

Яндекс

Yandex Translate

Neural Machine Translation

David Talbot

Яндекс Переводчик

ТЕКСТ

САЙТ

КАРТИНКА



• АНГЛИЙСКИЙ

The animal didn't cross the street because it was too tired.

60 / 10000

Яндекс Переводчик

ТЕКСТ САЙТ КАРТИНКА



● АНГЛИЙСКИЙ

The animal didn't cross the street because it was too tired.

60 / 10000

РУССКИЙ



Животное не пересечь улицу, потому что он слишком устал.

Яндекс Переводчик

ТЕКСТ САЙТ КАРТИНКА



• АНГЛИЙСКИЙ

The animal didn't cross the street because it was too tired.

РУССКИЙ



Животное не пересечь улицу, потому что он слишком устал.

РУССКИЙ



Животное не перешло улицу, потому что оно было слишком уставшим.

Problems with Phrase-based MT

- › Unrealistic independence assumptions
- › Curse of dimensionality
- › Impossible to optimize end-to-end
- › Reliance on hand-crafted language specific features

Neural Machine Translation

- › Direct model of conditional translation probability

$$e^* = \operatorname{argmax}_f \Pr(e) \Pr(f|e)$$

$$e^* = \operatorname{argmax}_e \Pr(e|f)$$

Neural Machine Translation

- › Direct model of conditional probability of translation given source

$$e^* = \operatorname{argmax}_f \Pr(e) \Pr(f|e)$$

$$e^* = \operatorname{argmax}_e \Pr(e|f)$$

$$= \prod_{i=1}^I \Pr(e_i | e_{j < i}, f)$$

Neural Machine Translation

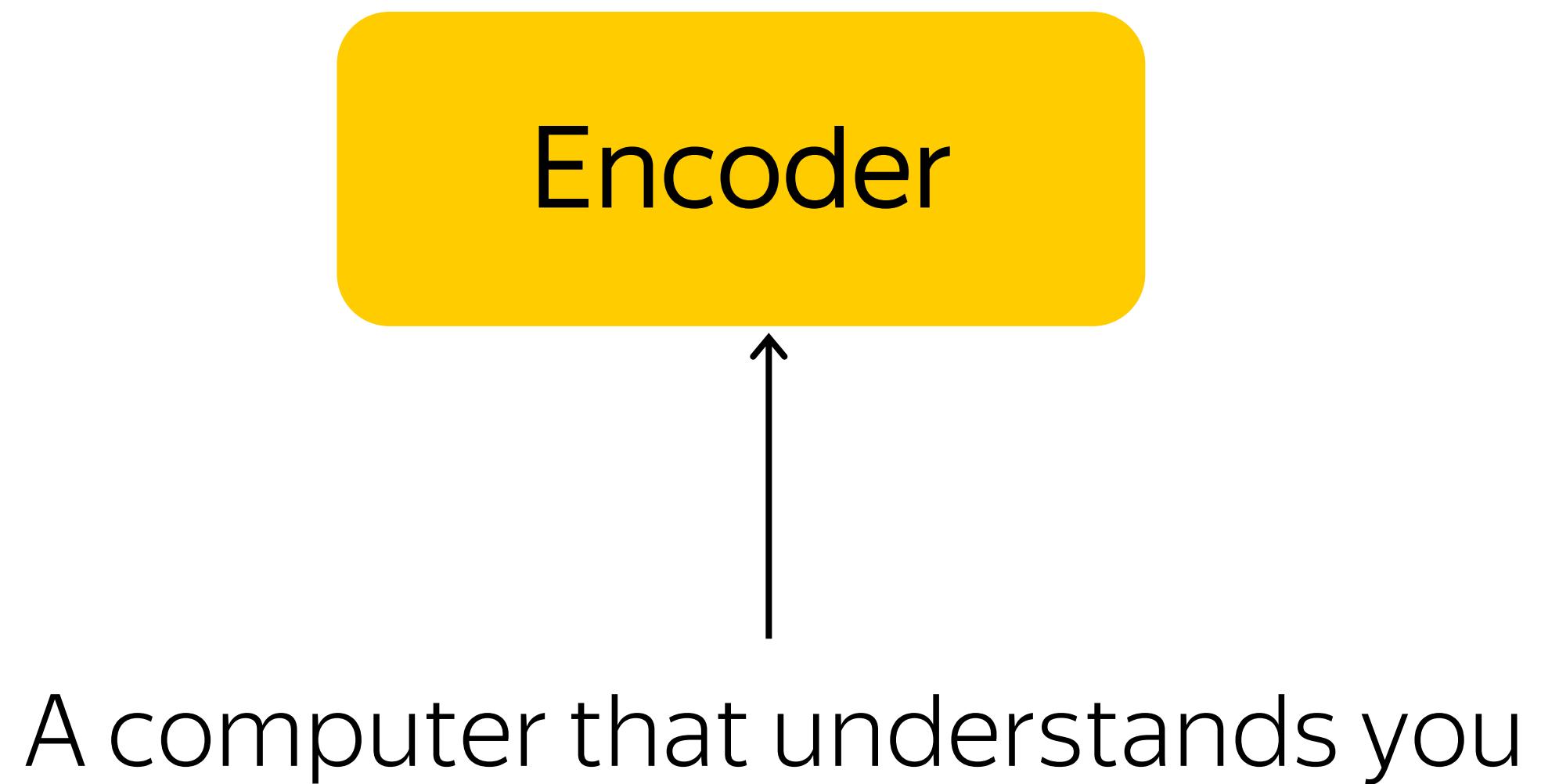
- › Direct model of conditional probability of translation given source

$$e^* = \operatorname{argmax}_f \Pr(e) \Pr(f|e)$$

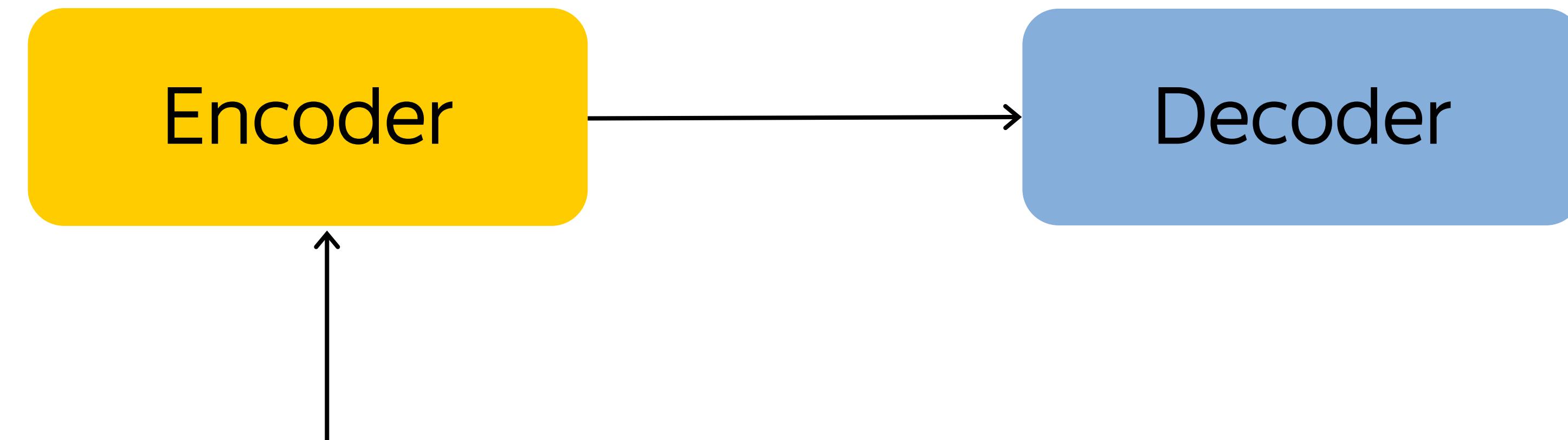
$$e^* = \operatorname{argmax}_e \Pr(e|f)$$

