EECS 498-004: Introduction to Natural Language Processing
Instructor: Prof. Lu Wang
Computer Science and Engineering
University of Michigan
https://web.eecs.umich.edu/~wangluxy/

Outline
• Word Senses and Word Relations
  • Word Similarity
  • Word Sense Disambiguation

Terminology: lemma and wordform
• A lemma or citation form
  • Same stem, part of speech, rough semantics
• A wordform
  • The inflected word as it appears in text

<table>
<thead>
<tr>
<th>Wordform</th>
<th>Lemma</th>
</tr>
</thead>
<tbody>
<tr>
<td>banks</td>
<td>bank</td>
</tr>
<tr>
<td>sung</td>
<td>sing</td>
</tr>
<tr>
<td>duermes</td>
<td>dormir</td>
</tr>
</tbody>
</table>

Lemmas have senses
• One lemma “bank” can have many meanings:
  Sense 1: “a bank can hold the investments in a custodial account…
  Sense 2: “as agriculture burgeons on the east bank the river will shrink even more”
• Sense (or word sense)
  • A discrete representation of an aspect of a word’s meaning.
  • The lemma bank here has two senses

Homonymy
Homonyms: words that share a form (spell or sound alike) but have unrelated, distinct meanings:
• bank₁: financial institution, bank₂: sloping land
• bat₁: club for hitting a ball, bat₂: nocturnal flying mammal
1. Homographs (bank/bank, bat/bat)
2. Homophones:
   1. Write and right
   2. Piece and peace

Homonymy causes problems for NLP applications
• Information retrieval
  • “bat care”
• Machine Translation
  • bat: murciélago (animal) or bate (for baseball)
• Text-to-Speech
  • bass (stringed instrument) vs. bass (fish)
1. The **bank** was constructed in 1875 out of local red brick.
2. I withdrew the money from the **bank**

---

**Polysemy**

1. The **bank** was constructed in 1875 out of local red brick.
2. I withdrew the money from the **bank**

*Are those the same sense?*
- Sense 1: "The building belonging to a financial institution"
- Sense 2: "A financial institution"

*A polysemous word has related meanings*
- Most non-rare words have multiple meanings

---

**Metonymy or Systematic Polysemy:**

A systematic relationship between senses

- Lots of types of polysemy are systematic:
  - School, university, hospital
  - All can mean the institution or the building.
- A systematic relationship:
  - Building: Organization
- Other such kinds of systematic polysemy:
  - Author (Jane Austen wrote Emma)
  - Works of Author (I love Jane Austen)
  - Tree (Plums have beautiful blossoms)
  - Fruit (I ate a preserved plum)

---

**How do we know when a word has more than one sense?**

- The "zeugma" test: Two senses of *serve*?
  - Which flights *serve* breakfast?
  - Does Lufthansa *serve* Philadelphia?

---

**Synonyms**

- Words that have the same meaning in some or all contexts.
  - filbert / hazelnut
  - couch / sofa
  - big / large
  - automobile / car
  - vomit / throw up
  - Water / H2O

- Two words are synonyms if they can be substituted for each other in all situations (strict/perfect definition).
Synonyms

- But there are few (or no) examples of perfect synonymy.
  - Even if many aspects of meaning are identical
  - Still may not preserve the acceptability based on notions of politeness, slang, register, genre, etc.
- Example:
  - Water/H₂O
  - Big/large
  - Brave/courageous

Synonymy is a relation between senses rather than words

- Consider the words big and large
- Are they synonyms?
  - How big is that plane?
  - Would I be flying on a large or small plane?
- How about here:
  - Miss Nelson became a kind of big sister to Benjamin.
  - Miss Nelson became a kind of large sister to Benjamin.
- Why?
  - big has a sense that means being older, or grown up
  - large lacks this sense

Antonyms

- Senses that are opposites with respect to one feature of meaning
- Otherwise, they are very similar!
- dark/light   short/long   fast/slow   rise/fall   hot/cold   up/down   in/out
- More formally: antonyms can
  - define a binary opposition or be at opposite ends of a scale
  - long/short, fast/slow
  - Be reversives:
    - rise/fall, up/down

Hyponymy and Hypernymy

- One sense is a **hyponym** of another if the first sense is more specific, denoting a subclass of the other
  - car is a hyponym of vehicle
  - mango is a hyponym of fruit
- Conversely **hypernym/superordinate** ("hyper is super")
  - vehicle is a hypernym of car
  - fruit is a hypernym of mango

