

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:

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