$$= \prod_{i=1}^I \Pr(e_i | e_{j < i}, f)$$

Neural Machine Translation

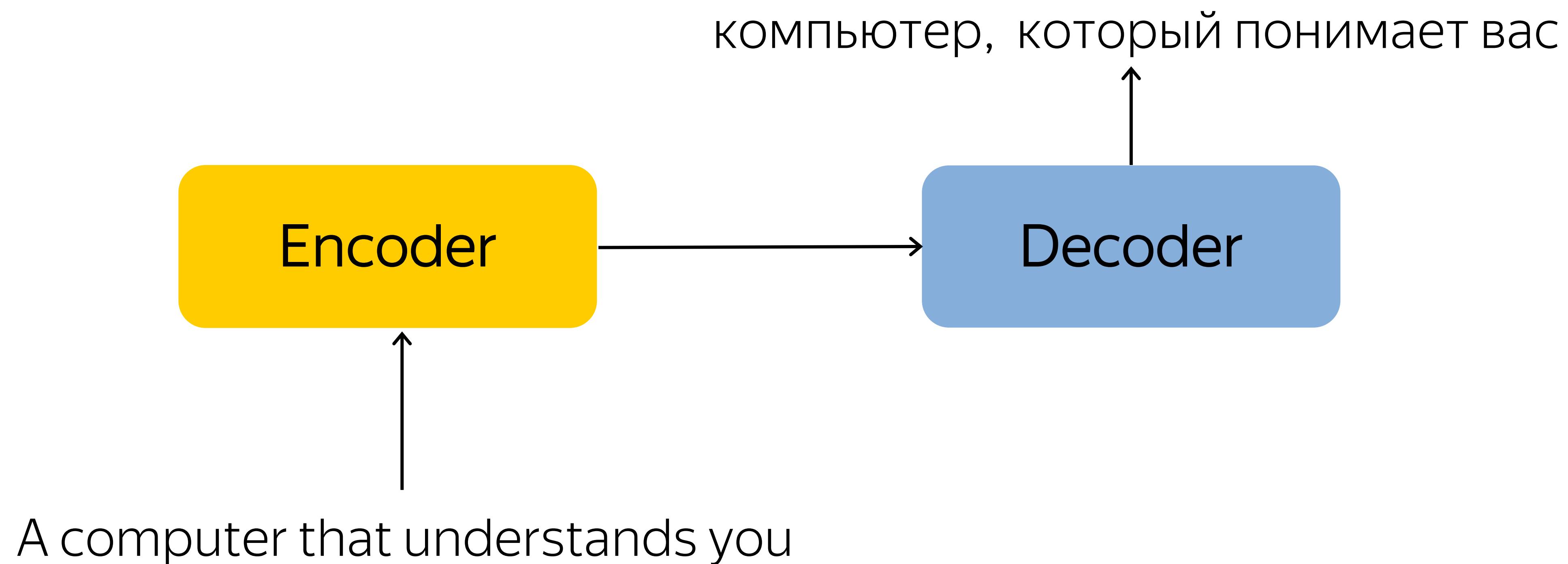


Neural Machine Translation

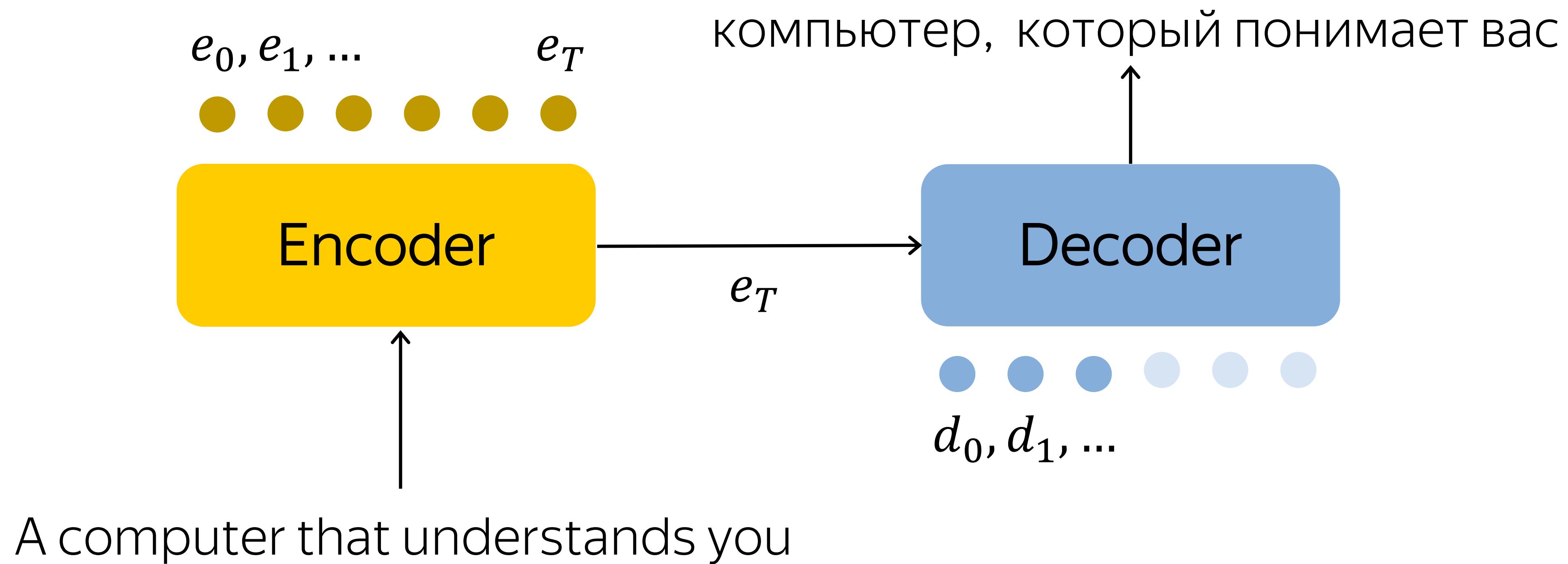


A computer that understands you

Neural Machine Translation



Neural Machine Translation



Encoder Decoder

- › No explicit independence assumptions
- › Learned representations (word embeddings)
- › End-to-end optimization

