

Distributional semantics: bee and honey vs. bee an bumblebee

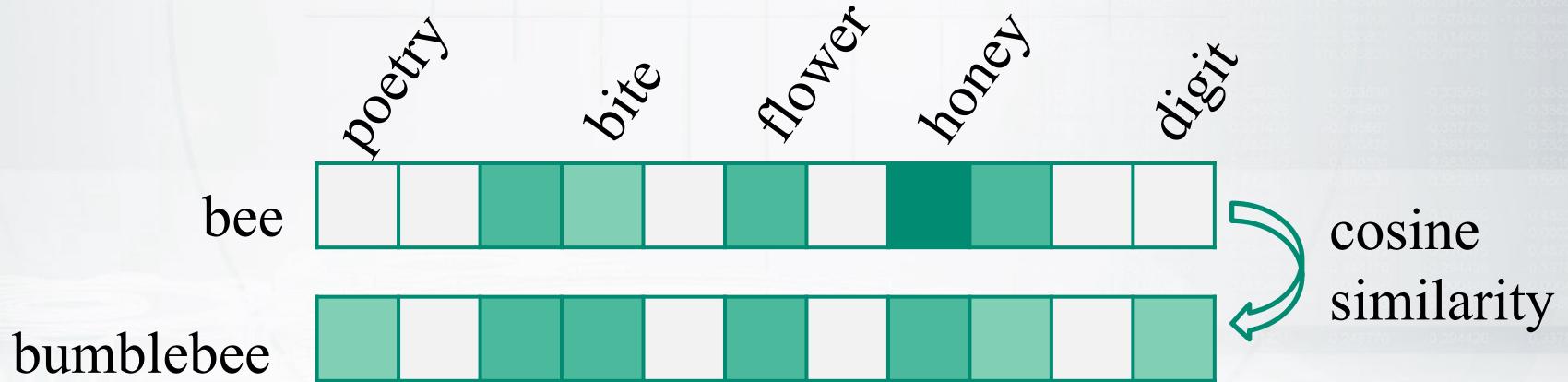
Word similarities

- First order co-occurrences

syntagmatic associates / relatedness (bee and honey)

- Second order co-occurrences

paradigmatic parallels / similarity (bee and bumblebee)



Schutze, H., & Pedersen, J. (1993). A vector model for syntagmatic and paradigmatic relatedness. In Making Sense of Words: Proceedings of the Conference, pp. 104-113, Oxford, England.

Distributional hypothesis

“*You shall know a word by the company it keeps.*”

— Firth, 1957.

- Use a sliding window of a fixed size
- Compute word co-occurrences n_{uv}

Distributional hypothesis

“*You shall know a word by the company it keeps.*”

— Firth, 1957.

- Use a sliding window of a fixed size
- Compute word co-occurrences n_{uv}
- **Better:** Pointwise Mutual Information:

$$PMI = \log \frac{p(u, v)}{p(u)p(v)} = \log \frac{n_{uv}n}{n_u n_v}$$

Distributional hypothesis

“*You shall know a word by the company it keeps.*”

— Firth, 1957.

- Use a sliding window of a fixed size
- Compute word co-occurrences n_{uv}
- **Better:** Pointwise Mutual Information:

$$PMI = \log \frac{p(u, v)}{p(u)p(v)} = \log \frac{n_{uv}n}{n_u n_v}$$

- **Even better:** positive Pointwise Mutual Information:

$$pPMI = \max(0, PMI)$$

Any problems here?

- First order co-occurrences

syntagmatic associates / relatedness (bee and honey)

- Second order co-occurrences

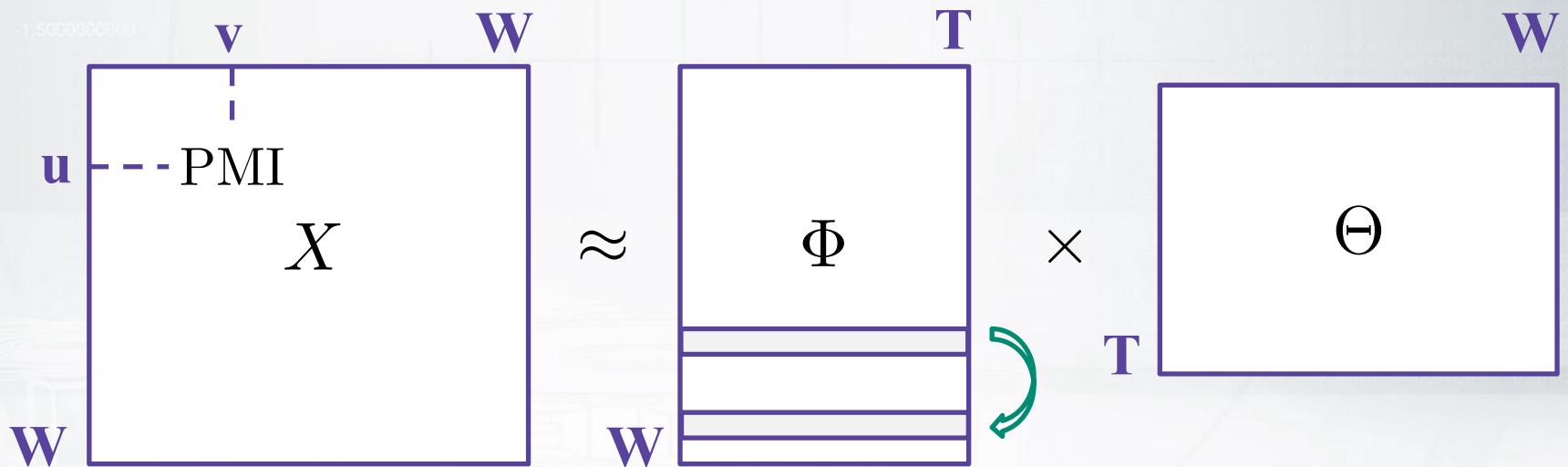
paradigmatic parallels / similarity (bee and bumblebee)



