

# Distributional Hypothesis

---

- Words that occur in *similar contexts* tend to have *similar meanings*.

“You shall know a word by the company it keeps.”  
(Firth, J. R. 1957:11)



# Two Kinds of Word Associations

---

- Two words have **first-order co-occurrence** (sometimes called syntagmatic association) if they are typically nearby each other.
  - For example, *wrote* is a first-order associate of *book* or *poem*.
- Two words have **second-order co-occurrence** (sometimes called paradigmatic association) if they have similar neighbours.
  - For example, *wrote* is a second-order associate of *said* or *remarked*.

What context window length should we choose accordingly:  
long or short?

# Context Window Length

---

- Shorter context windows tend to lead to representations that are a bit more syntactic.
  - When the vectors are computed from **short** context windows, the words closest to a target word  $w$  tend to be semantically similar words with the same parts-of-speech.
  - When vectors are computed from **long** context windows, the words closest to a target word  $w$  tend to be words that are topically related but not similar.