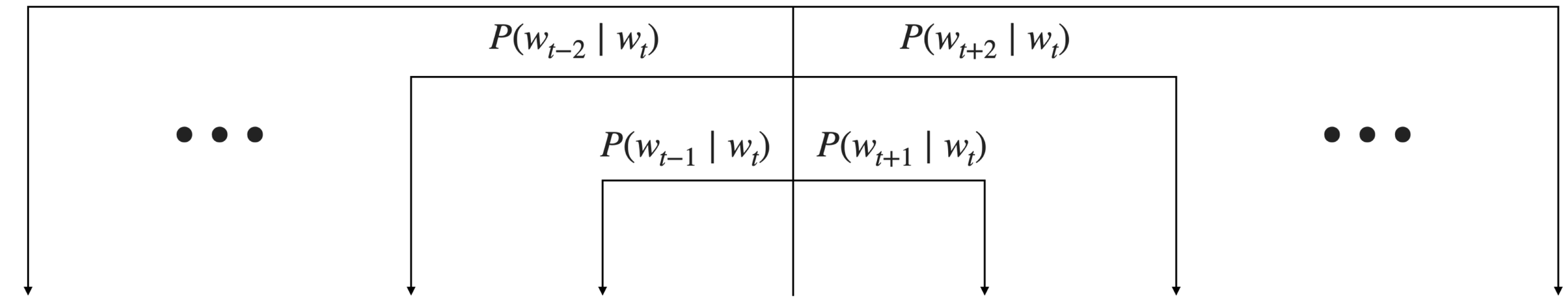
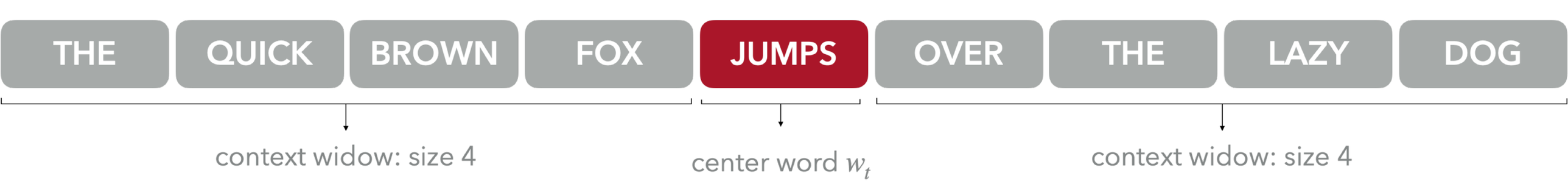
Word2vec





Probability of context given word:

$$P(w_{t-m}, \ldots, w_{t+m} \mid w_t) = \prod_{-m \leq j \leq m, j \neq 0} P(w_{t+j} \mid w_t; \theta)$$

Naïve Bayes assumption

model parameters

Loss function

- Likelihood (of entire document): -- maximize --

$$\prod_{t=1}^{T} \prod_{\substack{-m \le j \le m \\ j \neq 0}} P(w_{t+j} \mid w_t; \theta)$$

- Loss: negative log likelihood -- minimize --

$$-\frac{1}{T} \sum_{t=1}^{T} \sum_{\substack{-m \le j \le m \\ j \neq 0}} \log P(w_{t+j} \mid w_t; \theta)$$

Word-to-word probability

context word center word

$$P(o \mid c; \theta) = \frac{\exp(u_o^\top v_c)}{\sum_{w \in V} \exp(u_w^\top v_c)}$$

Softmax over all possible context words

Each word *w* gets 2 vectors:

- v_w when it is a center word
- u_w when it is a context word

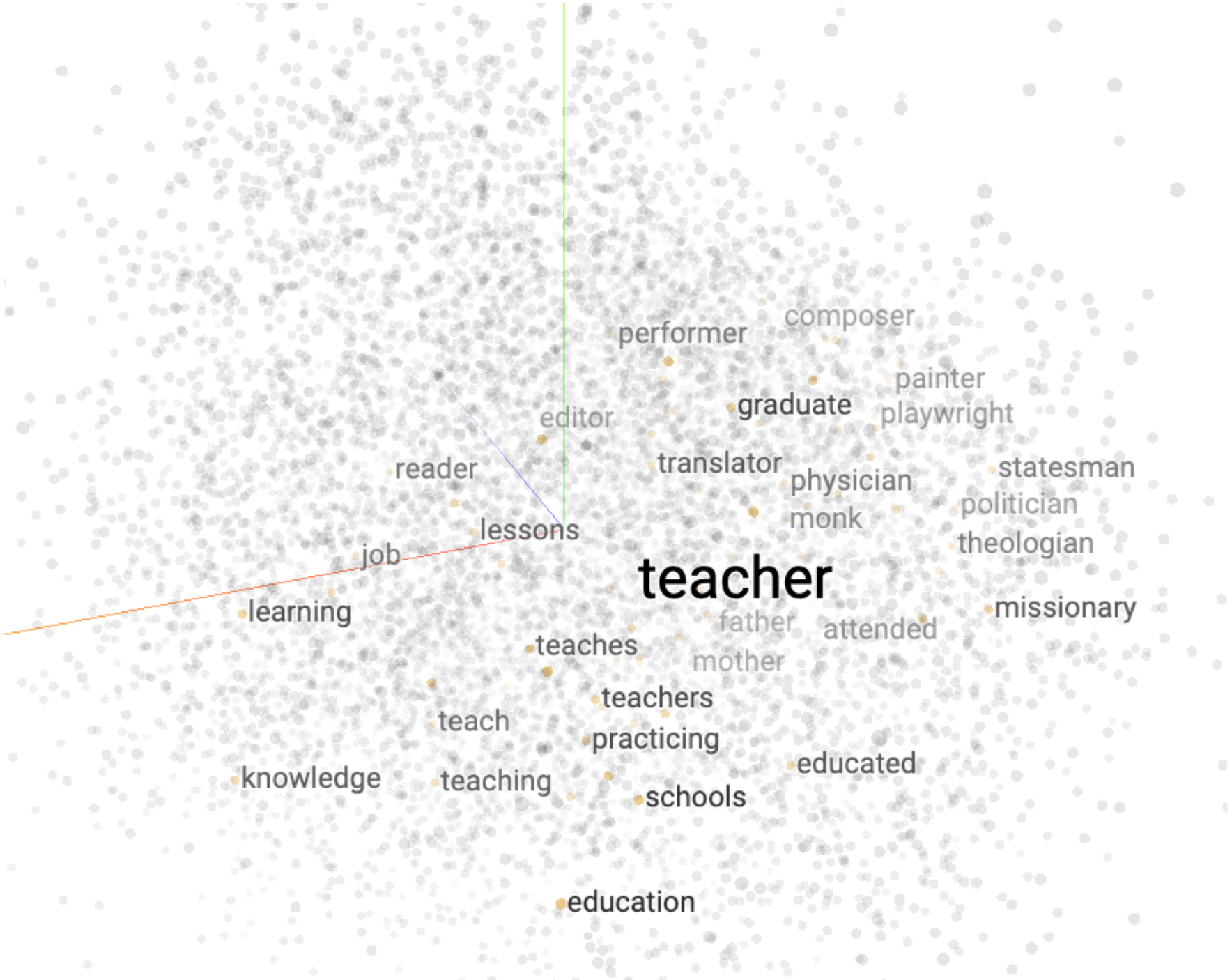
θ = all vectors v_w, u_w

Modeling document likelihood

$$\text{Loss}(U, V) = -\frac{1}{T}\sum_{t=1}^{T}\sum_{-m\leq j\leq m} \log P(w_{t+j} \mid w_t; \theta)$$

$$= -\frac{1}{T}\sum_{t=1}^{T}\sum_{-m\leq j\leq m} \log \frac{\exp(u_o^\top v_c)}{\sum_{w\in V}\exp(u_w^\top v_c)}$$

Exploring Word2vec embeddings



https://projector.tensorflow.org/

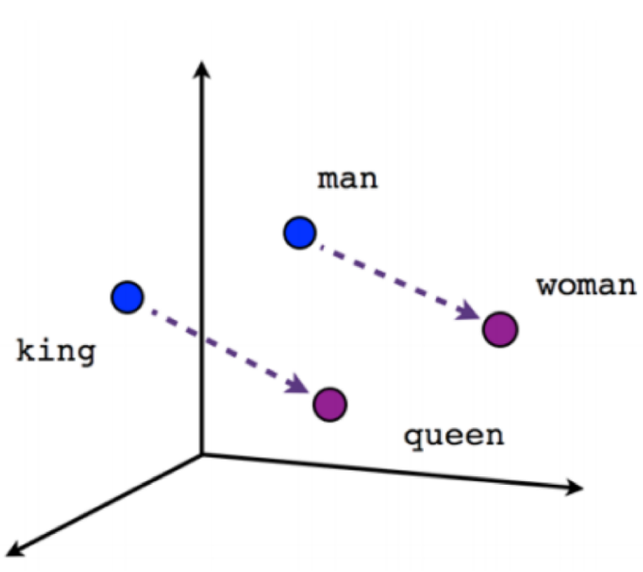
Exploring Word2vec embeddings

- Nearest neighbor (most similar) words:

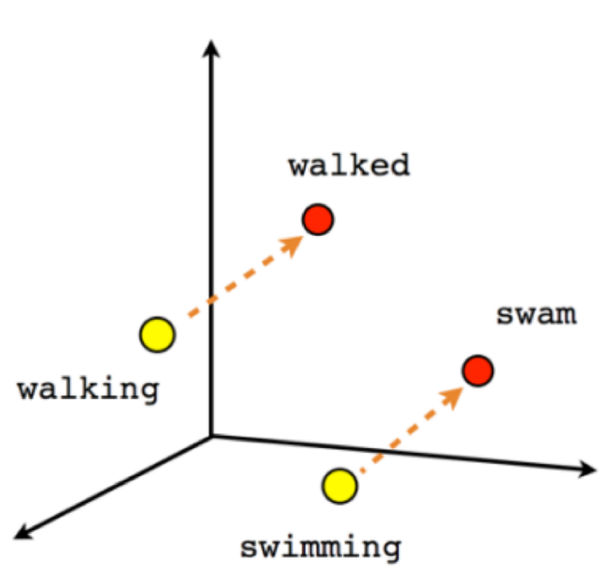
bottle		**teacher**		**price**	
glass	0.612	student	0.484	prices	0.314
klein	0.639	school	0.517	cost	0.471
plastic	0.642	scholar	0.550	market	0.485
juice	0.671	teaching	0.562	demand	0.507
liquid	0.674	theologian	0.578	consumer	0.535
ball	0.676	lawyer	0.587	wage	0.542
boiling	0.678	teachers	0.592	markets	0.551
brand	0.680	taught	0.600	wages	0.551
beer	0.681	translator	0.602	commodity	0.560
drunk	0.688	students	0.605	costs	0.564

Linear algebra with words

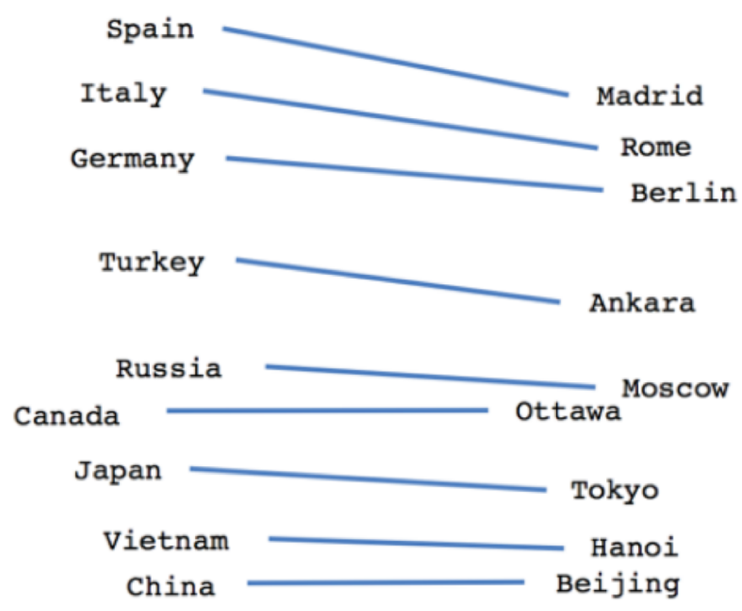
$$v_{\text{king}} + v_{\text{woman}} - v_{\text{man}} \approx v_{\text{queen}}$$



