про внимание
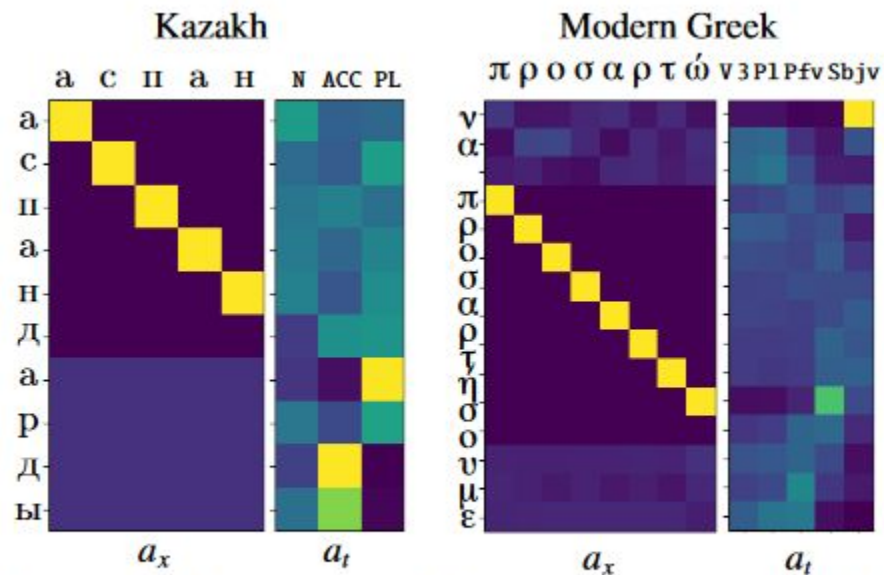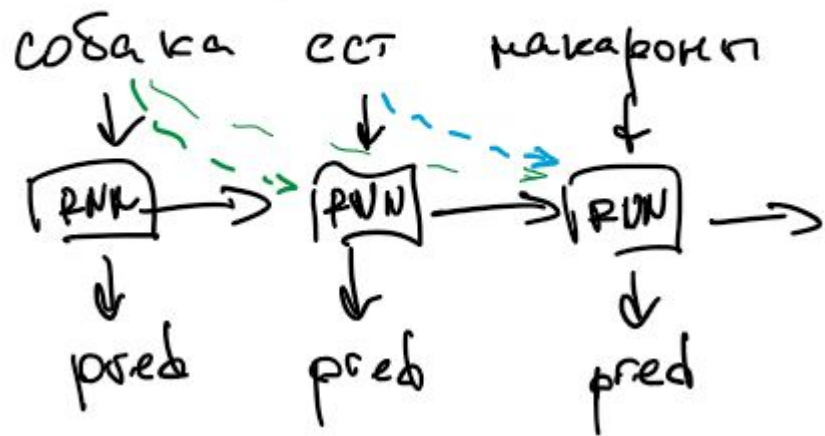
Figure 3: Attention visualization examples. The inflected form is generated from top to bottom.
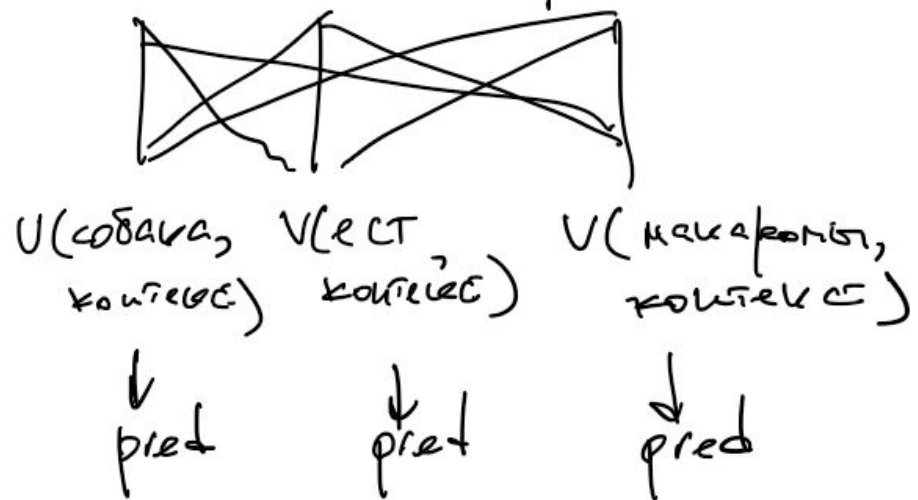
https://arxiv.org/pdf/1908.05838.pdf

Как работала RNN

собака   ест   макароны

```
собака          ест          макароны
  ↓               ↓               ↑
[RNN] ──→  →  [RNN] ──→  →  [RNN]  ──→
  ↓               ↓               ↓
pred           pred            pred
```

Как работала RNN

собака   ест   макароны

pred   pred   pred

Как работает АТN

собака ес макароны

U(собака,   V(ест   V(макароны,
  контекс)   контекс)   контекс)

pred   pret   pred

Как работает ATT N

собака ест макароши

$V(собака,\ контекс)$ $V(ест\ контекс)$ $V(макароши,\ контекс)$

$\downarrow$ pred $\downarrow$ pred $\downarrow$ pred

Как работает ATTN

собака ест макароны

V(собака, контекст)   V(ест, контекст)   V(макароны, контекст)

↓ pred   ↓ pred   ↓ pred

V(собака, контекст) ?