<table>
<thead>
<tr>
<th>Superordinate/hypernym</th>
<th>Subordinate/hyponym</th>
</tr>
</thead>
<tbody>
<tr>
<td>vehicle</td>
<td>fruit</td>
</tr>
<tr>
<td>car</td>
<td>mango</td>
</tr>
<tr>
<td>chair</td>
<td></td>
</tr>
</tbody>
</table>

Hyponymy more formally

- Extensional:
  - The class denoted by the superordinate extensionally includes the class denoted by the hyponym
- Entailment:
  - A sense A is a hyponym of sense B if being an A entails being a B
  - Hypernymy is usually transitive
    - (A hypo B and B hypo C entails A hypo C)
- Another name: the IS-A hierarchy
  - A IS-A B (or A ISA B)
  - B subsumes A

<table>
<thead>
<tr>
<th>Superordinate/hypernym</th>
<th>Subordinate/hyponym</th>
</tr>
</thead>
<tbody>
<tr>
<td>car</td>
<td>mango</td>
</tr>
<tr>
<td>chair</td>
<td></td>
</tr>
</tbody>
</table>
Hyponymy more formally

• Extensional:
  • The class denoted by the superordinate extensionally includes the class denoted by
    the hyponym

• Entailment:
  • A sense A is a hyponym of sense B if being an A entails being a B
  • (A hypo B and B hypo C entails A hypo C)

• Another name: the IS-A Hierarchy
  • A IS-A B (or A ISA B)
  • B subsumes A

<table>
<thead>
<tr>
<th>Superordinate/hyponym</th>
<th>Subordinate/hyponym</th>
</tr>
</thead>
<tbody>
<tr>
<td>vehicle</td>
<td>fruit</td>
</tr>
<tr>
<td>furniture</td>
<td>car</td>
</tr>
<tr>
<td>mango</td>
<td>chair</td>
</tr>
</tbody>
</table>

Hyponyms and Instances

• An instance is an individual, a proper noun that is a unique entity
  • San Francisco is an instance of city

• But city is a class
  • city is a hyponym of municipality...location...

Meronymy

• The part-whole relation
  • A leg is part of a chair; a wheel is part of a car.
  • Wheel is a meronym of car, and car is a holonym of wheel.

WordNet 3.0

• A hierarchically organized lexical database
  • On-line thesaurus + aspects of a dictionary
  • Noun: 117,798
  • Verb: 11,529
  • Adjective: 22,479
  • Adverb: 4,481

EuroWordNet

• WordNets for
  • Dutch
  • Italian
  • Spanish
  • German
  • French
  • Czech
  • Estonian

Senses of “bass” in Wordnet

Noun

• 5 (0) bass (the lowest part of the musical range)
• 5 (0) bass, bass part (the lowest part in polyphonic music)
• 5 (0) bass, bass (an adult male singer with the lowest voice)
• 5 (0) baso bass, bass (the lean flesh of a saltwater fish of the family Serranidae)
• 5 (0) freshwater bass, bass (any of various North American freshwater fish with lean flesh except the genus Micropterus)
• 5 (0) bass, bass voice, basso (the lowest adult male singing voice)
• 5 (0) bass, bass member, bass member (the member with the lowest range of a family of musical instruments)
• 5 (0) bass (nonscientific name for any of numerous edible marine and freshwater spiny-finned fishes)

Adjective

• 5 (0) bass, deep (having or denoting a low vocal or instrumental range; "a deep voice"); "a bass voice is lower than a baritone voice"); "a bass clarinet"
How is “sense” defined in WordNet?

• The synset (synonym set), the set of near-synonyms, instantiates a sense or concept, with a gloss:
  • Example: chump as a noun with the gloss:
    “a person who is gullible and easy to take advantage of”
  • This sense of “chump” is shared by 9 words:
    chump¹, fool², gull³, mark⁴, patsy⁵, fall guy⁶, sucker⁷, soft touch⁸, mug⁹
  • Each of these senses has the same gloss
    • (Not every sense; sense 2 of gull is the aquatic bird)

