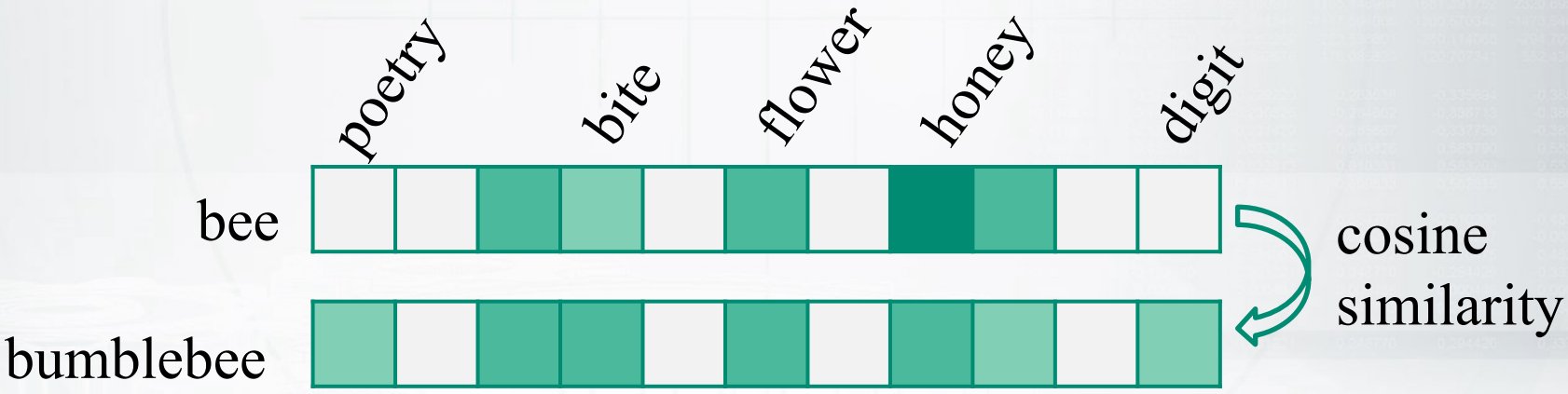
Distributional semantics: bee and honey vs. bee an bumblebee

Word similarities

- First order co-occurrences

  syntagmatic associates / relatedness (bee and honey)
- Second order co-occurences

  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:

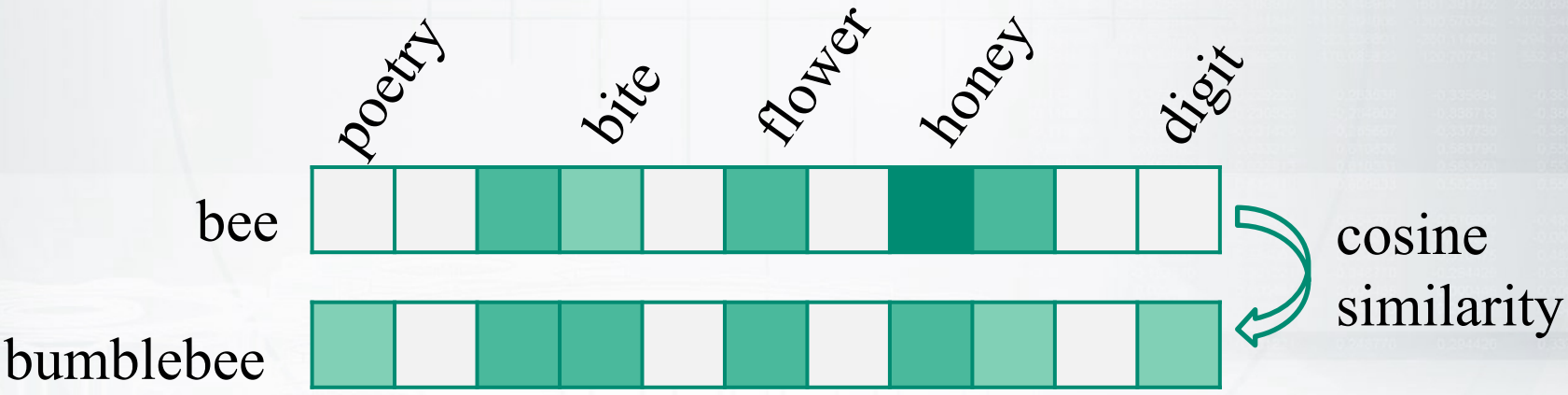
$$pPMI = \max(0, PMI)$$

Any problems here?

