Lexical Semantics and Word Sense Disambiguation
Announcements

• Midterm sample questions on website

• Next class: midterm review for part of the class. Post your wishes for topics for the review on Piazza

• HW1 grades out. Mean is 81. Nice going!

• Following topics: semantic parsing, then to distributed semantics and word embeddings, neural nets.
Polysemy

• The bank is constructed from red brick. I withdrew the money from the bank.
• Are those the same sense?
• Or consider the following WSJ example
  • While some banks furnish sperm only to married women, others are less restrictive.
  • Which sense of bank is this?
    • Is it distinct from (homonymous with) the river bank sense?
    • How about the savings bank sense?
The sense of "bank" is in the sentence "While some allow banks to use sperm only to married women, others are more restrictive?"

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>river</td>
<td>bank</td>
</tr>
<tr>
<td>savings</td>
<td>bank</td>
</tr>
<tr>
<td>other</td>
<td></td>
</tr>
</tbody>
</table>
Polysemy

• A single lexeme with multiple related meanings (bank the building, bank the financial institution)
• Most non-rare words have multiple meanings
  • The number of meanings is related to its frequency
  • Verbs tend more to polysemy
  • Distinguishing polysemy from homonymy isn’t always easy (or necessary)
Metaphor and Metonymy

• Specific types of polysemy

• Metaphor:
  • Germany will pull Slovenia out of its economic slump.
  • I spent 2 hours on that homework.

• Metonymy
  • The White House announced yesterday.
  • This chapter talks about part-of-speech tagging
  • Bank (building) and bank (financial institution)
How do we know when a word has more than one sense?

• ATIS examples
  • Which flights serve breakfast?
  • Does America West serve Philadelphia?

• The “zeugma” test:
  • ?Does United serve breakfast and San Jose?
Synonyms

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

- Two lexemes are synonyms if they can be successfully substituted for each other in all situations
  - If so they have the same *propositional meaning*
Synonyms

• But there are few (or no) examples of perfect synonymy.
  • Why should that be?
  • 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 and H₂O
Some more terminology

• Lemmas and wordforms
  • A **lexeme** is an abstract pairing of meaning and form
  • A **lemma** or **citation form** is the grammatical form that is used to represent a **lexeme**.
    - *Carpet* is the lemma for *carpets*
    - *Dormir* is the lemma for *duermes*.
  • Specific surface forms *carpets, sung, duermes* are called **wordforms**

• The lemma *bank* has two **senses**:
  • Instead, a **bank** can hold the investments in a custodial account in the client’s name
  • But as agriculture burgeons on the east **bank**, the river will shrink even more.

• A **sense** is a discrete representation of one aspect of the meaning of a word
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, for instance, became a kind of *big* sister to Benjamin.
- Miss Nelson, for instance, 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 their meaning
• Otherwise, they are very similar!
  • dark / light
  • short / long
  • hot / cold
  • up / down
  • in / out

• More formally: antonyms can
  • define a binary opposition or at opposite ends of a scale (long/short, fast/slow)
  • Be **reversives**: rise/fall, up/down
Could we say that antonyms are very similar? How are they similar?
Hyponymy

- 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*
  - *dog* is a hyponym of *animal*
  - *mango* is a hyponym of *fruit*
- Conversely
  - *vehicle* is a hypernym/superordinate of *car*
  - *animal* is a hypernym of *dog*
  - *fruit* is a hypernym of *mango*
Hypernymy 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

• Hyponymy is usually transitive
  • (A hypo B and B hypo C entails A hypo C)
Why would hypernyms/hyponyms be important in constructing a meaning representation?
II. WordNet

- A hierarchically organized lexical database
- On-line thesaurus + aspects of a dictionary
  - Versions for other languages are under development

<table>
<thead>
<tr>
<th>Category</th>
<th>Unique Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>117,097</td>
</tr>
<tr>
<td>Verb</td>
<td>11,488</td>
</tr>
<tr>
<td>Adjective</td>
<td>22,141</td>
</tr>
<tr>
<td>Adverb</td>
<td>4,601</td>
</tr>
</tbody>
</table>
WordNet

- Where it is:
  - [https://wordnet.princeton.edu/](https://wordnet.princeton.edu/)
Format of Wordnet Entries

The noun “bass” has 8 senses in WordNet.
1. bass¹ - (the lowest part of the musical range)
2. bass², bass part¹ - (the lowest part in polyphonic music)
3. bass³, basso¹ - (an adult male singer with the lowest voice)
4. sea bass¹, bass⁴ - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass¹, bass⁵ - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass⁶, bass voice¹, basso² - (the lowest adult male singing voice)
7. bass⁷ - (the member with the lowest range of a family of musical instruments)
8. bass⁸ - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

