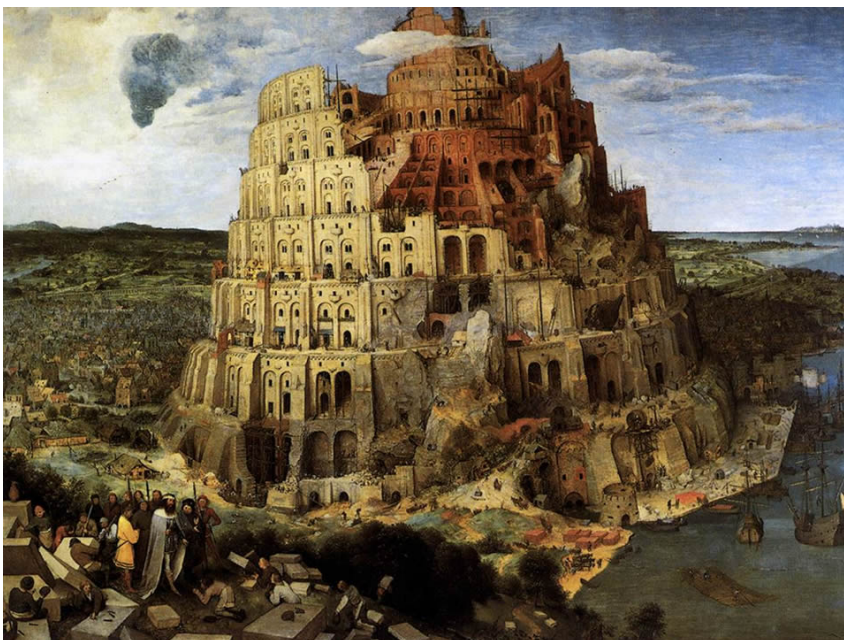


Pervasive Attention: 2D Convolutional Neural Networks for Sequence-to-Sequence Prediction

Подготовил: Самсонов Степан БПМИ151

Как понять друг друга?

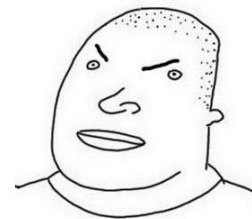


С чего всё
началось?



Штоooooooooo?

Как переводит машина?



- Прямой перевод
- Основанный на правилах
- Корпуса текстов
- Знания
- Гибриды

С чем имеем дело?

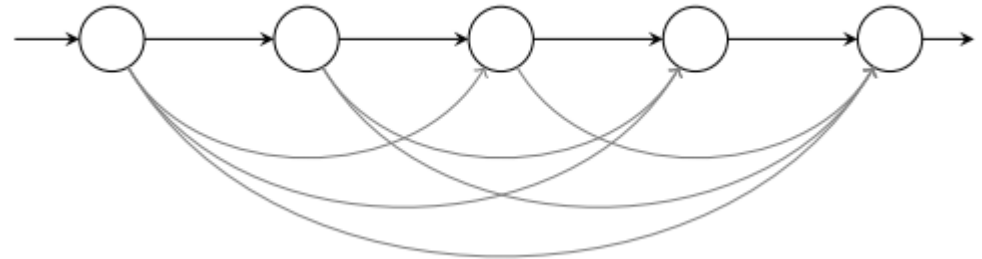
- CNN
- Обучаем на:
 - German-to-English (De-En)
 - English-to-German(En-De)