Male-Female



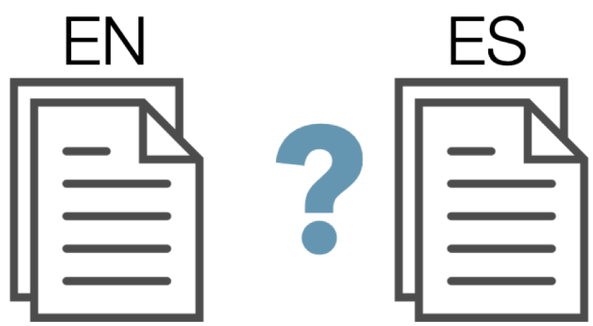
Verb tense



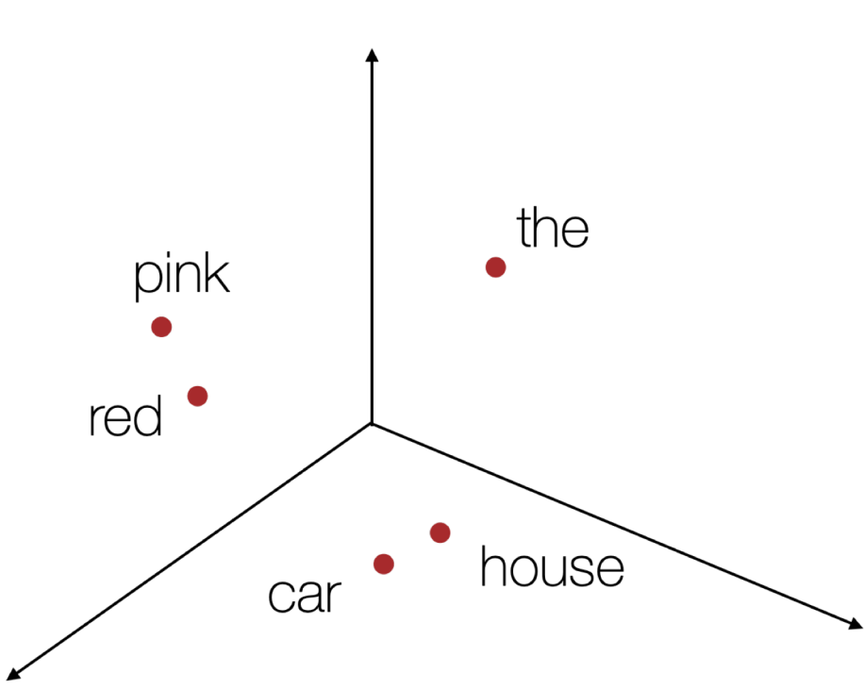
Country-Capital

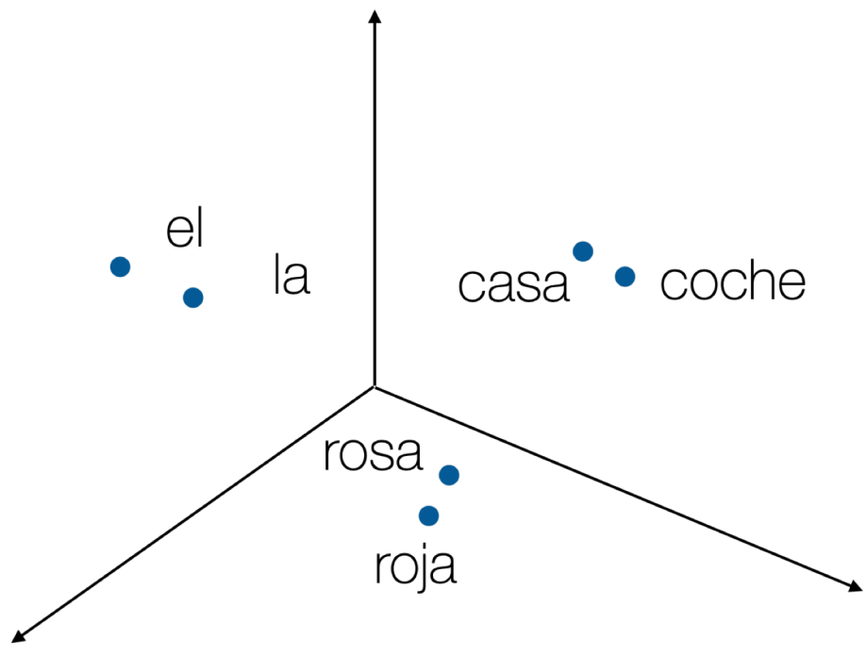
Source: tensorflow.org/tutorials

Unsupervised word translation



Idea: Similar word co-occurrence patterns
for a word and its translation
=> **Similar geometry across languages** *(Mikolov et al 2013)*





Caution: amplifying data biases

Man is to Computer Programmer as Woman is to Homemaker?
Debiasing Word Embeddings

Tolga Bolukbasi[1], Kai-Wei Chang[2], James Zou[2], Venkatesh Saligrama[1,2], Adam Kalai[2]
[1]Boston University, 8 Saint Mary's Street, Boston, MA
[2]Microsoft Research New England, 1 Memorial Drive, Cambridge, MA
tolgab@bu.edu, kw@kwchang.net, jamesyzou@gmail.com, srv@bu.edu, adam.kalai@microsoft.com

$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{computer programmer}} - \overrightarrow{\text{homemaker}}.$$

Extreme *she* occupations

1. homemaker	2. nurse	3. receptionist
4. librarian	5. socialite	6. hairdresser
7. nanny	8. bookkeeper	9. stylist
10. housekeeper	11. interior designer	12. guidance counselor

Extreme *he* occupations

1. maestro	2. skipper	3. protege
4. philosopher	5. captain	6. architect
7. financier	8. warrior	9. broadcaster
10. magician	11. figher pilot	12. boss

Caution: amplifying data biases