The adjective “bass” has 1 sense in WordNet.
1. bass¹, deep⁶ - (having or denoting a low vocal or instrumental range)
   "a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"
# WordNet Noun Relations

<table>
<thead>
<tr>
<th>Relation</th>
<th>Also called</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyponym</td>
<td></td>
<td>From concepts to superordinates</td>
<td><em>breakfast</em>¹ → <em>meal</em>¹</td>
</tr>
<tr>
<td>Hyponym</td>
<td></td>
<td>From concepts to subtypes</td>
<td><em>meal</em>¹ → <em>lunch</em>¹</td>
</tr>
<tr>
<td>Member Meronym</td>
<td>Superordinate</td>
<td>From groups to their members</td>
<td><em>faculty</em>² → <em>professor</em>¹</td>
</tr>
<tr>
<td>Member Meronym</td>
<td>Subordinate</td>
<td>From concepts to instances of the concept</td>
<td><em>composer</em>¹ → <em>Bach</em>¹</td>
</tr>
<tr>
<td>Has-Instance</td>
<td>Has-Member</td>
<td>From instances to their concepts</td>
<td><em>Austen</em>¹ → <em>author</em>¹</td>
</tr>
<tr>
<td>Has-Instance</td>
<td></td>
<td></td>
<td><em>copilot</em>¹ → <em>crew</em>¹</td>
</tr>
<tr>
<td>Instance</td>
<td></td>
<td>From members to their groups</td>
<td><em>table</em>² → <em>leg</em>³</td>
</tr>
<tr>
<td>Member Holonym</td>
<td>Member-Of</td>
<td>From wholes to parts</td>
<td><em>course</em>⁷ → <em>meal</em>¹</td>
</tr>
<tr>
<td>Part Holonym</td>
<td>Has-Part</td>
<td>From parts to wholes</td>
<td><em>leader</em>¹ → <em>follower</em>¹</td>
</tr>
<tr>
<td>Part Holonym</td>
<td>Part-Of</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# WordNet Verb Relations

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypernym</td>
<td>From events to superordinate events</td>
<td><em>fly</em>⁹ → <em>travel</em>⁵</td>
</tr>
<tr>
<td>Troponym</td>
<td>From a verb (event) to a specific manner elaboration of that verb</td>
<td><em>walk</em>¹ → <em>stroll</em>¹</td>
</tr>
<tr>
<td>Entails</td>
<td>From verbs (events) to the verbs (events) they entail</td>
<td><em>snore</em>¹ → <em>sleep</em>¹</td>
</tr>
<tr>
<td>Antonym</td>
<td>Opposites</td>
<td><em>increase</em>¹ ↔ <em>decrease</em>¹</td>
</tr>
</tbody>
</table>
WordNet Hierarchies

Sense 3
bass, basso --
(an adult male singer with the lowest voice)
=> singer, vocalist, vocalizer, vocaliser
  => musician, instrumentalist, player
  => performer, performing artist
  => entertainer
  => person, individual, someone...
    => organism, being
      => living thing, animate thing,
        => whole, unit
          => object, physical object
            => physical entity
              => entity

=> causal agent, cause, causal agency
  => physical entity
  => entity

Sense 7
bass --
(the member with the lowest range of a family of musical instruments)
=> musical instrument, instrument
  => device
    => instrumentality, instrumentation
      => artifact, artefact
        => whole, unit
          => object, physical object
            => physical entity
How is “sense” defined in WordNet?

• The set of near-synonyms for a WordNet sense is called a **synset (synonym set)**; it’s their version of a sense or a concept

• Example: *chump* as a noun to mean
  • ‘a person who is gullible and easy to take advantage of’

\{chump^1, fool^2, gull^1, mark^9, patsy^1, fall guy^1, sucker^1, soft touch^1, mug^2\}

• Each of these senses share this same gloss
• Thus for WordNet, the meaning of this sense of *chump* is this list.
Wordnet example
Word Sense Disambiguation
Word Sense Disambiguation (WSD)

- Given
  - a word in context,
  - A fixed inventory of potential word senses
- decide which sense of the word this is.
  - English-to-Spanish MT
    - Inventory is set of Spanish translations
  - Speech Synthesis
    - Inventory is homographs with different pronunciations like *bass* and *bow*
  - Automatic indexing of medical articles
    - MeSH (Medical Subject Headings) thesaurus entries
Two variants of WSD task

- **Lexical Sample task**
  - Small pre-selected set of target words
  - And inventory of senses for each word

- **All-words task**
  - Every word in an entire text
  - A lexicon with senses for each word
  - Sort of like part-of-speech tagging
    - Except each lemma has its own tagset
Approaches

• Supervised

• Semi-supervised
  • Unsupervised
    • Dictionary-based techniques
    • Selectional Association
  • Lightly supervised
    • Bootstrapping
    • Preferred Selectional Association
Supervised Machine Learning Approaches

• Supervised machine learning approach:
  • a training corpus of ?
  • used to train a classifier that can tag words in new text
  • Just as we saw for part-of-speech tagging, statistical MT.