WordNet: Viewed as a graph

WordNet Verb Relations

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyponym</td>
<td>From verbs to subordinate events</td>
<td>jey³ → travel⁶</td>
</tr>
<tr>
<td>Troponym</td>
<td>From events to subordinate event</td>
<td>wash¹ → soil⁸</td>
</tr>
<tr>
<td>Entails</td>
<td>From verbs (events) to the verbs (events) they entail</td>
<td>snore¹ → sleep¹</td>
</tr>
<tr>
<td>Antonym</td>
<td>Semantic opposition between lemmas</td>
<td>increase¹ → decrease¹</td>
</tr>
<tr>
<td>Derivationally</td>
<td>Lemmas with same morphological root</td>
<td>destroy¹ → destruction¹</td>
</tr>
</tbody>
</table>

WordNet 3.0

• Where it is:
  • http://wordnetweb.princeton.edu/perl/webwn
• Libraries
  • Python: WordNet from NLTK
    • http://www.nltk.org/Home
Outline

- Word Senses and Word Relations
  - Word Similarity
  - Word Sense Disambiguation

Why word similarity

- A practical component in lots of NLP tasks
  - Question answering
  - Natural language generation
  - Automatic essay grading
  - Plagiarism detection

Word Similarity

- Synonymy: a binary relation
  - Two words are either synonymous or not
- Similarity (or distance): a looser metric (more useful in practice!)
  - Two words are more similar if they share more features of meaning
  - Similarity is properly a relation between senses
    - Bank is similar to fund
    - Bank is similar to slope
  - But we’ll compute similarity over both words and senses

WordNet: Viewed as a graph

Two classes of similarity algorithms

- Thesaurus-based algorithms
  - Are words “nearby” in hypernym hierarchy?
  - Do words have similar glosses (definitions)?
- Distributional algorithms
  - Do words have similar distributional contexts?

Path-based similarity

- Two concepts (senses/synsets) are similar if they are near each other in the thesaurus hierarchy
  - have a short path between them
  - concepts have path 1 to themselves
Refinements to path-based similarity

- \( \text{pathlen}(c_1, c_2) = 1 + \text{number of edges in the shortest path} \) in the hypernym graph between sense nodes \( c_1 \) and \( c_2 \)
- ranges from 0 to 1 (identity)

- \( \text{simpath}(c_1, c_2) = \frac{1}{\text{pathlen}(c_1, c_2)} \)

- \( \text{wordsim}(w_1, w_2) = \max_{c_1 \in \text{senses}(w_1), c_2 \in \text{senses}(w_2)} \text{sim}(c_1, c_2) \)

Example: path-based similarity

\[ \text{simpath}(\text{nickel}, \text{coin}) = \frac{1}{2} = 0.5 \]
\[ \text{simpath}(\text{fund}, \text{budget}) = \frac{1}{2} = 0.5 \]
\[ \text{simpath}(\text{nickel}, \text{currency}) = \frac{1}{4} = 0.25 \]
\[ \text{simpath}(\text{nickel}, \text{money}) = \frac{1}{6} = 0.17 \]
\[ \text{simpath}(\text{nickel}, \text{standard}) = \frac{1}{6} = 0.17 \]

Problem with basic path-based similarity

- Assumes each link represents a uniform distance
- But \( \text{nickel} \) to \( \text{money} \) seems to us to be closer than \( \text{nickel} \) to \( \text{standard} \)
- Nodes high in the hierarchy are very abstract

- We instead want a metric that
  - Represents the cost of each edge independently
  - Words connected only through abstract nodes are less similar

Information content similarity metrics

Let’s define \( P(c) \) as:

- The probability that a randomly selected word in a corpus is an instance of concept \( c \)
- Formally: there is a distinct random variable, ranging over words, associated with each concept in the hierarchy
  - for a given concept, each observed word is either
    - a member of that concept with probability \( P(c) \)
    - not a member of that concept with probability \( 1 - P(c) \)
- All words are members of the root node (Entity)
  - \( P(\text{Entity}) = 1 \) (in practice, it may not be 1)
  - The lower a node in hierarchy, the lower its probability

\[ P(c) = \frac{\sum_{w \in \text{words}(c)} \text{count}(w)}{N} \]
Information content similarity

- WordNet hierarchy augmented with probabilities P(c)

<table>
<thead>
<tr>
<th>Entity</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>entry</td>
<td>0.395</td>
</tr>
<tr>
<td>inanimate object</td>
<td>0.167</td>
</tr>
<tr>
<td>natural object</td>
<td>0.0163</td>
</tr>
<tr>
<td>geological formation</td>
<td>0.00176</td>
</tr>
</tbody>
</table>

- LIN (LCS(c1, c2)) = The most informative (lowest) node in the hierarchy subsuming both c1 and c2

Using information content for similarity: the Resnik method

- The similarity between two words is related to their common information
- The more two words have in common, the more similar they are
- Resnik: measure common information as:
  - The information content of the lowest common subsumer of the two nodes
  - \( \text{sim}_{\text{resnik}}(c_1, c_2) = \text{IC}(LCS(c_1, c_2)) = -\log P(LCS(c_1, c_2)) \)

Dekang Lin method

- Using information content to evaluate semantic similarity in a taxonomy. ICML 1995.