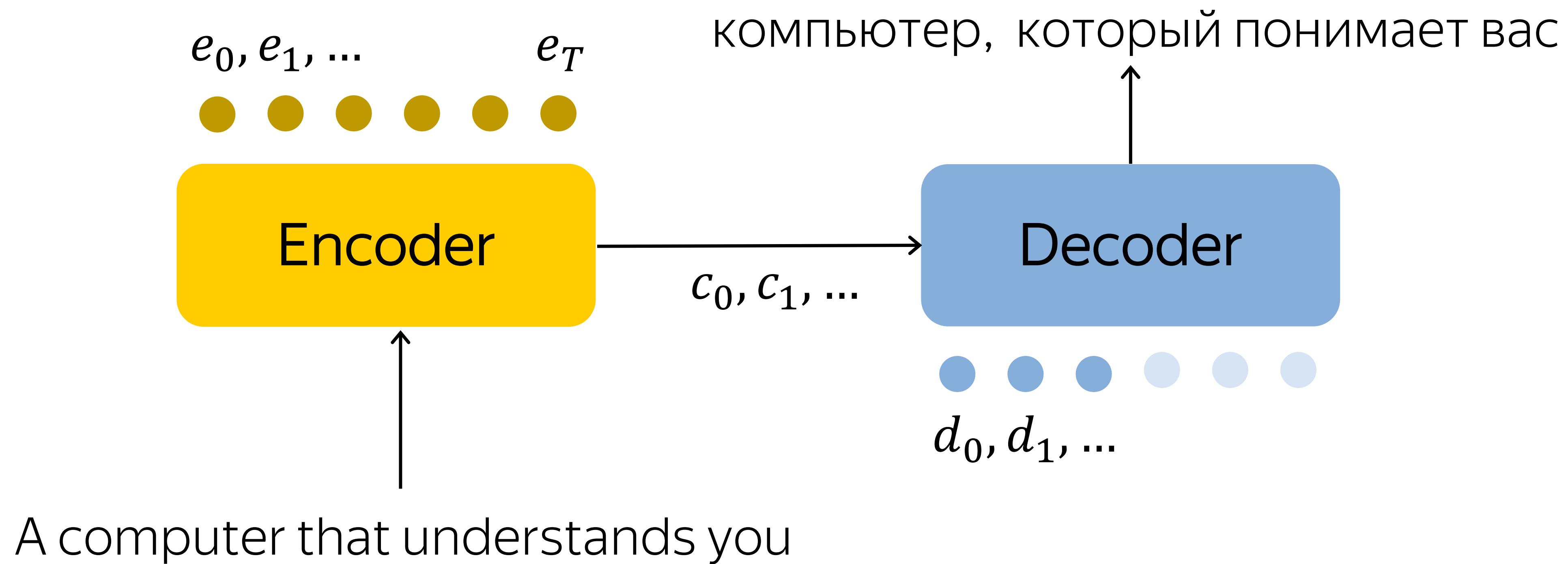
Problems with Vanilla Encoder Decoder

- › Poor performance on long sentences
- › Bias towards shorter candidates (?)
- › Fluent but inadequate output
- › No guarantee that all input words are translated

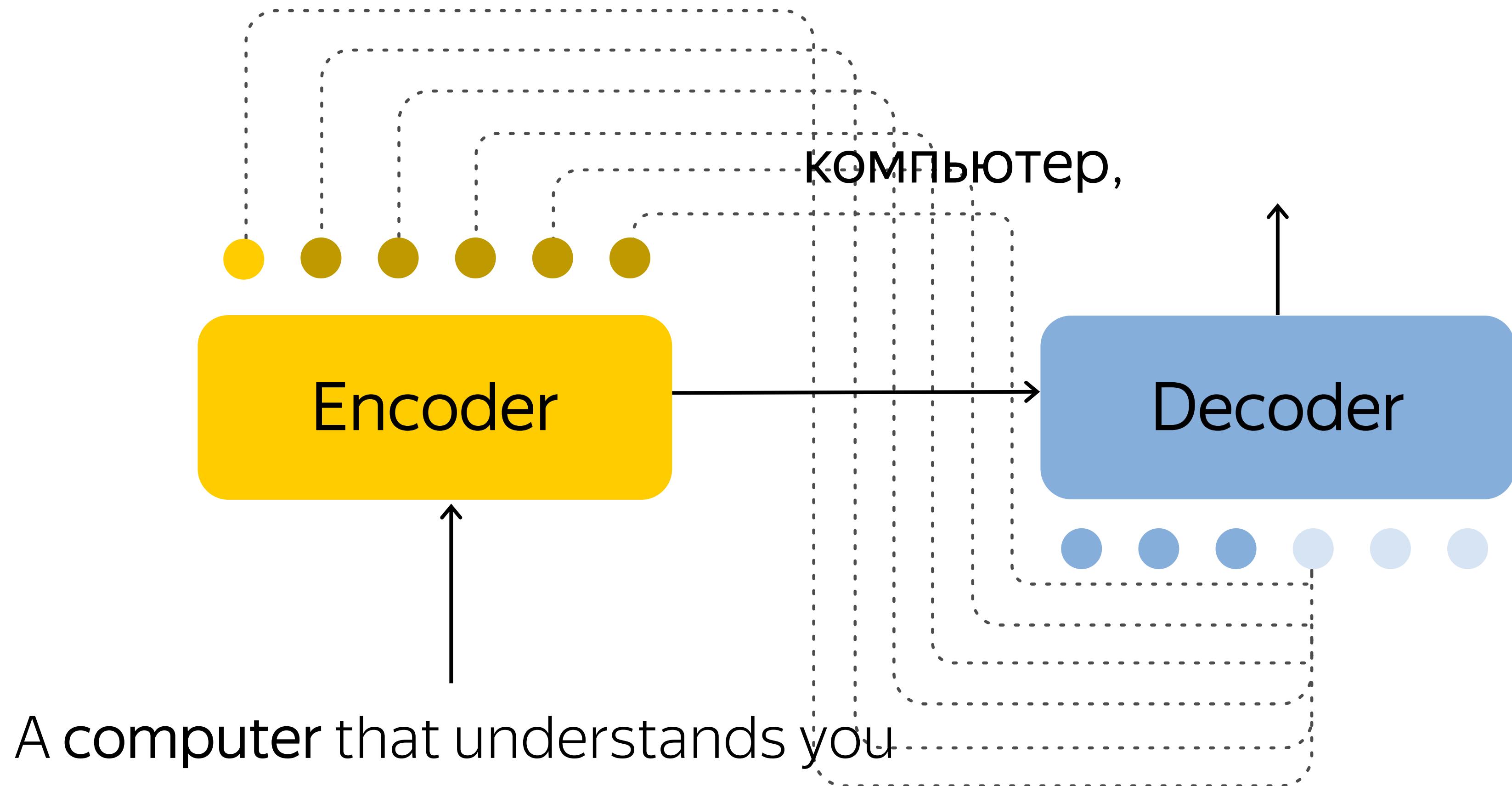
Attention

- › Allow model to scale with sentence length
- › Allow model to focus on specific information during translation
- › Not quite word alignments

