

# WordNet

## Lecture 7: October 2, 2013

CS886-2 Natural Language Understanding  
University of Waterloo

CS886 Lecture Slides (c) 2013 P. Poupart

1

## Word Semantics

- **WordNet**
  - Manually constructed database of word senses and relations
- **Latent Semantic Indexing**
  - Unsupervised learning technique based on singular value decomposition
- **Latent Dirichlet Allocation**
  - Unsupervised learning technique based on statistical learning

CS886 Lecture Slides (c) 2013 P. Poupart

2

## WordNet

- Free online database: <http://wordnet.princeton.edu/>
- Manually constructed
- Program searchable
- Primarily English
- Includes nouns, adjectives, verbs and adverbs
- It groups/relates words based on their meaning

CS886 Lecture Slides (c) 2013 P. Poupart

3

## Content

- **Synset**: set of synonyms (words with the same meaning)
  - E.g. {light, clean, clear, unclouded}
- **Gloss**: each synset has a short definition
  - E.g. {light, clean, clear, unclouded}: free from anything that dulls or dims.

CS886 Lecture Slides (c) 2013 P. Poupart

4

## Important Relations

- Superordinate/subordinate or Hyper/hyponymy or “**is a**” relation
  - Yields hierarchy of synsets
  - E.g., vehicle/car, fruit/mango, furniture/chair, mammal/dog
- Meronymy or “**part of**” relation
  - E.g. wheel/car, leg/chair

CS886 Lecture Slides (c) 2013 P. Poupart

5

## Common Uses of WordNet

- **Keyword expansion**
  - Keyword search: search for other words in the same synset
  - Question answering: rephrase question (and answer) by replace some words by other words from the same synset
- **Semantic relatedness**
  - The graph distance between two words in wordNet based on the “is a” relation provides a simple heuristic to indicate the degree of semantic relatedness

CS886 Lecture Slides (c) 2013 P. Poupart

6

## Challenges

- Manually maintaining WordNet is time consuming
- WordNet is not tailored to any domain or corpus
- WordNet will never be complete!