Dekang Lin similarity theorem

- The similarity between A and B is measured by the ratio between the amount of information needed to state the commonality of A and B and the information needed to fully describe what A and B are
  \[ \text{sim}_{\text{lin}}(A, B) = \frac{\text{IC}(\text{common}(A, B))}{\text{IC}(\text{description}(A, B))} \]
- Lin (altering Resnik) defines IC(common(A, B)) as 2 x information of the LCS
  \[ \text{sim}_{\text{lin}}(c_1, c_2) = \frac{2 \log P(LCS(c_1, c_2))}{\log P(c_1) + \log P(c_2)} \]

Lin similarity function

\[ \text{sim}_{\text{lin}}(A, B) = \frac{2 \log P(LCS(c_1, c_2))}{\log P(c_1) + \log P(c_2)} \]
\[ \text{sim}_{\text{lin}}(\text{hill}, \text{coast}) = \frac{2 \log P(\text{geological-formation})}{\log P(\text{hill}) + \log P(\text{coast})} \]
\[ = \frac{2 \ln 0.00176}{\ln 0.0000189 + \ln 0.000216} = 0.9 \]
Libraries for computing thesaurus-based similarity

- NLTK
- WordNet::Similarity
  - http://wn-similarity.sourceforge.net/
  - Web-based interface:
    - http://marimba.d.umn.edu/cgi-bin/similarity/similarity.cgi

Evaluating similarity

- Extrinsic (task-based, end-to-end) Evaluation:
  - Question answering
  - Spell checking
  - Essay grading
  - Word sense disambiguation
- Intrinsic Evaluation:
  - Correlation between algorithm and human word similarity ratings (e.g. using Pearson correlation, Spearman correlation)
  - Wordsim353: 353 noun pairs rated 0-10, sim(plane,car)=5.77
  - Taking multiple-choice vocabulary tests
    - lexic is closest in meaning to:
      - imposed, believed, requested, correlated

Outline

- Word Senses and Word Relations
- Word Similarity
  - Word Sense Disambiguation

Lexical Ambiguity

- Most words in natural languages have multiple possible meanings.
  - "pen" (noun)
    - The dog is in the pen.
    - The ink is in the pen.
  - "take" (verb)
    - Take one pill every morning.
    - Take the first right past the stoplight.

Lexical Ambiguity

- Most words in natural languages have multiple possible meanings.
  - "pen" (noun)
    - The dog is in the pen.
    - The ink is in the pen.
  - "take" (verb)
    - Take one pill every morning.
    - Take the first right past the stoplight.
  - Syntax helps distinguish meanings for different parts of speech of an ambiguous word.
    - "conduct" (noun or verb)
      - John's conduct in class is unacceptable.
      - John will conduct the orchestra on Thursday.
Motivation for Word Sense Disambiguation (WSD)

- Many tasks in natural language processing require disambiguation of ambiguous words.
  - Question Answering
  - Information Retrieval
  - Machine Translation
  - Text Mining
  - Phone Help Systems

Senses Based on Needs of Translation

- Only distinguish senses that are translated to different words in some other language,
  - play: tocar vs. jugar
  - know: conocer vs. saber
  - be: ser vs. estar
  - leave: salir vs dejar
  - take: llevar vs. tomar vs. sacar
- May still require overly fine-grained senses
  - river in French is either:
    - fleuve: flows into the ocean
    - rivièrè: does not flow into the ocean

Word Sense Disambiguation (WSD)

- Given
  - A word in context (The dog is in the pen)
  - A fixed inventory of potential word senses (pen¹, pen²)
- Decide which sense of the word this is
- What set of senses?
  - In general: the senses in a thesaurus like WordNet
    - English-to-Spanish MT: set of Spanish translations
    - Speech Synthesis: homographs like bass and bow