Schütze, H., & Pedersen, J. (1993). A vector model for syntagmatic and paradigmatic relatedness. In Making Sense of Words: Proceedings of the Conference, pp. 104-113, Oxford, England.

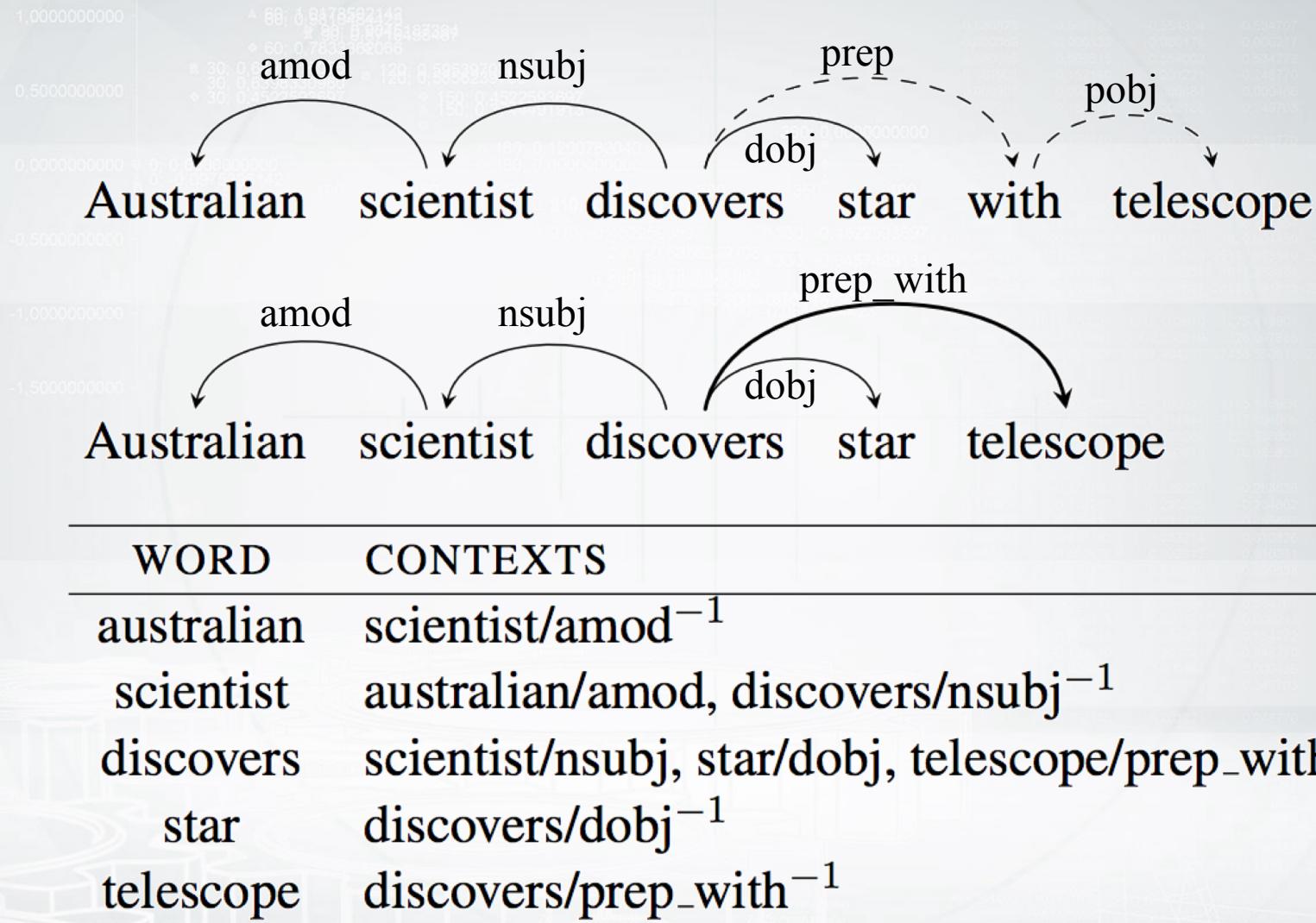
Vector Space Models of Semantics

- **Input:** word-word co-occurrences (counts, PMI, ...)
- **Method:** dimensionality reduction (SVD, ...)
- **Output:** similarity between vector representations of words



Turnay, P.D., Pantel, P.: from Frequency to Meaning: Vector Space Models of Semantics, 2010.

What is a context?



What is a context?

- \mathbf{C} is a vocabulary of contexts, e.g. word/dependency
- But usually contexts are words form a sliding window
- Then $\mathbf{W} = \mathbf{C}$ and \mathbf{X} is a symmetric matrix