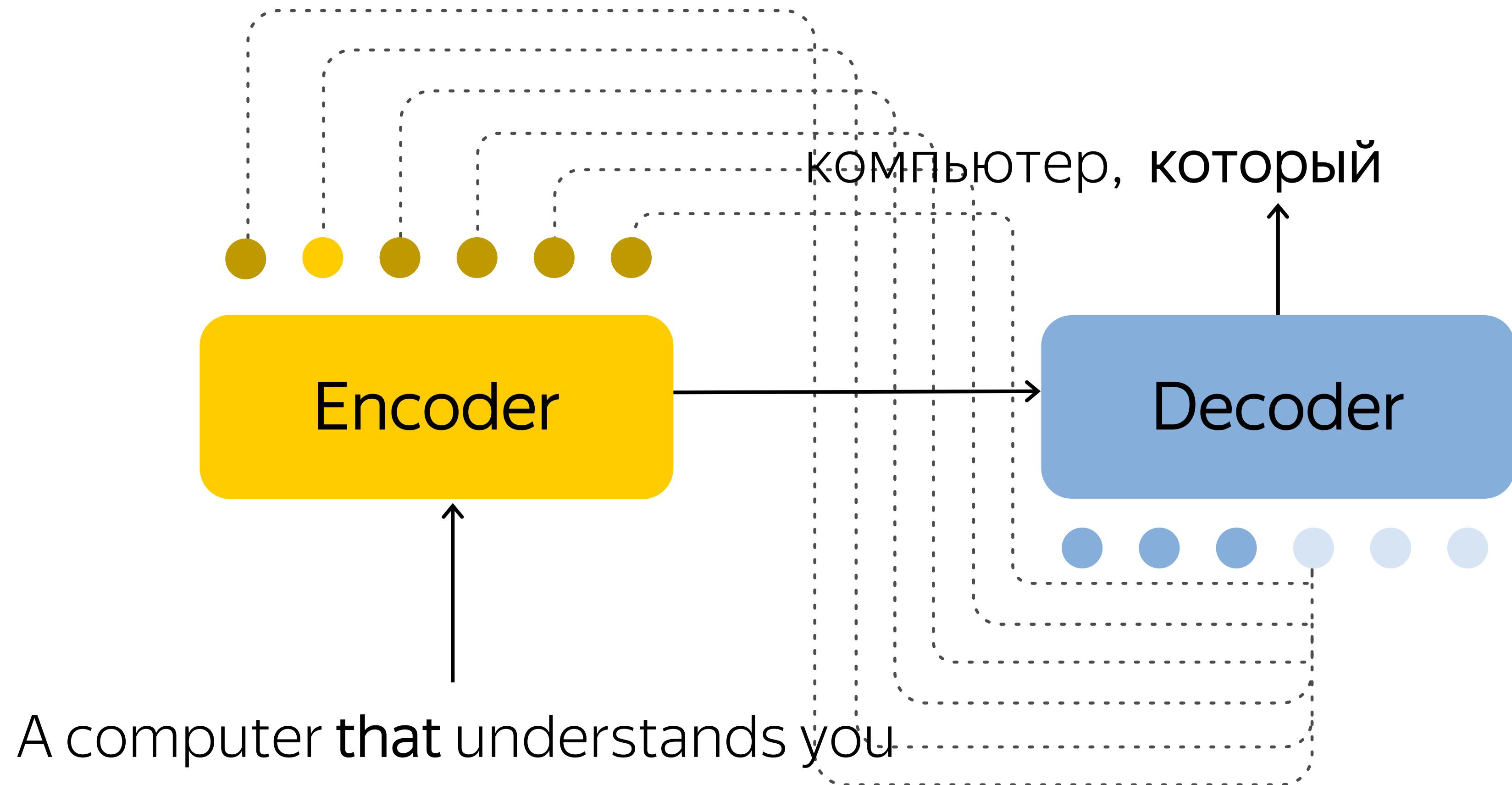
Neural Machine Translation



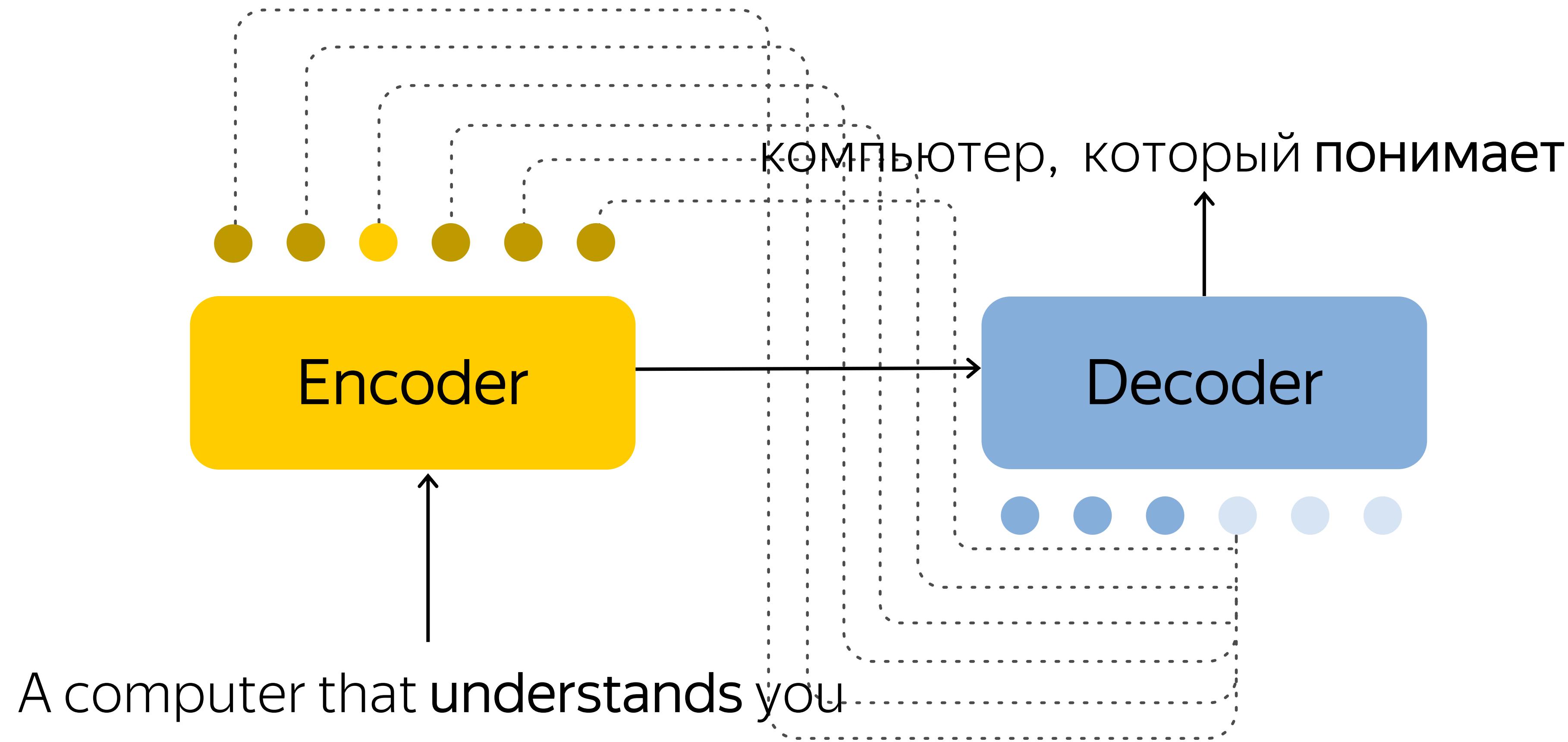
Neural Machine Translation



