-based Neural Machine Translation – Luong et al.
- Визуал: https://vas3k.ru/blog/machine_translation/
- Визуал: https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks
- Визуал: https://guillaumegenthial.github.io/sequence-to-sequence.html
- Визуал: https://lena-voita.github.io/nlp_course/seq2seq_and_attention.html

Extras 1: another seq2seq scheme



Extras 2: score function examples



Extras 3: Bahdanau et al. Model