V(собака, контекст) =

$$\sum_{\substack{\text{слова} \\ \text{предл-я}}} \text{score}(\text{слово}_{\text{предл-я}}, \text{собака}) \cdot \text{emb}(\text{слово}_{\text{предл-я}})$$

$\underbrace{\qquad\qquad}$ насколько слово релевантно собаке.
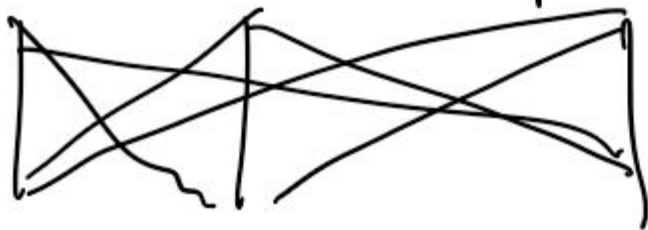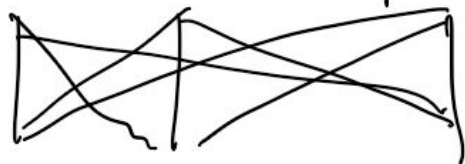
Name	Alignment score function	Citation
Content-base attention	$\text{score}(s_t, h_i) = \text{cosine}[s_t, h_i]$	Graves2014
Additive(*)	$\text{score}(s_t, h_i) = \mathbf{v}_a^\top \tanh(\mathbf{W}_a[s_t; h_i])$	Bahdanau2015
Location-Base	$\alpha_{t,i} = \text{softmax}(\mathbf{W}_a s_t)$ Note: This simplifies the softmax alignment to only depend on the target position.	Luong2015
General	$\text{score}(s_t, h_i) = s_t^\top \mathbf{W}_a h_i$ where $\mathbf{W}_a$ is a trainable weight matrix in the attention layer.	Luong2015
Dot-Product	$\text{score}(s_t, h_i) = s_t^\top h_i$	Luong2015
Scaled Dot-Product(^)	$\text{score}(s_t, h_i) = \frac{s_t^\top h_i}{\sqrt{n}}$ Note: very similar to the dot-product attention except for a scaling factor; where n is the dimension of the source hidden state.	Vaswani2017

$$V(\text{собака}, \text{контекст}) \quad ?$$

$$V(\text{собака}, \text{контекст}) =$$

$$\sum_{\substack{\text{слова} \\ \text{предл-я}}} \text{score}(\underset{\text{предл-я}}{\text{слово}}, \text{собака}) \cdot \text{emb}(\underset{\text{предл-я}}{\text{слово}})$$

насколько слово
релевантно собаке.

Figure 3: Attention visualization examples. The inflected form is generated from top to bottom.

$V(\text{собака}, \text{контекст})$?

$V(\text{собака}, \text{контекст}) =$

$$\sum_{\substack{\text{слова} \\ \text{предл-я}}} \text{score}(\text{слово}_{\text{предл-я}}, \text{собака}) \cdot \text{emb}(\text{слово}_{\text{предл-я}})$$

$\underbrace{}$

насколько слова
релевантно собаке.
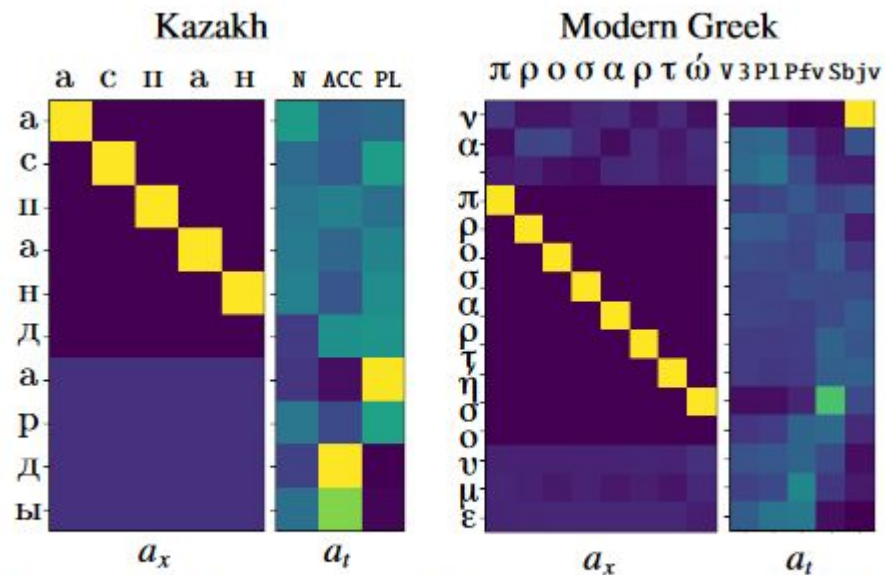
Пример гифок про то, как вычислять атеншн скор

https://towardsdatascience.com/illustrated-self-attention-2d627e33b20a

Две задачки

Задачка 1: https://gist.github.com/oserikov/13c706f7c8265b8d52d2f59e9a4d914d

Задачка 2: https://gist.github.com/oserikov/1b582a473aafdfe382487f0165f572ee