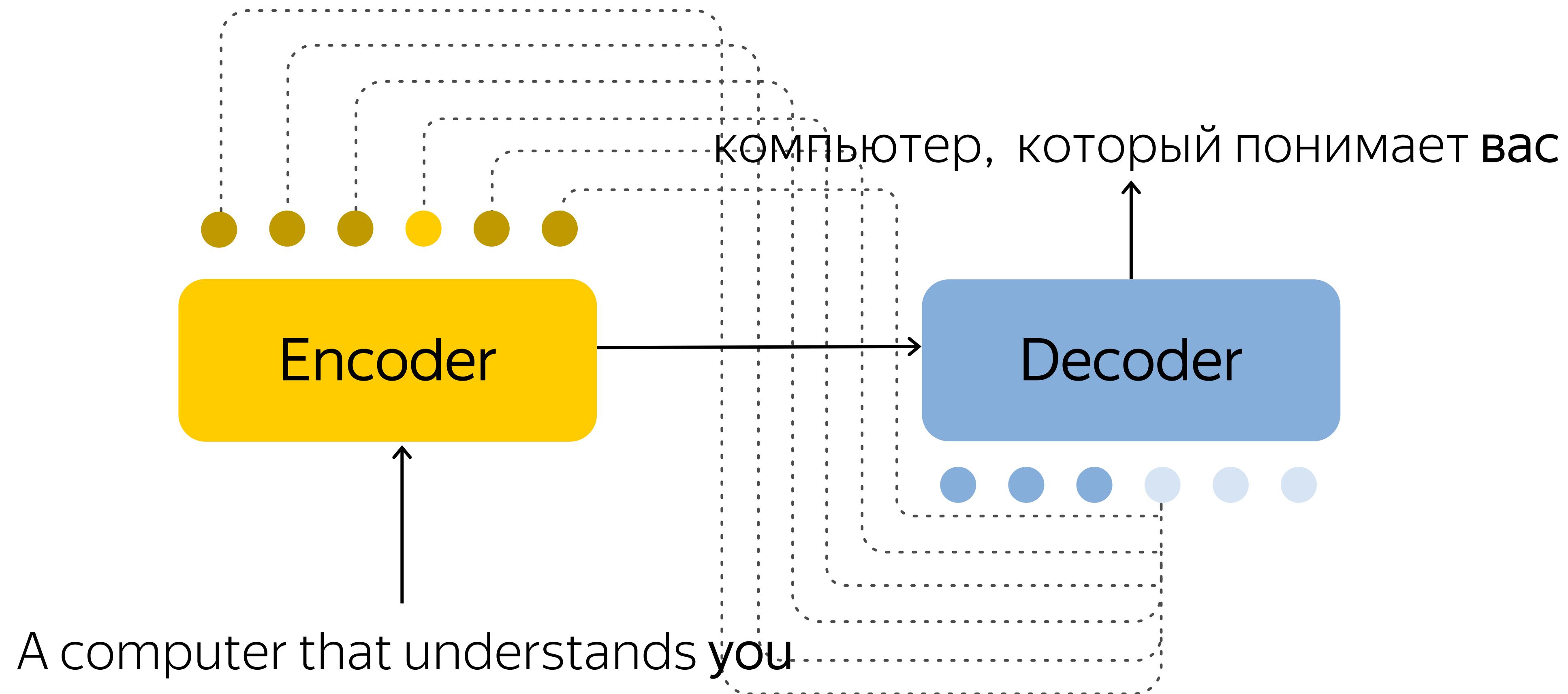
Neural Machine Translation



Neural Machine Translation

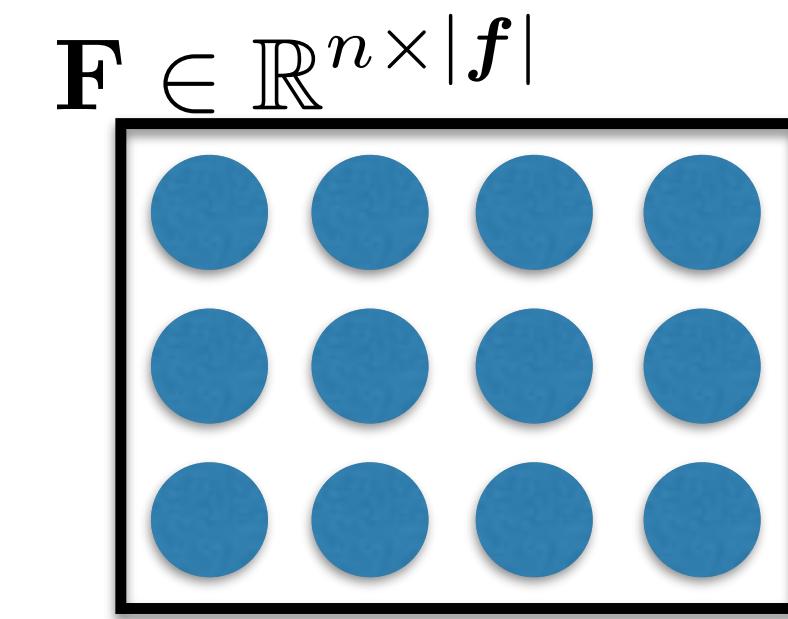
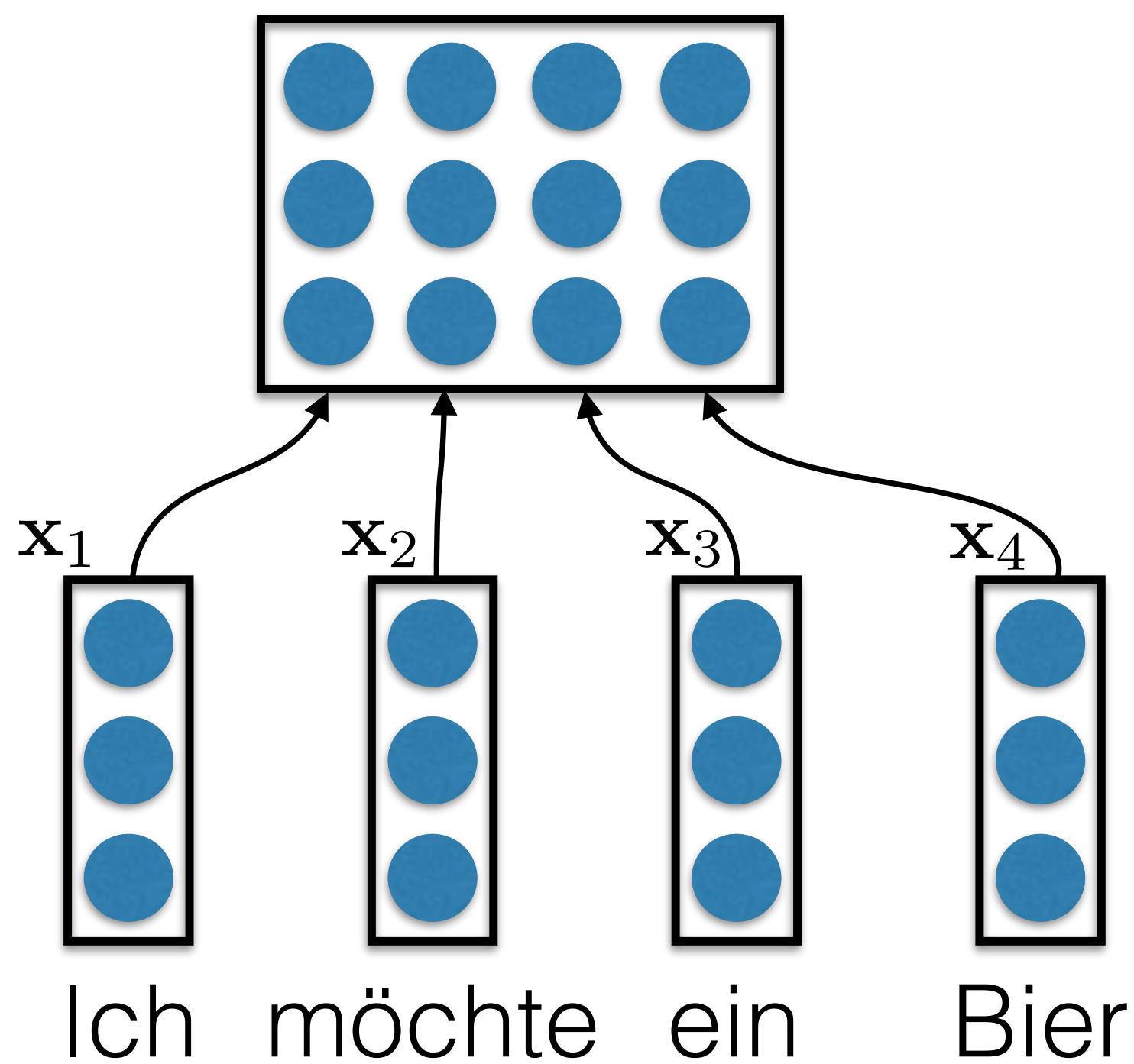