Two variants of WSD task

- Lexical Sample task
  - Small pre-selected set of target words (line, plant)
  - And inventory of senses for each word
  - Supervised machine learning: train a classifier for each word
- All-words task
  - Every word in an entire text
  - A lexicon with senses for each word
  - Data sparseness: can’t train word-specific classifiers

WSD Methods

- Supervised Machine Learning
- Thesaurus/Dictionary Methods
- Semi-Supervised Learning

Supervised Machine Learning Approaches

- Supervised machine learning approach:
  - a training corpus of words tagged in context with their sense
  - used to train a classifier that can tag words in new text
- Summary of what we need:
  - the tag set ("sense inventory")
  - the training corpus
  - A set of features extracted from the training corpus
  - A classifier
Supervised WSD 1: WSD Tags

- What’s a tag?
  A dictionary sense?
- For example, for WordNet an instance of “bass” in a text has 8 possible tags or labels (bass1 through bass8, as noun).

8 senses of “bass” in WordNet

1. bass - (the lowest part of the musical range)
2. bass, bass part - (the lowest part in polyphonic music)
3. bass, basso - (an adult male singer with the lowest voice)
4. sea bass, bass - (flesh of lean-fleshed saltwater fish of the family Serranidae)
5. freshwater bass, bass - (any of various North American lean-fleshed freshwater fishes especially of the genus Micropterus)
6. bass, bass voice, basso - (the lowest adult male singing voice)
7. bass - (the member with the lowest range of a family of musical instruments)
8. bass - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Supervised WSD 2: Get a corpus

- Lexical sample task:
  Line-hard serve corpus - 4000 examples of each
  Interest corpus - 2369 sense-tagged examples
- All words:
  Semantic concordance: a corpus in which each open-class word is labeled with a sense from a specific dictionary/thesaurus.
  Senseval: 234,000 words from Brown Corpus, manually tagged with WordNet senses
  Senseval: 3 competition corpora - 2081 tagged word tokens

Supervised WSD 3: Extract feature vectors

Feature vectors

- A simple representation for each observation (each instance of a target word)
- Vectors of sets of feature/value pairs
- Represented as an ordered list of values
- These vectors represent, e.g., context—the window of words around the target

Lexical Ambiguity

- Most words in natural languages have multiple possible meanings.
  - “pen” (noun)
    - The dog is in the pen.
  - “pen” (verb)
    - Take one pill every morning.
  - “take” (verb)
    - Take the first right past the stoplight.
Two kinds of features in the vectors

- **Collocational features** and **bag-of-words features**
  - **Collocational**
    - Features about words at specific positions near target word
      - Often limited to just word identity and POS
  - **Bag-of-words**
    - Features about words that occur anywhere in the window (regardless of position)
      - Typically limited to frequency counts

Examples

- Example text (WSJ):
  An electric guitar and **bass** player stand off to one side not really part of the scene
  - Assume a window of +/- 2 from the target

Bag-of-words features

- "an unordered set of words" – position ignored
- Counts of words occur within the window.
- First choose a vocabulary
- Then count how often each of those terms occurs in a given window
  - sometimes just a binary "indicator" 1 or 0

Co-Occurrence Example

- Assume we've settled on a possible vocabulary of 12 words in "bass" sentences:
  [fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band]

  - The vector for:
    - **guitar and bass player stand**
      [0,0,0,1,0,0,0,0,0,1,0]
Syntactic Relations (Ambiguous Verbs)

• For an ambiguous verb, it is very useful to know its direct object.
  • 1. “played the game”
  • 2. “played the guitar”
  • 3. “played the risky and long-lasting card game”
  • 4. “played the beautiful and expensive guitar”
  • 5. “played the big brass tuba at the football game”
  • 6. “played the game listening to the drums and the tubas”

• May also be useful to know its subject:
  • “The game was played while the band played.”
  • “The game that included a drum and a tuba was played on Friday.”