• Summary of what we need:
  • the tag set ("sense inventory")
  • the training corpus
  • A set of features extracted from the training corpus
  • A classifier
What would be a good training corpus for semantic disambiguation?
Supervised WSD 1: WSD Tags

• What’s a tag?
WordNet

- http://www.cogsci.princeton.edu/cgi-bin/webwn
The noun ``bass'' has 8 senses 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)
Inventory of sense tags for *bass*

<table>
<thead>
<tr>
<th>WordNet Sense</th>
<th>Spanish Translation</th>
<th>Roget Category</th>
<th>Target Word in Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>bass^4</td>
<td>lubina</td>
<td>FISH/INSECT</td>
<td>...fish as Pacific salmon and striped <strong>bass</strong> and...</td>
</tr>
<tr>
<td>bass^4</td>
<td>lubina</td>
<td>FISH/INSECT</td>
<td>...produce filets of smoked <strong>bass</strong> or sturgeon...</td>
</tr>
<tr>
<td>bass^7</td>
<td>bajo</td>
<td>MUSIC</td>
<td>...exciting jazz <strong>bass</strong> player since Ray Brown...</td>
</tr>
<tr>
<td>bass^7</td>
<td>bajo</td>
<td>MUSIC</td>
<td>...play <strong>bass</strong> because he doesn't have to solo...</td>
</tr>
</tbody>
</table>
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.
    • SemCor: 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

• Weaver (1955)
• If one examines the words in a book, one at a time as through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of the words. [...] But if one lengthens the slit in the opaque mask, until one can see not only the central word in question but also say N words on either side, then if N is large enough one can unambiguously decide the meaning of the central word. [...] The practical question is: ``What minimum value of N will, at least in a tolerable fraction of cases, lead to the correct choice of meaning for the central word?''
• dishes

• bass
• washing dishes.
• simple dishes including
• convenient dishes to
• of dishes and

• free bass with
• pound bass of
• and bass player
• his bass while
plates or food: "convenient dishes to"

plates

food

Start the presentation to activate live content
If you see this message in presentation mode, install the add-in or get help at PollEv.com/app
• “In our house, everybody has a career and none of them includes washing dishes,” he says.

• In her tiny kitchen at home, Ms. Chen works efficiently, stir-frying several simple dishes, including braised pig’s ears and chicken livers with green peppers.

• Post quick and convenient dishes to fix when you’re in a hurry.

• Japanese cuisine offers a great variety of dishes and regional specialties.
• We need more good teachers – right now, there are only a half a dozen who can play the free bass with ease.

• Though still a far cry from the lake’s record 52-pound bass of a decade ago, “you could fillet these fish again, and that made people very, very happy.” Mr. Paulson says.

• 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 again.

• Lowe caught his bass while fishing with pro Bill Lee of Killeen, Texas, who is currently in 144th place with two bass weighing 2-09.
Feature vectors

• A simple representation for each observation (each instance of a target word)
  • Vectors of sets of feature/value pairs
    • I.e. files of comma-separated values
  • These vectors should represent the window of words around the target

How big should that window be?
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, just as a sort of nod to gringo expectations perhaps
  • Assume a window of +/- 2 from the target
Examples

• Example text
  • 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
  • Assume a window of +/- 2 from the target
Collocational

- Position-specific information about the words in the window
- **guitar and bass player stand**
  - [guitar, NN, and, CC, player, NN, stand, VB]
  - $\text{Word}_{n-2}, \text{POS}_{n-2}, \text{word}_{n-1}, \text{POS}_{n-1}, \text{Word}_{n+1} \text{ POS}_{n+1}$...
  - In other words, a vector consisting of
  - [position n word, position n part-of-speech...]
Bag-of-words

• Information about the words that occur within the window.
• First derive a set of terms to place in the vector.
• Then note how often each of those terms occurs in a given window.
Co-Occurrence Example

• Assume we’ve settled on a possible vocabulary of 12 words that includes guitar and player but not and and stand

