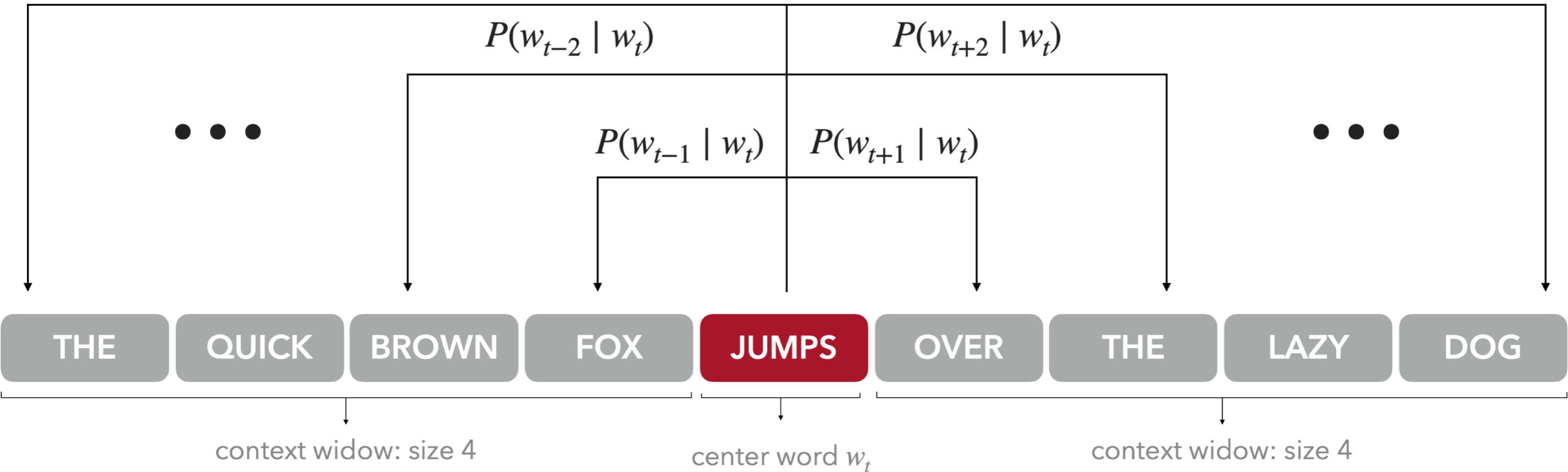


Word2vec



Probability of
context
given word:

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

↑
Naïve Bayes assumption

↑
model parameters

Loss function

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

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

- Loss: negative log likelihood -- minimize --

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

Word-to-word probability

context word center word

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

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

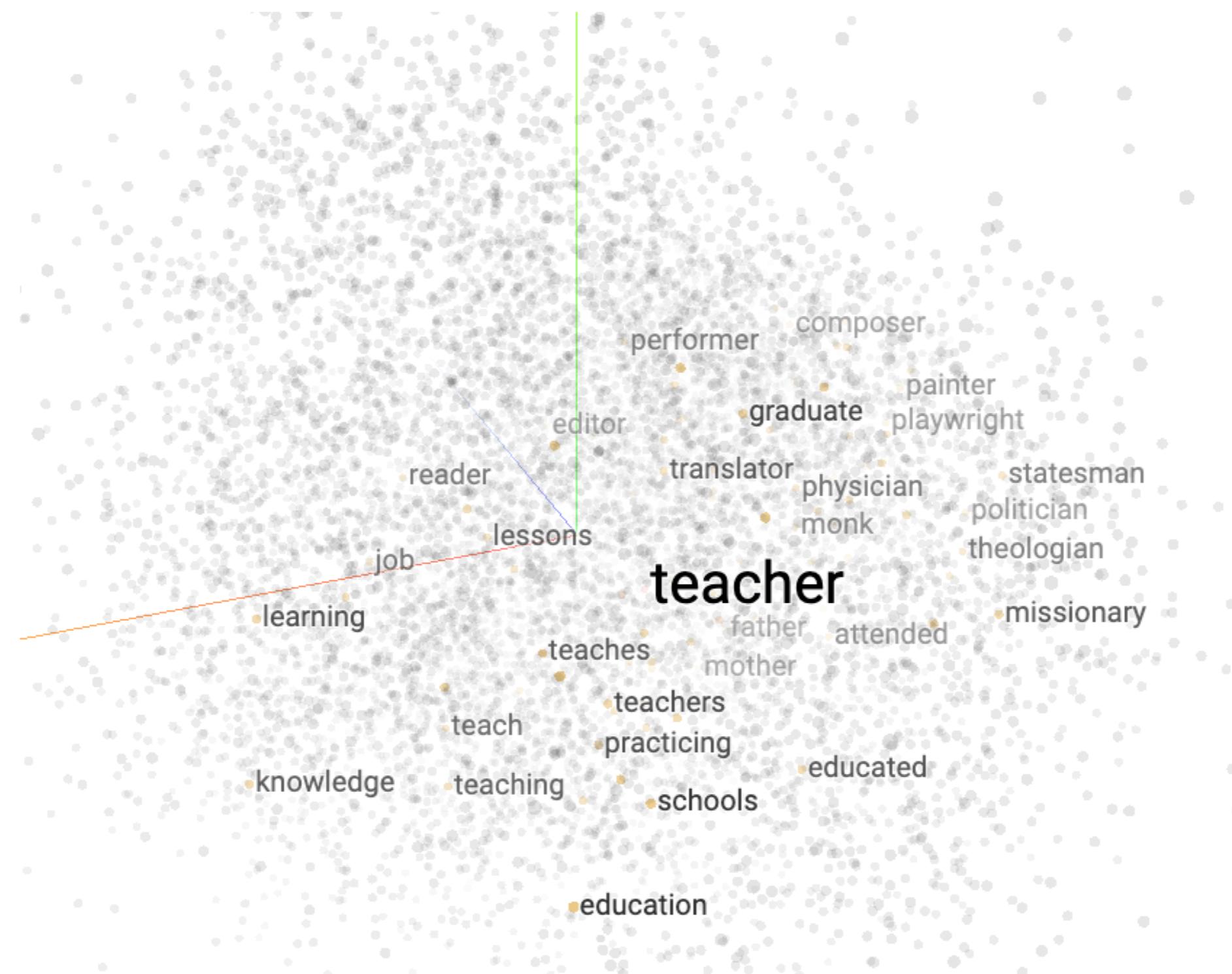
θ = all vectors v_w, u_w

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



Exploring Word2vec embeddings

- Nearest neighbor (most similar) words:

bottle

glass	0.612
klein	0.639
plastic	0.642
juice	0.671
liquid	0.674
ball	0.676
boiling	0.678
brand	0.680
beer	0.681
drunk	0.688

teacher

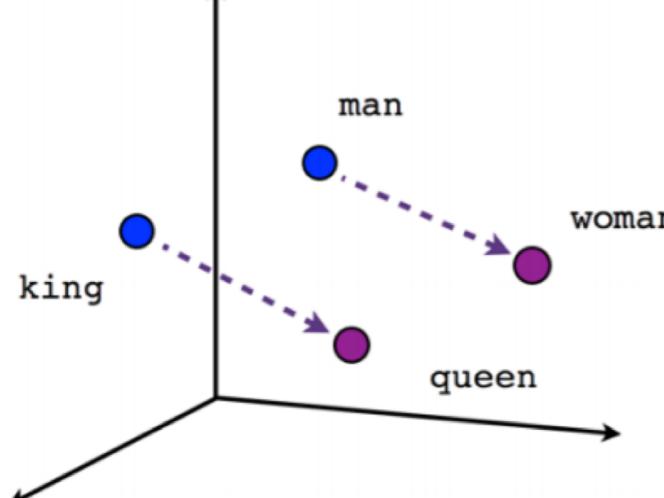
student	0.484
school	0.517
scholar	0.550
teaching	0.562
theologian	0.578
lawyer	0.587
teachers	0.592
taught	0.600
translator	0.602
students	0.605