Neural Machine Translation



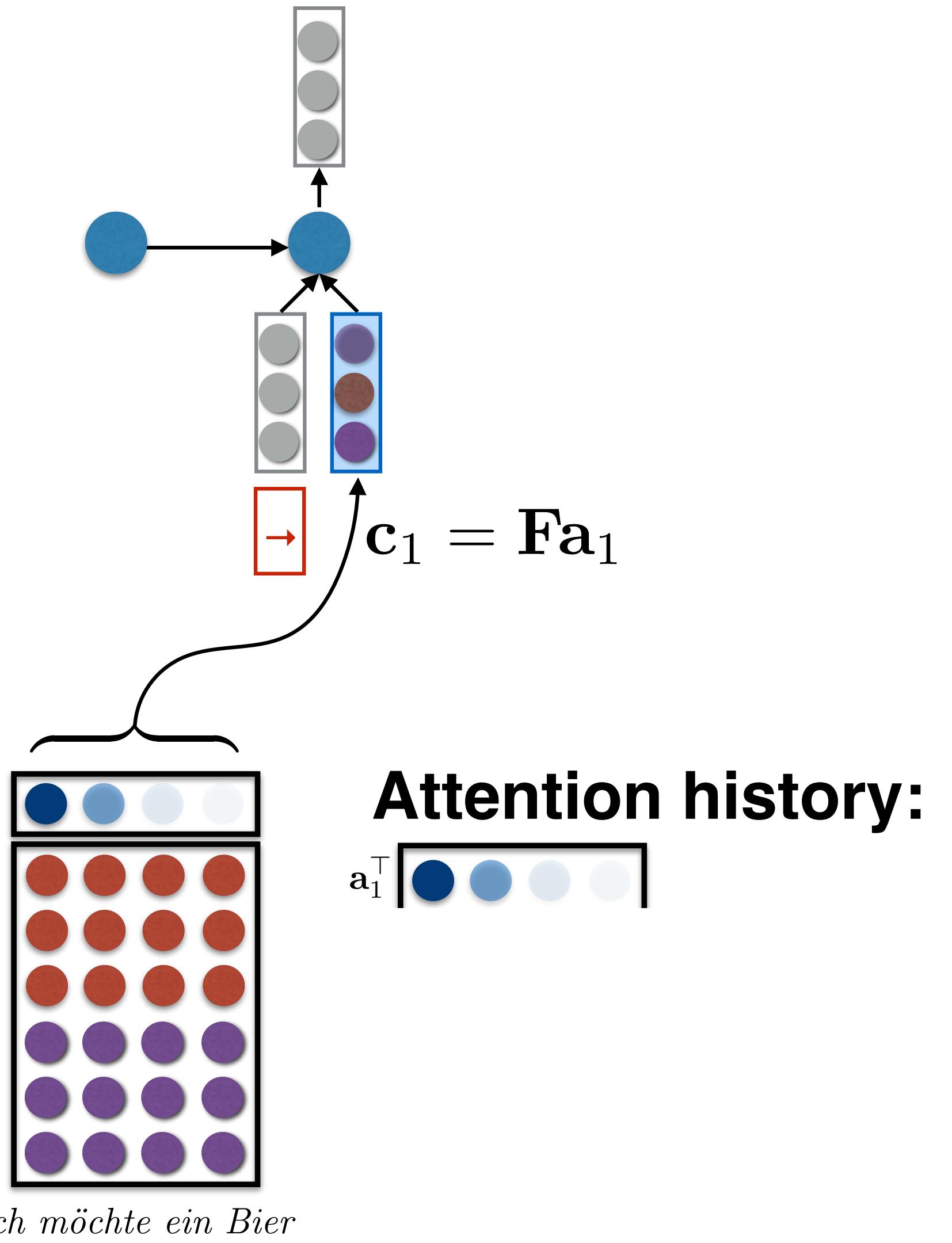
Encode Source into Matrix

$$\mathbf{f}_i = \mathbf{x}_i$$



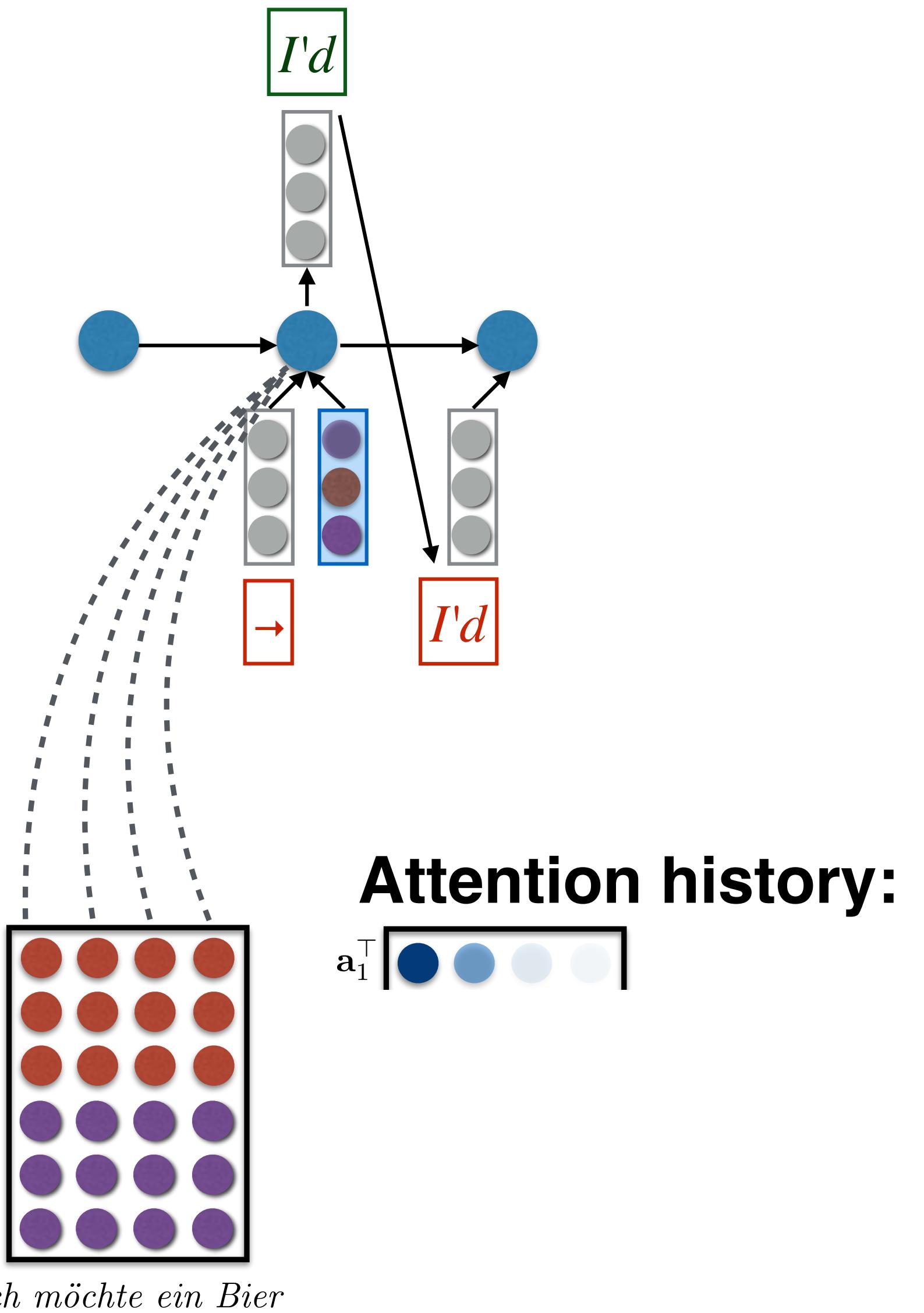
Ich möchte ein Bier

Decode With Attention



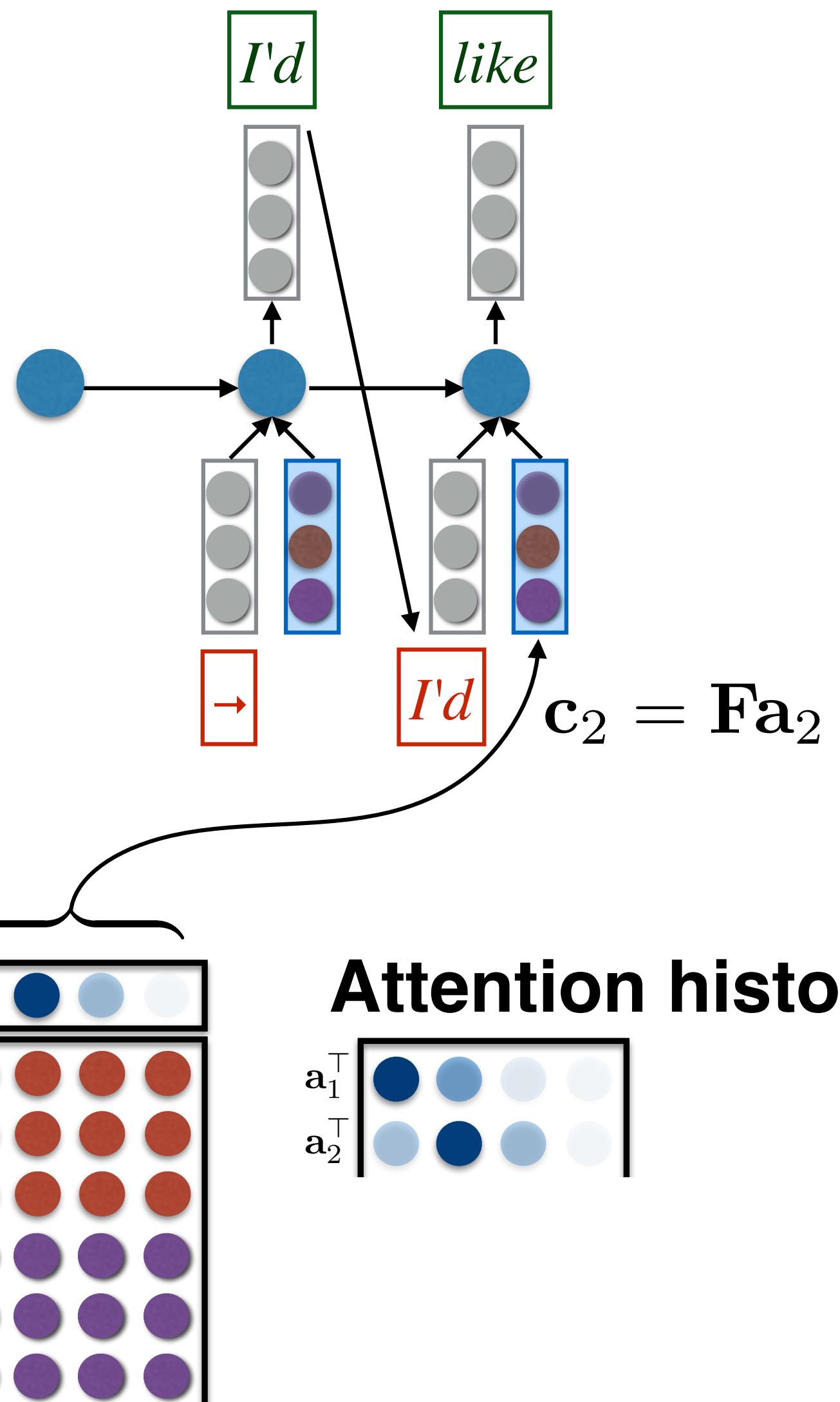
Images from
Chris Dyer

Decode with Attention



