NLP
Text Similarity

Semantic Similarity: Synonymy and Other Semantic Relations
Synonyms and paraphrases

• Example: post–close market announcements

The S&P 500 climbed 6.93, or 0.56 percent, to 1,243.72, its best close since June 12, 2001.

The Nasdaq gained 12.22, or 0.56 percent, to 2,198.44 for its best showing since June 8, 2001.

The DJIA rose 68.46, or 0.64 percent, to 10,705.55, its highest level since March 15.
Synonyms

• Different words (and also word compounds) can have similar meanings.
  – *tepid* and *lukewarm* have very similar meanings and can be substituted for one another (*tepid water* vs. *lukewarm water*).

• True synonyms are actually relatively rare.
  – even though *big* and *large* are often thought of as synonyms, consider the difference between *Big Leagues* and *Large Leagues.*

• The verbs *sweat* and *perspire* are also near synonyms.
  – However, they differ in their frequency of use and the type of text in which they are likely to appear.
Polysemy

- Polysemy is the property of words to have multiple senses.
- For example, the noun *book* can refer to the following:
  - A literary work (e.g., “Anna Karenina”)
  - A stack of pages (e.g., a notebook)
  - A record of business transactions (think “bookkeeper”)
  - A record of bets (think “bookmaker”)
  - A list of buy and sell orders in a financial market
Polysemy

• The same word can also have multiple parts of speech, each with its own set of senses. For example, the word *book*, as a verb can mean “make a reservation for” or “occupy”.
• The different senses of the same word don’t have to be equally frequent.
• Some of the senses may overlap (e.g., the first two senses of *book* on the previous slide). That’s partially why different dictionaries list different sets of word senses for the same word.
  – “My favorite books are Anna Karenina and my father’s checkbook” 😊
• Some words can be highly polysemous (e.g., the verb “get” has at least 35 different meanings, according to Wordnet).
Other Semantic Relations

• Antonymy (near opposites)
  – *raise–lower*

• Hypernymy
  – a *deer* is a hypernym for *elk*

• Hyponymy (the inverse of hypernymy)

• Membership Meronymy:
  – a *flock* includes *sheep* (or *birds*)

• Part Meronymy:
  – a *table* has *legs*
Synsets

• Semantic relations hold between word senses, not between words.

• Examples:
  – the antonym of *hot* can be either *mild* or *cold* (or *unattractive*) depending on the specific sense of *hot*.
  – the immediate hypernym of *bar* can be one of the following, among others: *room*, *musical notation*, *obstruction*, *profession*, depending on the sense of *bar*.

• The term *synset* is used to group together all synonyms of the same word. If a word is polysemous, it may be associated with multiple synsets.