price

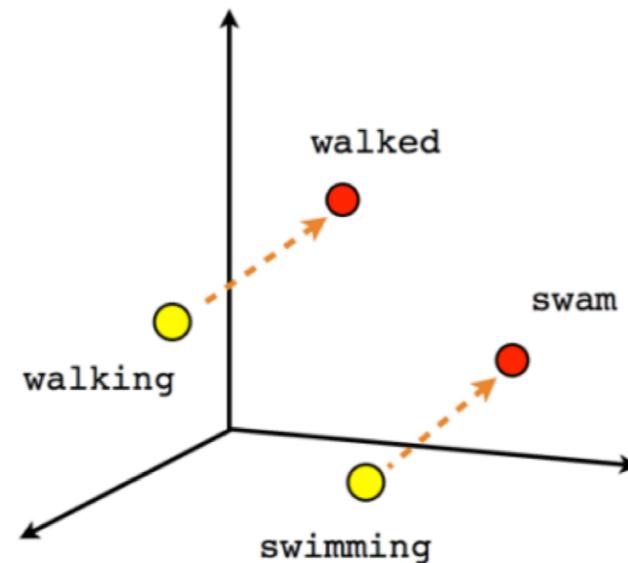
prices	0.314
cost	0.471
market	0.485
demand	0.507
consumer	0.535
wage	0.542
markets	0.551
wages	0.551
commodity	0.560
costs	0.564

Linear algebra with words

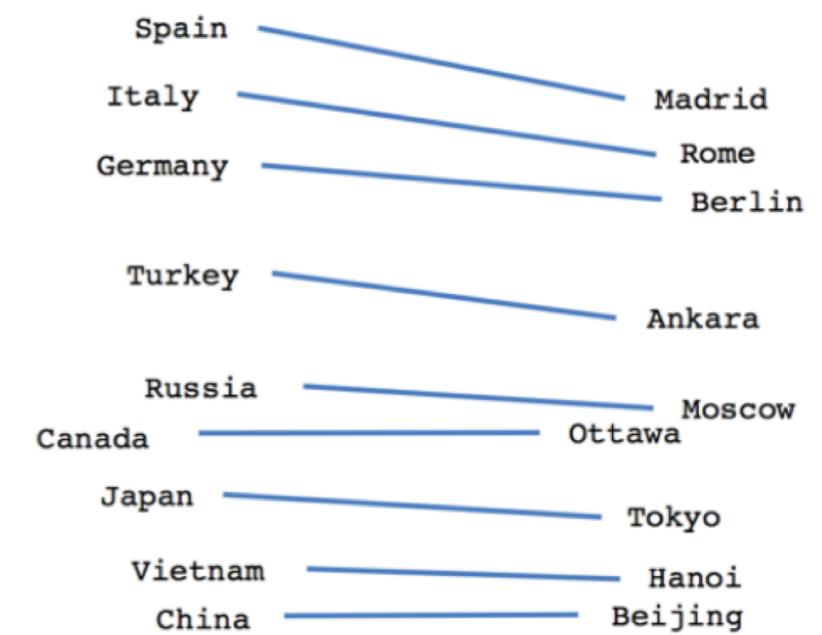
$$v_{\text{king}} + v_{\text{woman}} - v_{\text{man}} \approx v_{\text{queen}}$$