Подаём на вход

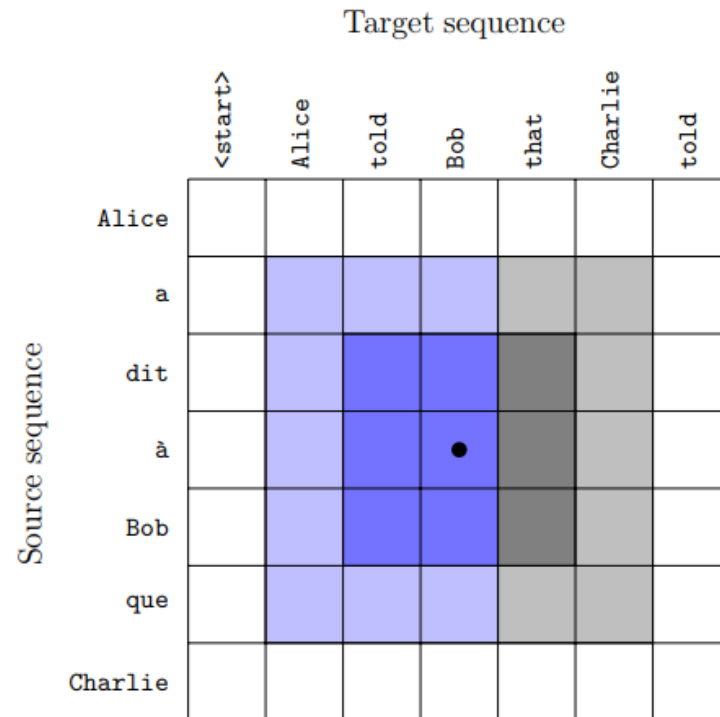
- Есть пара (s, t)
- Есть длины $|s|$ и $|t|$
- Берём токены
- Получаем тензор $R^{|t| \times |s| \times f_0}$
- $f_0 = d_s + d_t$

А что внутри?

- Dense net

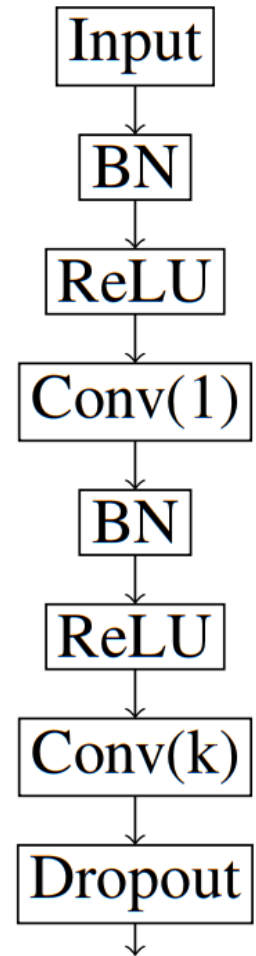


- CNN



Как сворачиваем?

- g – growth rate
- Свёртка 1:
 - Используем 1×1
 - Получаем $4g$ из $f_0 + (l - 1)g$
- Свёртка 2:
 - Используем $k \times \lceil \frac{k}{2} \rceil$
 - Получаем g



Что получаем в итоге?

- Тензор $H^{|t| \times |s| \times fL}$

- Max-pooling

$$H_{id}^{\text{pool}} = \max_{j \in \{1, \dots, |s|\}} H_{ijd}^L$$

- Average-pooling

$$H_{id}^{\text{pool}} = \frac{1}{\sqrt{|s|}} \sum_{j \in \{1, \dots, |s|\}} H_{ijd}^L$$