Syntactic Relations (Ambiguous Nouns)

• For an ambiguous noun, it is useful to know what verb it is an object of:
  • “played the piano and the horn”
  • “wounded by the rhinoceros’ horn”

• May also be useful to know what verb it is the subject of:
  • “the bank near the river loaned him $100”
  • “the bank is eroding and the bank has given the city the money to repair it”

Syntactic Relations (Ambiguous Adjectives)

• For an ambiguous adjective, it useful to know the noun it is modifying.
  • “a brilliant young man”
  • “a brilliant yellow light”
  • “a wooden writing desk”
  • “a wooden acting performance”

Classification: definition

• Input:
  • a word w in a text window d (which we’ll call a “document”)”
  • a fixed set of classes C = {c1, c2, ..., cJ}
  • A training set of m hand-labeled text windows again called “documents” \( (d_1, y_1), ... , (d_m, y_m) \), \( y_m \) is in C

• Output:
  • a predicted class \( c \in C \)

Classification Methods: Supervised Machine Learning

• Input:
  • a word w in a text window d (which we’ll call a “document”)”
  • a fixed set of classes C = {c1, c2, ..., cJ}
  • A training set of m hand-labeled text windows again called “documents” \( (d_1, y_1), ... , (d_m, y_m) \), \( y_m \) is in C

• Output:
  • a learned classifier \( \gamma : d \rightarrow c \)
Applying Naive Bayes to WSD

• \( P(c) \) is the prior probability of that sense
  • Counting in a labeled training set.
• \( P(w|c) \) conditional probability of a word given a particular sense
  • \( P(w|c) = \frac{\text{count}(w,c)}{\text{count}(c)} \)
  • We get both of these from a tagged corpus like SemCor

Choosing a class:
\[
P(f|d5) = \frac{1}{4} \times \frac{2}{9} \times \left(\frac{2}{9}\right)^2 \approx 0.00003
\]

WSD Evaluations and baselines

• Best evaluation: **extrinsic** (**end-to-end**, **task-based**) evaluation
  • Embed WSD algorithm in a task and see if you can do the task better!
• What we often do for convenience: **intrinsic evaluation**
  • Exact match sense accuracy
  • % of words tagged identically with the human-manual sense tags
  • Usually evaluate using **held-out data/test data** from same labeled corpus

Most Frequent Sense

• WordNet senses are ordered in frequency order
• So “most frequent sense” in WordNet = “take the first sense”
• Sense frequencies come from the SemCor corpus
The Simplified Lesk algorithm

- Let's disambiguate "bank" in this sentence:
  The bank can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.
- given the following two WordNet senses:

<table>
<thead>
<tr>
<th>Sense</th>
<th>Gloss</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>bank1</td>
<td>a financial institution that accepts deposits and channels the money into lending activities</td>
<td>&quot;he cashed a check at the bank&quot;, &quot;that bank holds the mortgage on my home&quot;</td>
</tr>
<tr>
<td>bank2</td>
<td>sloping land (especially the slope beside a body of water)</td>
<td>&quot;they pulled the canoe up on the bank&quot;, &quot;he sat on the bank of the river and watched the currents&quot;</td>
</tr>
</tbody>
</table>

The bank can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.

Semi-Supervised Learning

Problem: supervised and dictionary-based approaches require large hand-built resources. What if you don’t have so much training data?
Solution: Bootstrapping
  Generalize from a very small hand-labeled seed-set.

Sentences extracting using "fish" and "play"

The Simplified Lesk algorithm

Choose sense with most word overlap between gloss and context (not counting function words)

<table>
<thead>
<tr>
<th>Sense</th>
<th>Gloss</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>bank1</td>
<td>a financial institution that accepts deposits and channels the money into lending activities</td>
<td>&quot;he cashed a check at the bank&quot;, &quot;that bank holds the mortgage on my home&quot;</td>
</tr>
<tr>
<td>bank2</td>
<td>sloping land (especially the slope beside a body of water)</td>
<td>&quot;they pulled the canoe up on the bank&quot;, &quot;he sat on the bank of the river and watched the currents&quot;</td>
</tr>
</tbody>
</table>

The bank can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.

Summary: generating seeds

1) Hand labeling
2) "One sense per collocation":
   - A word reoccurring in collocation with the same word will almost surely have the same sense.
   - the word play occurs with the music sense of bass
   - the word fish occurs with the fish sense of bass
3) "One sense per discourse":
   - The sense of a word is highly consistent within a document - Yarowsky (1995)
   - (At least for non-function words, and especially topic-specific words)
Summary

• Word Sense Disambiguation: choosing correct sense in context
• Applications: MT, QA, etc.
• Three classes of Methods
  • Supervised Machine Learning: Naive Bayes classifier
  • Thesaurus/Dictionary Methods
  • Semi-Supervised Learning
• Main intuition
  • There is lots of information in a word's context
  • Simple algorithms based just on word counts can be surprisingly good