- First order co-occurrences

  syntagmatic associates / relatedness (bee and honey)
- Second order co-occurences

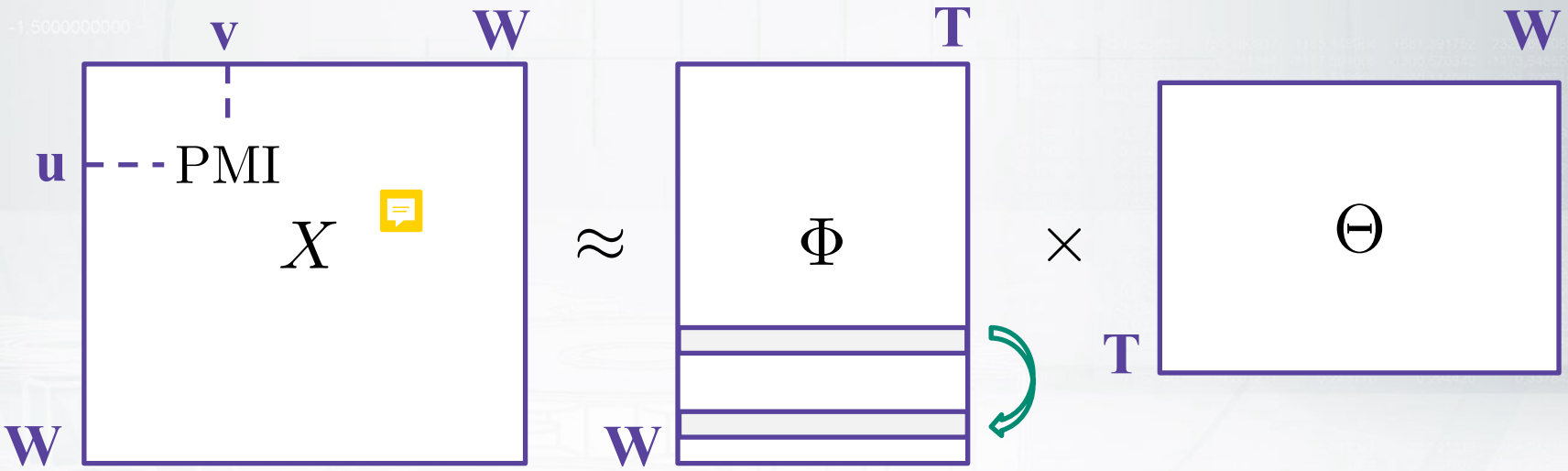
  paradigmatic parallels / similarity (bee and bumblebee)



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.

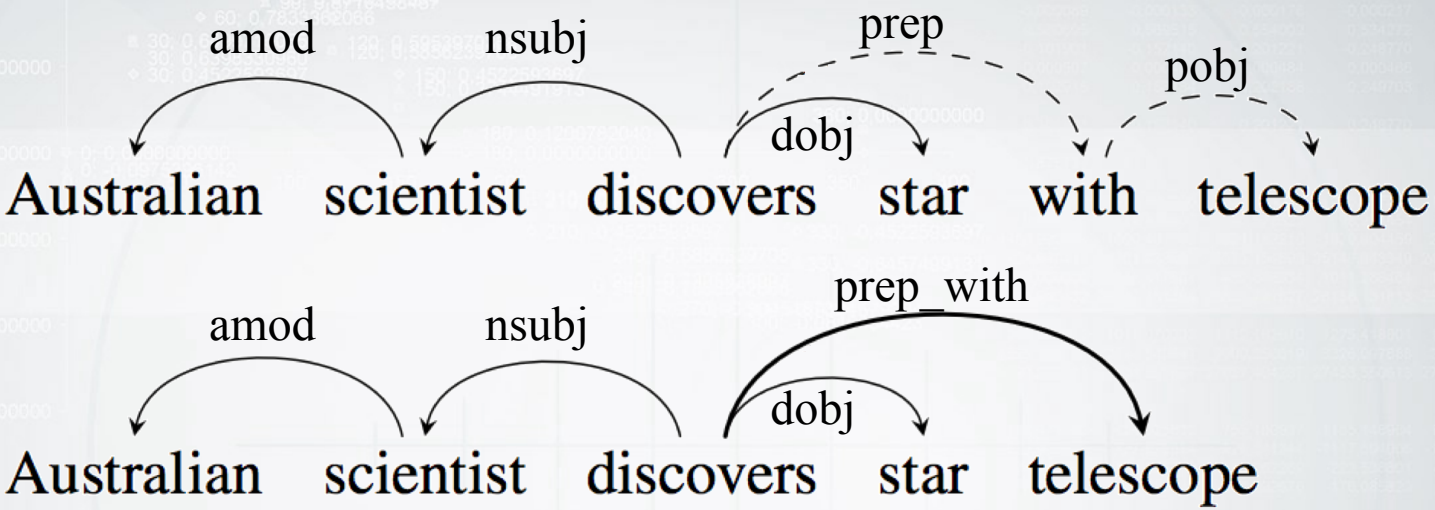
Vector Space Models of Semantics

- **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?



WORD	CONTEXTS
australian	scientist/amod^{-1}
scientist	australian/amod, discovers/nsubj^{-1}
discovers	scientist/nsubj, star/dobj, telescope/prep_with
star	discovers/dobj^{-1}
telescope	discovers/prep_with^{-1}

Omer Levy, Yoav Goldber, Dependency-Based Word Embeddings, ACL-2014.

What is a context?

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