- Словарь V

- $E \in R^{|V| \times fL}$

- $p_i = \text{SoftMax}(EH_i^{\text{pool}})$

Implicit sentence alignment

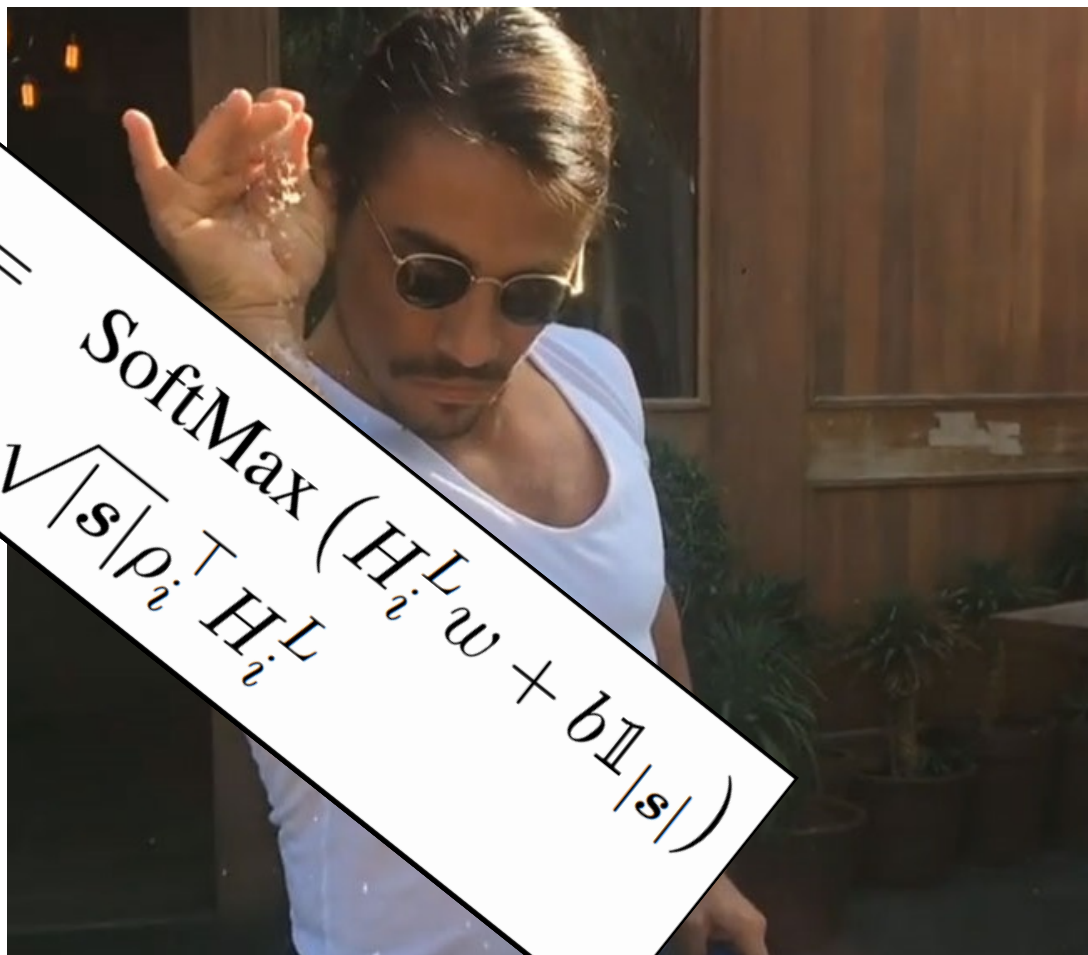
- Давайте взглянем на:

$$B_{ij} = \{d \in \{1, \dots, f_L\} \mid j = \arg \max(H_{ijd}^L)\}$$

- Это может оказаться полезным:

$$\alpha_{ij} = \sum_{d \in B_{ij}} E_{wd} H_{ijd}^L$$

Self attention



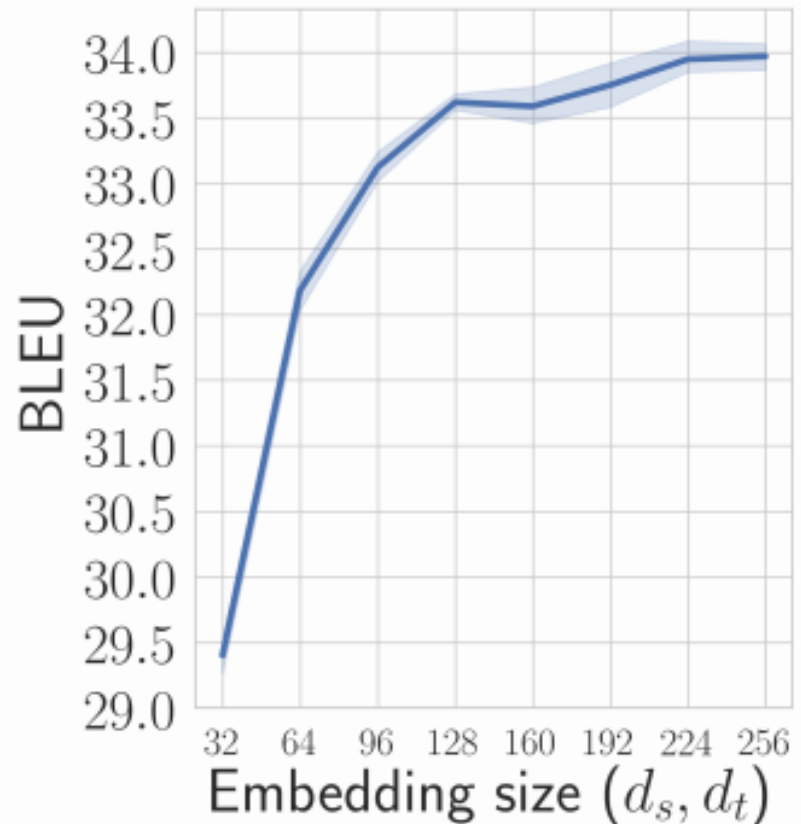
$$\begin{aligned} \rho_i &= \text{SoftMax} (H_i^L w + b \mathbb{I}_{|s|}) \\ H_i^{\text{att}} &= \frac{\rho_i}{\sqrt{|s|} \rho_i^T H_i^L} \end{aligned}$$

Можем обучать



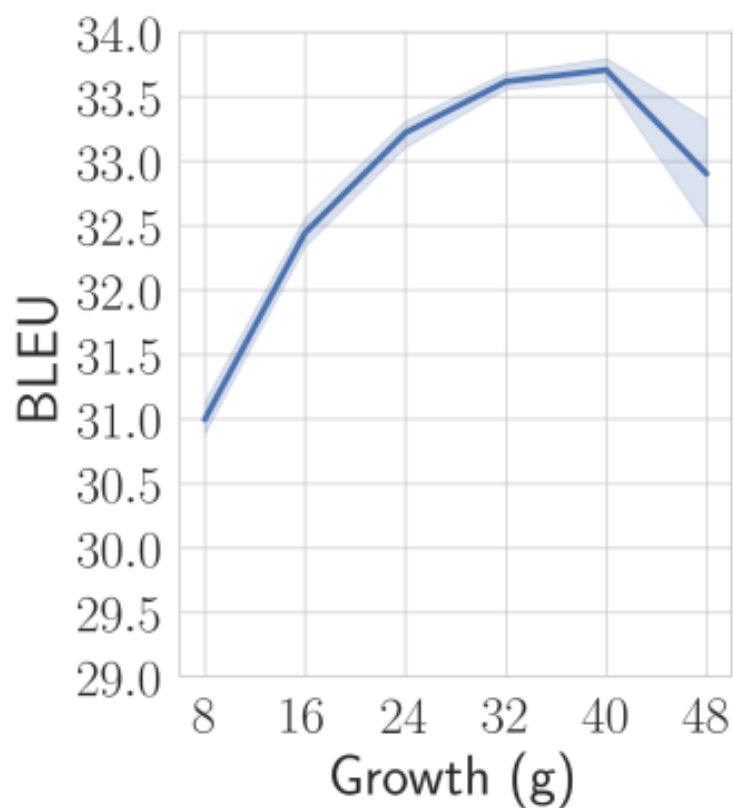
tooth	$\leftarrow *h_3dónts \rightarrow$	dēns, dentis \rightarrow dental
heart	$\leftarrow *k̑erd \rightarrow$	cor, cordis \rightarrow cordial
foot	$\leftarrow *pōds \rightarrow$	pēs, pedis \rightarrow pedestrian
star	$\leftarrow *H_2ster \rightarrow$	stēlla \rightarrow constellation
young	$\leftarrow *H_2yuH̑ \rightarrow$	juvenis \rightarrow juvenile
wind	$\leftarrow *H_2weH_1̑to \rightarrow$	ventus \rightarrow ventilation