Images from
Chris Dyer

Decode with Attention



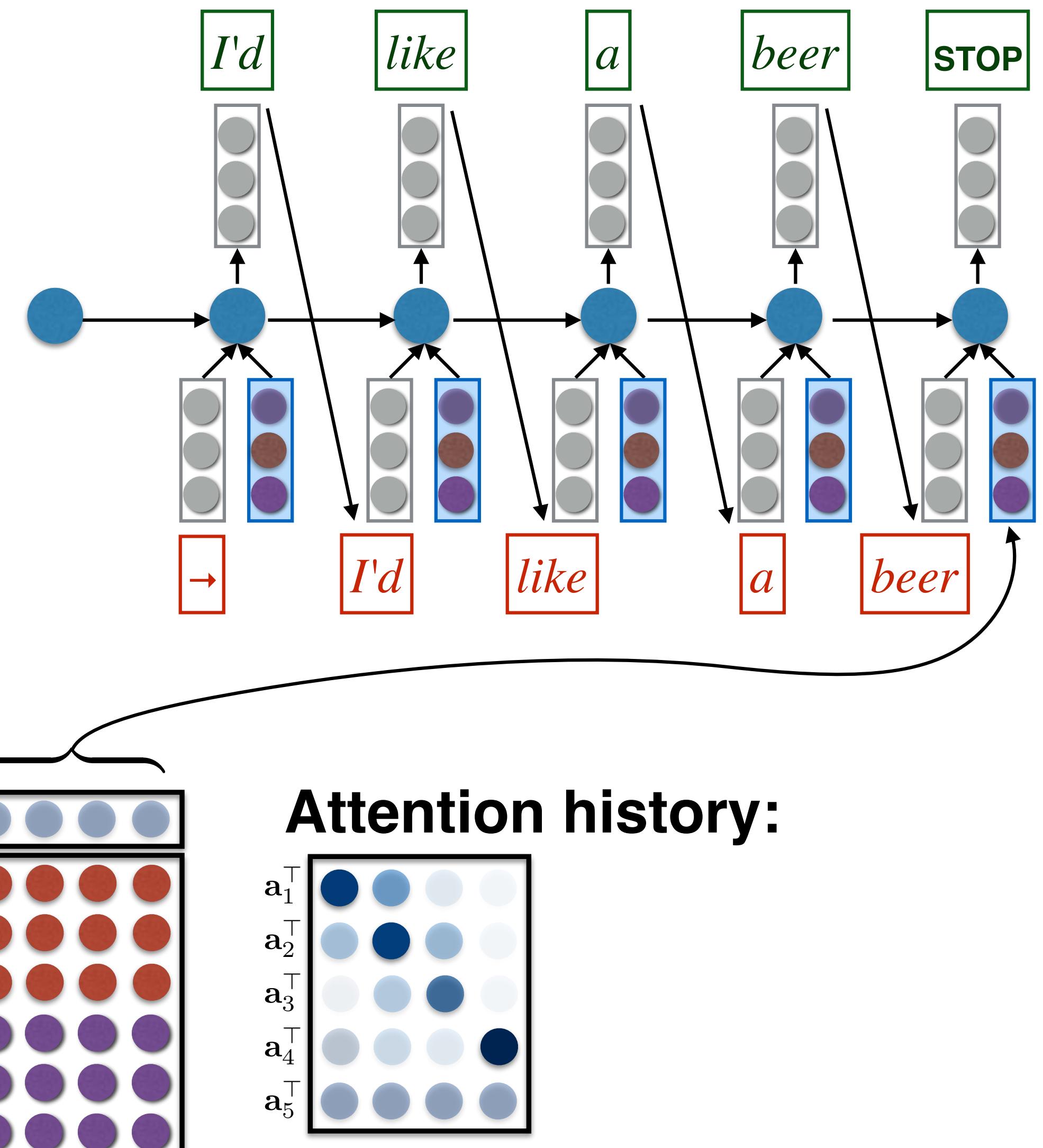
Ich möchte ein Bier

Attention history:

$$\begin{matrix} \mathbf{a}_1^\top & \text{blue} & \text{red} & \text{purple} \\ \mathbf{a}_2^\top & \text{blue} & \text{red} & \text{purple} \end{matrix}$$