Male-Female



Verb tense



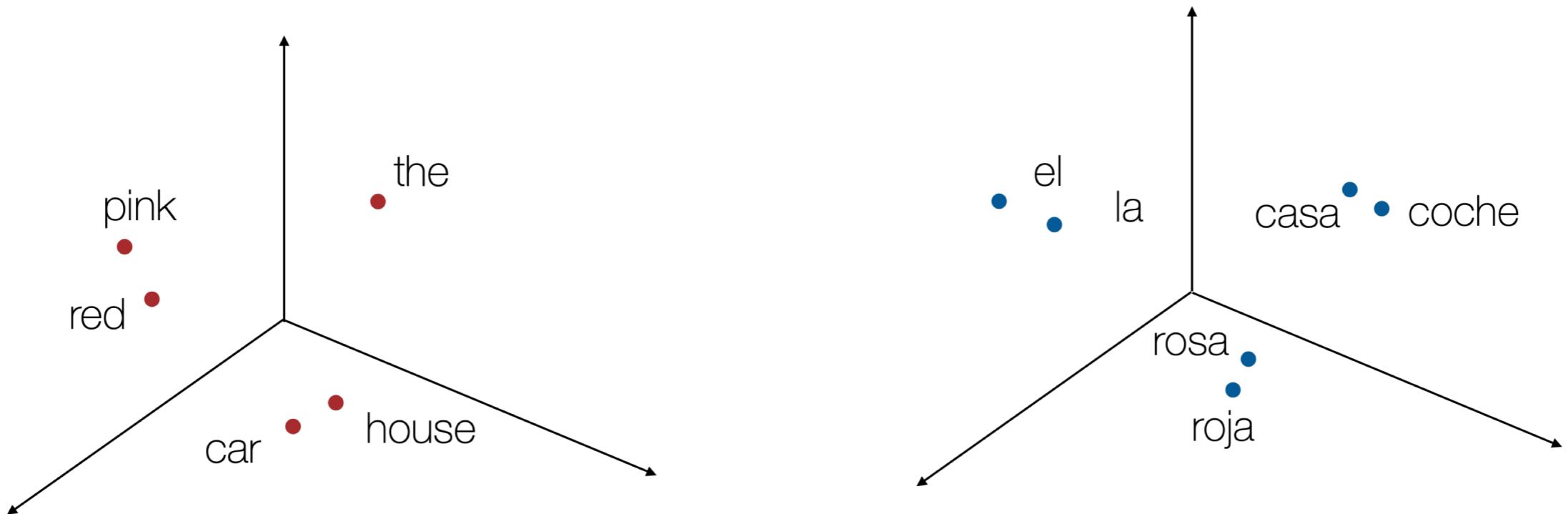
Country-Capital

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 Bolukbasi¹, Kai-Wei Chang², James Zou², Venkatesh Saligrama^{1,2}, Adam Kalai²

¹Boston University, 8 Saint Mary's Street, Boston, MA

²Microsoft 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

The screenshot shows a translation interface with two panels. The left panel has source text in Hungarian and target text in English. The right panel has source text in English and target text in Spanish. The English input is identical in both panels.

Left Panel (Hungarian to English):

- Bengali English Hungarian Detect language ▾
- English Spanish Hungarian ▾
- Translate

Source (Hungarian):

- ő egy ápoló.
- ő egy tudós.
- ő egy mérnök.
- ő egy pék.
- ő egy tanár.
- ő egy esküvői szervező.
- ő egy vezérigazgatója.

Target (English):

- she's a nurse.
- he is a scientist.
- he is an engineer.
- she's a baker.
- he is a teacher.
- She is a wedding organizer.
- he's a CEO.

Right Panel (English to Spanish):

- Bengali English Hungarian Detect language ▾
- English Spanish Hungarian ▾
- Translate

Source (English):

- she's a nurse.
- he is a scientist.
- he is an engineer.
- she's a baker.
- he is a teacher.
- She is a wedding organizer.
- he's a CEO.

Target (Spanish):

- she's a nurse.
- he is a scientist.
- he is an engineer.
- she's a baker.
- he is a teacher.
- She is a wedding organizer.
- he's a CEO.