

- Интеграл слов с одним значением в разных языках почти не встречаются в одинаковом контексте
- Интеграл слов с одним значением почти не встречаются

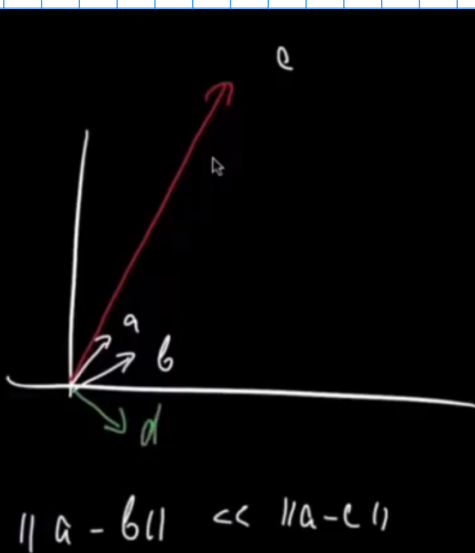
## Word embeddings in different languages

- Word embeddings are quite similar for different languages
- Assume there  $n = 5000$  word-translation pairs  $\{x_i, y_i\}_{i \in \{1, n\}}$
- Learn linear mapping between the source and target spaces

$$W^* = \operatorname{argmin}_{W \in M_d(\mathbb{R})} \|WX - Y\|_F$$

- The translation of source word is  $t = \operatorname{argmax}_t \cos(Wx_s, y_t)$ .

- ① получаем по отдельности эмбединги для 2-х языков.
- ② т.к. эмбед. будут почти, мы можем попробовать обучить матрицу перевода (поворот, растяжение, сжатие) для некоторого к-ва слов и считать, что она подойдет и для остальных +

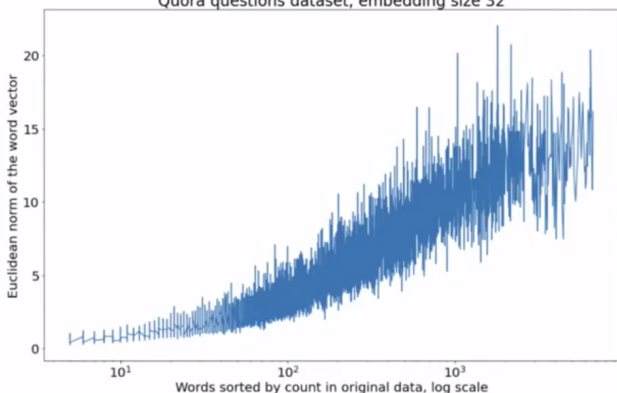


почему cos fit, а не евклидово?

мысли могут иметь схожую тематику, но быть разного размера

⇓  
иметь разные нормы

How word frequency affects the embedding vector norm  
Quora questions dataset, embedding size 32



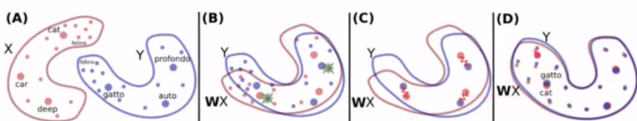
- Чем выше частота встречаемости слова, тем больше его норма

### Vector norms for words with no specific context

word	count	vector norm
overheat	11	0.81233
enormous	12	0.807057
dog	1212	11.2591
cat	1545	10.3738
laptop	1906	14.5192
phone	4124	15.7901
a	155726	11.4656
the	252068	8.47355

Пример перевода

## Word embeddings in different languages



Comment: mapping between two languages can be done completely in unsupervised manner with GANs.

We will meet later.

## Word embeddings in different languages

- Word embeddings are quite similar for different languages
- Assume there  $n = 5000$  word-translation pairs  $\{x_i, y_i\}_{i \in \{1, n\}}$
- Learn linear mapping between the source and target spaces  
**enforcing an orthogonality constraint on  $W$ :**

$$W^* = \underset{W \in O_d(\mathbb{R})}{\operatorname{argmin}} \|WX - Y\|_F = UV^T, \text{ with } U\Sigma V^T = \operatorname{SVD}(YX^T).$$

- The translation of source word is  $t = \operatorname{argmax}_t \cos(Wx_s, y_t)$ .

Sources: [Normalized word embedding and orthogonal transform for bilingual word translation, NAACL 2015](#) <sup>8</sup>  
[Word translation without parallel data, ICLR 2018](#)

• улучшение через нормализацию ограничения на матрицу  $W$  - пусть она будет ортогональной.

- Кросс-язык - в RNN-х не коммутируются введ./вывед. эмбедингами из отдельного Word2Vec