• guitar and bass player stand
  • [0,0,0,1,0,0,0,0,0,1,0,0]
  • Which are the counts of words predefined as e.g.,
  • [fish,fishing,viol, guitar, double,cello...
Classifiers

- Once we cast the WSD problem as a classification problem, then all sorts of techniques are possible
  - Naïve Bayes (the easiest thing to try first)
  - Decision lists
  - Decision trees
  - Neural nets
  - Support vector machines
  - Nearest neighbor methods...
Classifiers

• The choice of technique, in part, depends on the set of features that have been used
  • Some techniques work better/worse with features with numerical values
  • Some techniques work better/worse with features that have large numbers of possible values
    • For example, the feature the word to the left has a fairly large number of possible values
Naïve Bayes

• \( \hat{s} = \arg \max_{s \in S} p(s | V) \), or \( \arg \max_{s \in S} \frac{p(V | s)p(s)}{p(V)} \)

• Where \( s \) is one of the senses \( S \) possible for a word \( w \) and \( V \) the input vector of feature values for \( w \)

• Assume features independent, so probability of \( V \) is the product of probabilities of each feature, given \( s \), so

\[
p(V | s) = \prod_{j=1}^{n} p(v_j | s)
\]

\( p(V) \) same for any \( \hat{s} \)

• Then

\[
\hat{s} = \arg \max_{s \in S} p(s) \prod_{j=1}^{n} p(v_j | s)
\]
• How do we estimate $p(s)$ and $p(v_j | s)$?
  • $p(s_i)$ is max. likelihood estimate from a sense-tagged corpus ($\text{count}(s_i, w_j)/\text{count}(w_j)$) – how likely is bank to mean ‘financial institution’ over all instances of bank?
  • $P(v_j | s)$ is max. likelihood of each feature given a candidate sense ($\text{count}(v_j, s)/\text{count}(s)$) – how likely is the previous word to be ‘river’ when the sense of bank is ‘financial institution’

• Calculate

$$\hat{s} = \arg\max_{s \in S} p(s) \prod_{j=1}^{n} p(v_j | s)$$

take the highest scoring sense as the most likely choice
Naïve Bayes Test

• On a corpus of examples of uses of the word line, naïve Bayes achieved about 73% correct

• Good?
Decision Lists: another popular method

- A case statement....

<table>
<thead>
<tr>
<th>Rule</th>
<th>Sense</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>fish</em> within window</td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>striped bass</em></td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>guitar</em> within window</td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>bass player</em></td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>piano</em> within window</td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>tenor</em> within window</td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>sea bass</em></td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>play/N bass</em></td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>river</em> within window</td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>violin</em> within window</td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>salmon</em> within window</td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>on bass</em></td>
<td>$\Rightarrow$</td>
</tr>
<tr>
<td><em>bass are</em></td>
<td>$\Rightarrow$</td>
</tr>
</tbody>
</table>
Learning Decision Lists

• Restrict the lists to rules that test a single feature (1-decisionlist rules)
• Evaluate each possible test and rank them based on how well they work.
• Glue the top-N tests together and call that your decision list.
Yarowsky

• On a binary (homonymy) distinction used the following metric to rank the tests

\[
\frac{P(\text{Sense}_1 \mid \text{Feature})}{P(\text{Sense}_2 \mid \text{Feature})}
\]

• This gives about 95% on this test...
WSD Evaluations and baselines

- *In vivo* versus *in vitro* evaluation
- In vitro evaluation is most common now
  - Exact match **accuracy**
    - % of words tagged identically with manual sense tags
  - Usually evaluate using held-out data from same labeled corpus
    - Problems?
    - Why do we do it anyhow?

- Baselines
  - Most frequent sense
  - The Lesk algorithm
Most Frequent Sense

• Wordnet senses are ordered in frequency order

• So “most frequent sense” in wordnet = “take the first sense”

<table>
<thead>
<tr>
<th>Freq</th>
<th>Synset</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>338</td>
<td>plant^1, works, industrial plant</td>
<td>buildings for carrying on industrial labor</td>
</tr>
<tr>
<td>207</td>
<td>plant^2, flora, plant life</td>
<td>a living organism lacking the power of locomotion</td>
</tr>
<tr>
<td>2</td>
<td>plant^3</td>
<td>something planted secretly for discovery by another</td>
</tr>
<tr>
<td>0</td>
<td>plant^4</td>
<td>an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience</td>
</tr>
</tbody>
</table>
Ceiling

• Human inter-annotator agreement
  • Compare annotations of two humans
  • On same data
  • Given same tagging guidelines