Images from
Chris Dyer

Decode with Attention



Ich möchte ein Bier

Images from
Chris Dyer

Encoder-Decoder With Attention

- › Removes bottleneck of vanilla Encoder-Decoder
- › Weakens correlation between sentence length and quality
- › Improves adequacy (but not completely)
- › Encoder-decoder with attention became SOTA in 2015

Application Specific Features

- › Length normalization
- › Coverage penalty

Length Normalization

- › Overcome bias towards shorter sentences
- › Normalize probabilities of candidates by their length $L(e)$

$$e^* = \operatorname{argmax}_e \frac{\Pr(e|f)}{L(e)}$$

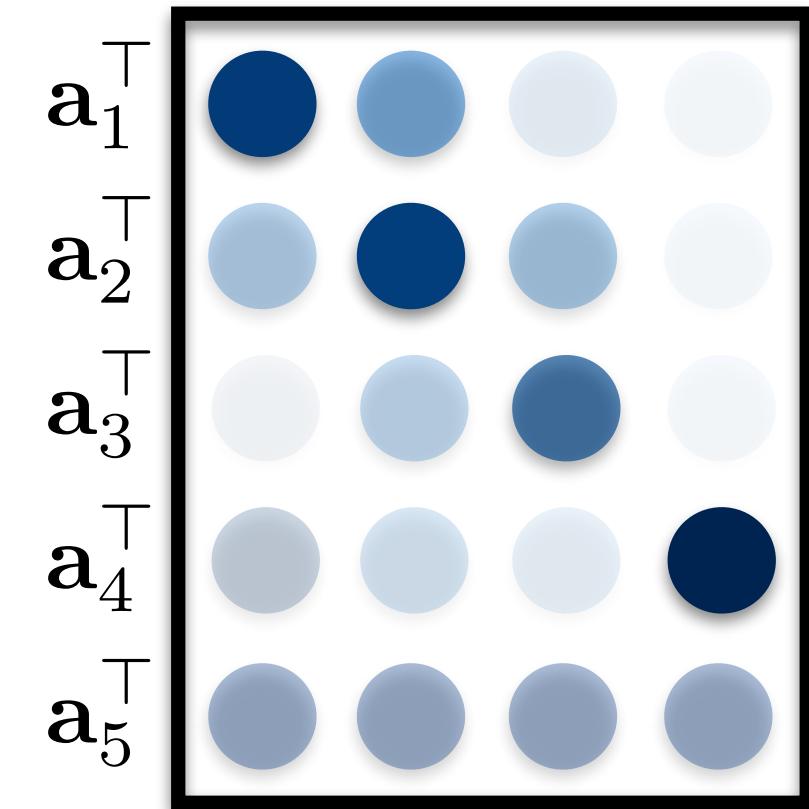
- › *Explicitly model length

$$e^* = \operatorname{argmax}_e \Pr(L(e)|f) \Pr(e|f)$$

Coverage Penalty

- › Make sure all source words were translated
- › Use $\sum_j \alpha_{i,j}$ as indicator of how much we translated the i -th source word

Attention history



$$e^* = \operatorname{argmax}_e \frac{\log \Pr(e|f)}{L(e)} + C(e, f)$$

How good is NMT?

- › Significantly more fluent: lemmatized and raw BLEU scores are closer

The animal didn't cross the street because it was too tired

Животное не **пересечь** улицу потому, что он слишком устал

Животное не **перешло** улицу потому, что оно было слишком уставшим

How good is NMT?

- › Rather amazingly good at translation from very distant languages

私は交通がとても重かったので、道路を渡ることはできませんでした。

I couldn't cross the road because the traffic was so heavy.

How good is NMT?

- › Rather amazingly good at translation from very distant languages

私は交通がとても重かったので、道路を渡ることはできませんでした。

I couldn't cross the road because the traffic was so heavy.

How good is NMT?

- › Rather amazingly good at translation from very distant languages

私は交通がとても重かったので、道路を渡ることはできませんでした。

I couldn't cross the road because the traffic was so heavy.

How good is NMT?

- › Rather amazingly good at translation from very distant languages

私は交通がとても重かったので、道路を渡ることはできませんでした。

I couldn't cross the road because the traffic was so heavy.

How good is NMT?

- › Rather amazingly good at translation from very distant languages

私は交通がとても重かったので、道路を渡ることはできませんでした。

I couldn't cross the road because the traffic was so heavy.

How good is NMT?

- › Rather amazingly good at translation from very distant languages

私は交通がとても重かったので、道路を渡ることはできませんでした。

I couldn't cross the road because the traffic was so heavy.

How good is NMT?

- › Rather amazingly good at translation from very distant languages

私は交通がとても重かったので、道路を渡ることはできませんでした。

I couldn't cross the road because the traffic was so heavy.

How good is NMT?

- › Rather amazingly good at translation from very distant languages

私は交通がとても重かったので、道路を渡ることはできませんでした。

I couldn't cross the road because the traffic was so heavy.

How good is NMT?

- › Rather amazingly good at translation from very distant languages

私は交通がとても重かったので、道路を渡ることとはできませんでした。

I couldn't cross the road because the traffic was so heavy.

How good is NMT?

- › Amazingly good at rephrasing

私は交通がとても重かったので、道路を渡ることはできませんでした。

I couldn't cross the road because the traffic was so heavy.

How good is NMT?

- › Amazingly good at rephrasing

私は交通がとても重かったので、道路を渡ることはできませんでした。

The traffic was so heavy that it was not possible to cross the road.

What Problems does NMT have?

- › Word sense disambiguation stands out



● АНГЛИЙСКИЙ

This store sells unsold stock at a discount.

45 / 10000

Этот магазин продает непроданные акции со скидкой.

RNN



Suggest an edit



• АНГЛИЙСКИЙ

This store sells unsold stock at a discount.

45 / 10000

РУССКИЙ



Этот магазин продает непроданные товары со скидкой.

Transformer

Перевести в Google Bing



• АНГЛИЙСКИЙ

According to some analysts the stock is selling at a large discount.

68 / 10000

По мнению некоторых аналитиков, акции продаются с большой скидкой.

RNN



Suggest an edit



• АНГЛИЙСКИЙ

According to some analysts the stock is selling at a large discount.

68 / 10000

РУССКИЙ



По мнению некоторых аналитиков акции продаются с большим дисконтом.

Transformer

[Перевести в Google Bing](#)



● АНГЛИЙСКИЙ

He made the stock from leftovers.

33 / 10000

Он сделал запас из остатков.

RNN



Suggest an edit



● АНГЛИЙСКИЙ

He made the stock from leftovers.

33 / 10000

РУССКИЙ



Он сделал запасы из остатков.

Transformer

Перевести в Google Bing



• АНГЛИЙСКИЙ

He made the stock from the leftover chicken.

44 / 10000

Он сделал запас из оставшейся курицы.

RNN



Suggest an edit



• АНГЛИЙСКИЙ

He made the stock from the leftover chicken.

44 / 10000

РУССКИЙ



Он сделал бульон из остатков курицы.

Transformer

Перевести в Google Bing