(a) $L = 20, g = 32$

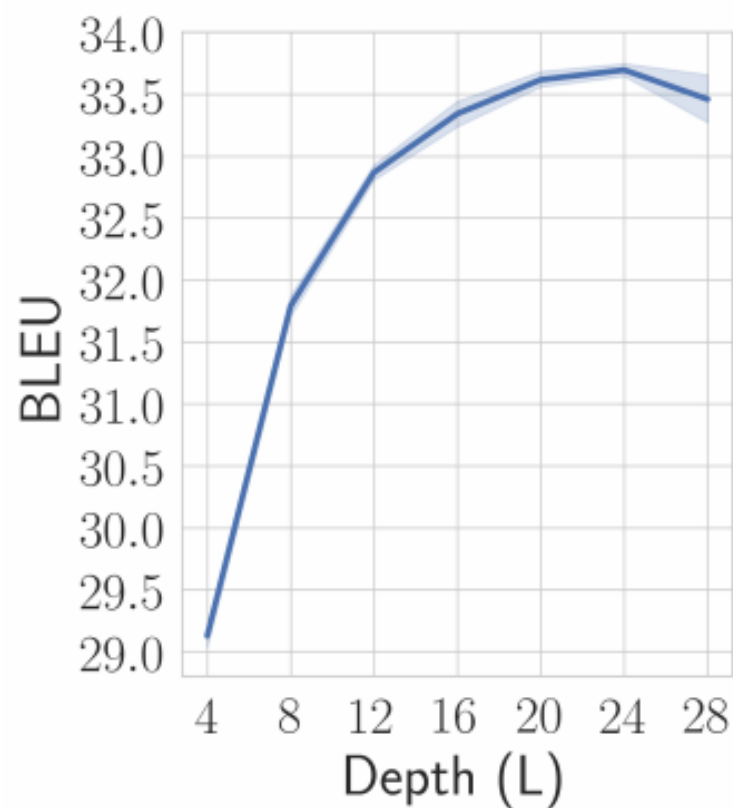


И ещё

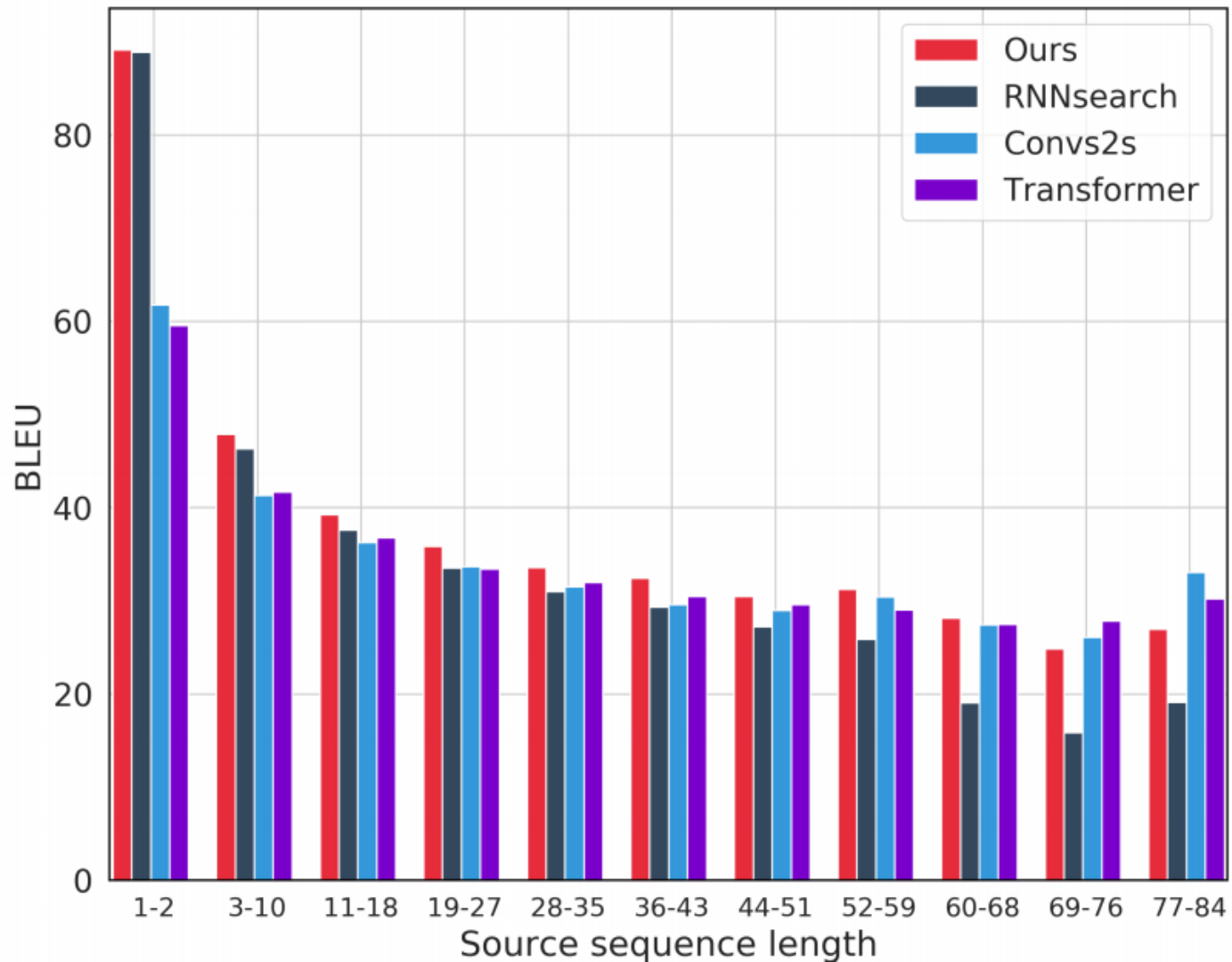
(b) $L = 20, d = 128$



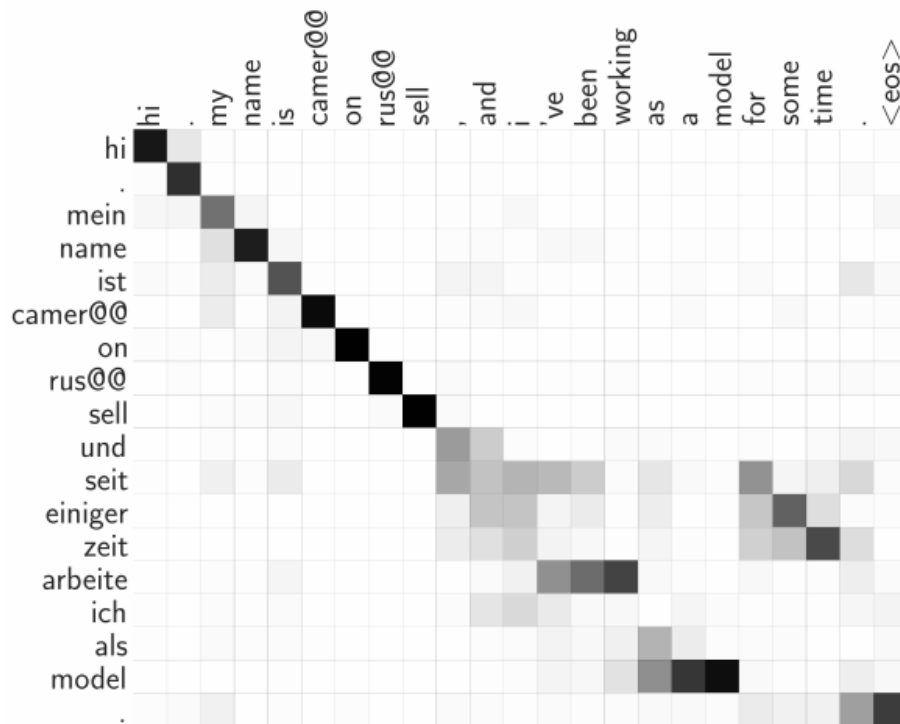
(c) $d = 128, g = 32$



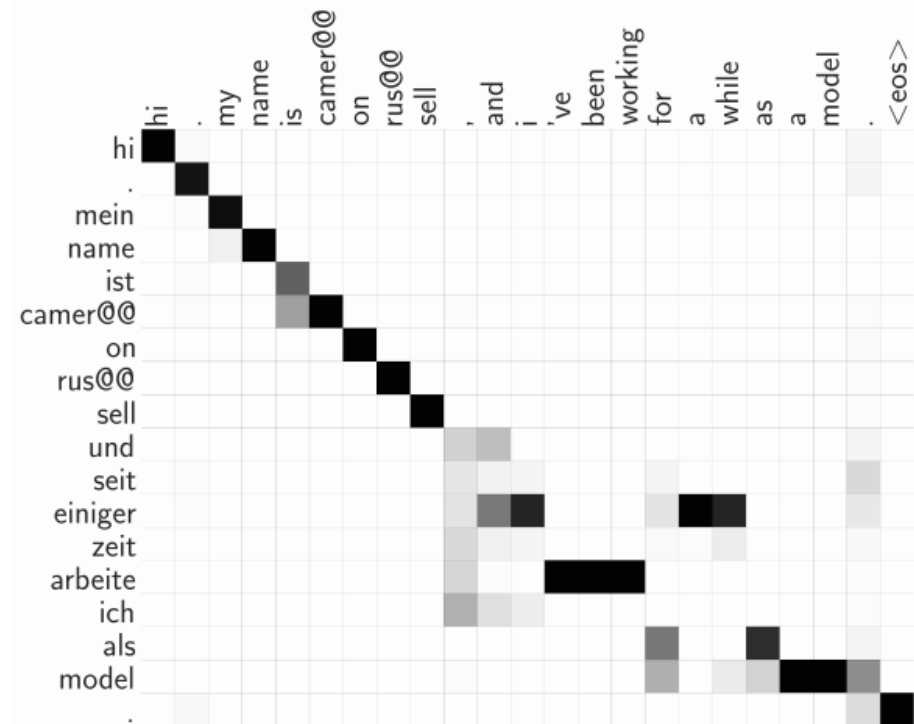
Сравним



