Semantics:
Word Sense Disambiguation

Word Sense Disambiguation

- Definition
  - Correct selection of the appropriate sense / meaning of a polysemous word in context
- In English, the most frequently occurring nouns have 7 senses and the most frequently occurring verbs have 11 senses
- How can we define different word senses?
  - Give a list of synonyms
  - Give a definition, which will necessarily use words that will have different senses, and these will (perhaps circularly) use words for definitions
- Coarse-grained senses distinguish core aspects of meaning
- Fine-grained senses also distinguish periphal aspects of meaning

Difficulties with synonyms

- True synonyms non-existent, or very rare
- Near-synonyms (Edmonds and Hirst)
  - Examples:
    - Error, blunder, mistake
    - Order, command, bid, enjoin, direct
  - Dimensions of synonym differentiation
    - Stylistic variation
      - Pissed, drunk, inebriated
    - Expressive variation
      - Attitude: skinny, thin, slim
      - Emotion: father, dad, daddy
    - . . .

Human Sense Disambiguation

- Sources of influence known from psycholinguistics research:
  - local context
    - the sentence or other surrounding text containing the ambiguous word restricts the interpretation of the ambiguous word
  - domain knowledge
    - the fact that a text is concerned with a particular domain activates only the sense appropriate to that domain
  - frequency data
    - the frequency of each sense in general usage affects its accessibility to the mind

Lesk Algorithm

- Original Lesk definition: measure overlap between sense definitions for all words in context. (Michael Lesk 1986)
  - Identify simultaneously the correct senses for all words in context
- Simplified Lesk (Kilgarriff & Rosensweig 2000): measure overlap between sense definitions of a word and current context
  - Identify the correct sense for one word at a time
  - Current context is the set of words in the surrounding sentence/paragraph/document.

Lesk Algorithm: A Simplified Version

- Algorithm for simplified Lesk:

  1. Retrieve from machine readable dictionary all sense definitions of the word to be disambiguated

  2. Determine the overlap between each sense definition and the current context

  3. Choose the sense that leads to highest overlap

Example: disambiguate PINE in

"Pine cones hanging in a tree"

- PINE

  1. kinds of evergreen tree with needle-shaped leaves

  2. waste away through sorrow or illness

Pine#1 ∩ Sentence = 1
Pine#2 ∩ Sentence = 0

Evaluations of Lesk Algorithm

- Initial evaluation by M. Lesk
  - 50-70% on short samples of text manually annotated set, with respect to Oxford Advanced Learner's Dictionary
  - Set of senses are "coarse-grained"

- Senseval conferences have shared tasks involving data for word sense disambiguation
  - Uses WordNet senses (more fine-grained and thus more difficult)
  - Evaluation on Senseval-2 all-words data, with back-off to most frequent sense (Vasilescu, Langlais, Lapalme 2004)
    - Original Lesk: 42%
    - Simplified Lesk: 58%

WSD algorithm development in Senseval

- All-word task
  - Given an entire text, disambiguate every content word in the text
  - Use general-purpose lexicon with senses
  - Can use a labeled corpus
    - SemCor is a subset of the Brown corpus with 234,000 words labeled with WordNet senses
    - Additional corpora developed through Senseval

Sense Tagged Corpus

- Examples of text where words are annotated with their sense from WordNet

Bonnie and Clyde are two really famous criminals, I think they were **bank/1** robbers
My **bank/1** charges too much for an overdraft.
I went to the **bank/1** to deposit my check and get a new ATM card.
The University of Minnesota has an East and a West **Bank/2** campus right on the Mississippi River.
My grandfather planted his pole in the **bank/2** and got a great big catfish!
The **bank/2** is pretty muddy, I can't walk there.

Classification approach to WSD

- Train a classification algorithm that can label each (open-class) word with the correct sense, given the context of the word
- Training set is the hand-labeled corpus of senses
- The context is represented as a set of "features" of the word and includes information about the surrounding words
  - Features are similar to those used directly by the Lesk algorithm
- Result of training is a model that is used by the classification algorithm to label words in the test set, and ultimately, in new text examples
- In the Senseval conferences, a number of systems in range of 70-80% accuracy for English Lexical Sample task

Word Similarity Features

- For each word in the context, compute a similarity measure between that word and the words in the definitions to be disambiguated
- Similarity measures
  - Can be defined from a semantic relation lexicon, such as WordNet
  - One example is path similarity
    - For any two words, gives a number between 0 and 1 based on the shortest path between the two words in the WordNet hypernym/hyponym hierarchy

WSD classification features

- Collocational features
  - Information about words in specific positions (i.e. previous word)
  - Typical features include the word itself, its stem and its POS tag
  - Example feature set:
    2 words to the left and right of the target word and their POS tags

    *An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.*

    [guitar, NN, and, CC, player, NN, stand, VB]
- Syntactic features
  - Predicate-argument relations
    - Verb-object, subject-verb,
  - Heads of Noun and Verb Phrases

WSD classification features

- Associated words features: for each word to be disambiguated, collect a small number of frequently-used context words.
  - Example: for each word, collect the 12 most frequent words from a collection of sentences drawn from the corpus as the limited set.

    For bass, the 12 most frequent context words from the WSJ are:

    [fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band]
- Represent these words as a bag-of-words feature:
  - The features of bass in the previous sentence (represented as 1 or 0 indicating the presence or not of the word in a window of size 10):

    [0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0]