Max-pooling VS self-attention

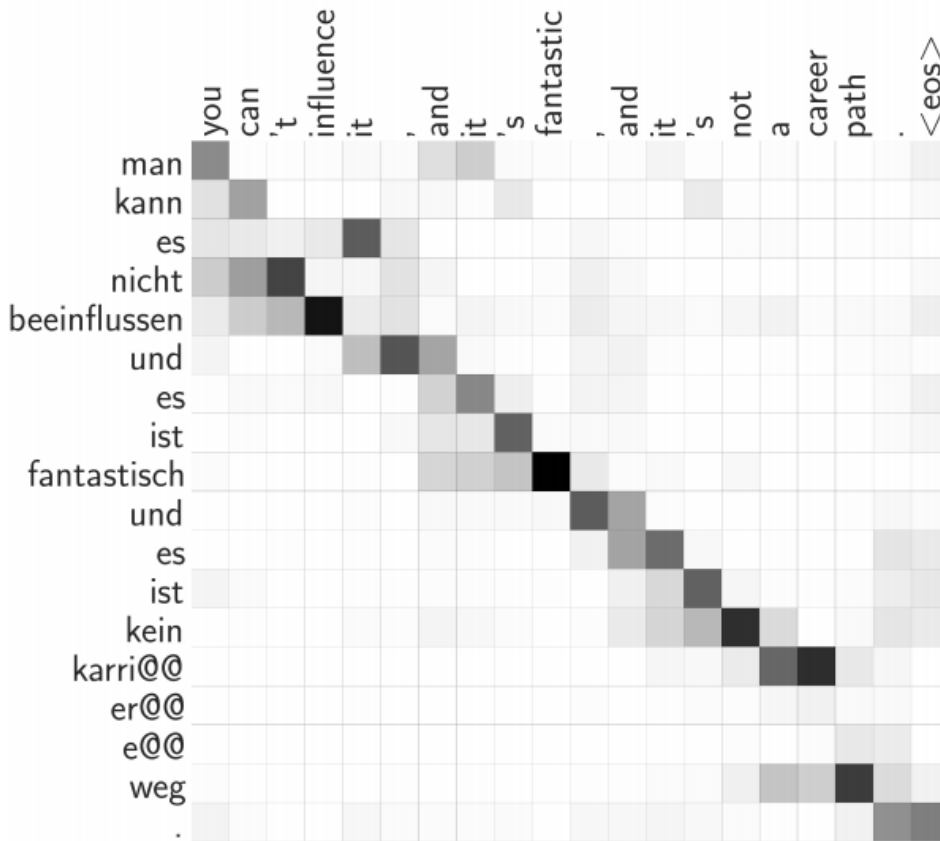


(a) Max-pooling

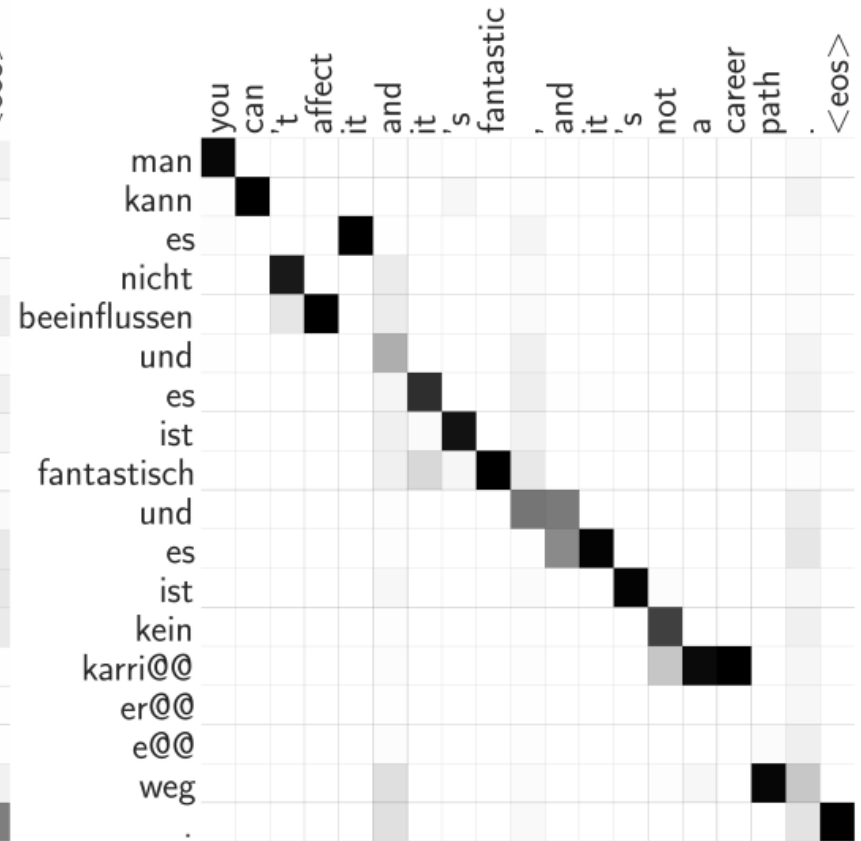


(b) Self-attention

Max-pooling VS self-attention



(c) Max-pooling



(d) Self-attention

Спасибо

Google Переводчик

Защищено | https://translate.google.ru/?hl=ru&tab=wT&authuser=0&google_abuse=GOOGLE_ABUSE_EXEMPTION%3DID%3De36618d69ddc5842:TM%3D1541594564:C%3...


Google

Переводчик

Отключить моментальный перевод

английский русский немецкий Определить язык

русский английский украинский Перевести



Beesechurger

Ты нормальный вообще?

Предложить исправление

Ты normal'nyy voobshche?

32/5000

О Переводчике Google Сообщества Мобильная версия

О Google Правила и принципы Справка Отправить отзыв

23:26 07.11.2018