• Human agreements on all-words corpora with Wordnet style senses
  • 75%-80%
Unsupervised Methods

WSD: Dictionary/Thesaurus methods

- The Lesk Algorithm
- Selectional Restrictions
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:

| bank¹ | Gloss: | a financial institution that accepts deposits and channels the money into lending activities  
“he cashed a check at the bank”, “that bank holds the mortgage on my home” |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Examples:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“he cashed a check at the bank”, “that bank holds the mortgage on my home”</td>
</tr>
</tbody>
</table>

| bank² | Gloss: | sloping land (especially the slope beside a body of water)  
“They pulled the canoe up on the bank”, “he sat on the bank of the river and watched the currents” |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Examples:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>“they pulled the canoe up on the bank”, “he sat on the bank of the river and watched the currents”</td>
</tr>
</tbody>
</table>
Original Lesk: pine cone

pine  1  kinds of evergreen tree with needle-shaped leaves
     2  waste away through sorrow or illness
cone  1  solid body which narrows to a point
      2  something of this shape whether solid or hollow
      3  fruit of certain evergreen trees
Corpus Lesk

• Add corpus examples to glosses and examples
• The best performing variant
Disambiguation via Selectional Restrictions

• “Verbs are known by the company they keep”
  • Different verbs select for different thematic roles
    wash the *dishes* (takes washable-thing as patient)
    serve delicious *dishes* (takes food-type as patient)

• Method: another semantic attachment in grammar
  • Semantic attachment rules are applied as sentences are syntactically parsed, e.g.
    \[ VP \rightarrow V \ NP \]
    \[ V \rightarrow \text{serve } \text{<theme>} \{\text{theme:food-type}\} \]
  • Selectional restriction violation: no parse
• But this means we must:
  • Write selectional restrictions for each sense of each predicate – or use FrameNet
    • Serve alone has 15 verb senses
  • Obtain hierarchical type information about each argument (using WordNet)
    • How many hypernyms does dish have?
    • How many words are hyponyms of dish?

• But also:
  • Sometimes selectional restrictions don’t restrict enough (Which dishes do you like?)
  • Sometimes they restrict too much (Eat dirt, worm! I’ll eat my hat!)

• Can we take a statistical approach?
Semi-supervised Bootstrapping

• What if you don’t have enough data to train a system...

• Bootstrap
  • Pick a word that you as an analyst think will co-occur with your target word in particular sense
  • Grep through your corpus for your target word and the hypothesized word
  • Assume that the target tag is the right one
Bootstrapping

• For bass
  • Assume play occurs with the music sense and fish occurs with the fish sense
Sentences extracting using “fish” and “play”

more good teachers – right now, there are only a half a dozen who can
with ease.

ic guitar and bass player stand off to one side, not really part of the scene
nod to gringo expectations perhaps.

The New Jersey Jazz Society, in a fund-raiser for the American Jazz Hall
his historic night next Saturday, Harry Goodman, Mr. Goodman’s bro-
ver at the original concert, will be in the audience with other family me-
archers said the worms spend part of their life cycle in such fish as Pacific
ed bass and Pacific rockfish or snapper.

It started when fishermen decided the striped bass in Lake Mead were to
still a far cry from the lake’s record 52-pound bass of a decade ago, “y
the fish again, and that made people very, very happy,” Mr. Paulson says.
Where do the seeds come from?

1) Hand labeling

2) “One sense per discourse”:
   - The sense of a word is highly consistent within a document - Yarowsky (1995)
   - True for topic dependent words
   - Not so true for other POS like adjectives and verbs, e.g. make, take
   - Krovetz (1998) “More than one sense per discourse” argues it isn’t true at all once you move to fine-grained senses

3) One sense per collocation:
   - A word reoccurring in collocation with the same word will almost surely have the same sense.

Slide adapted from Chris Manning
Stages in the Yarowsky bootstrapping algorithm
Problems

- Given these general ML approaches, how many classifiers do I need to perform WSD robustly
  - One for each ambiguous word in the language
- How do you decide what set of tags/labels/senses to use for a given word?
  - Depends on the application
WordNet Bass

• Tagging with this set of senses is an impossibly hard task that’s probably overkill for any realistic application

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)
Senseval History

• ACL-SIGLEX workshop (1997)
  • Yarowsky and Resnik paper
• SENSEVAL-I (1998)
  • Lexical Sample for English, French, and Italian
• SENSEVAL-II (Toulouse, 2001)
  • Lexical Sample and All Words
  • Organization: Kilkgarriff (Brighton)
• SENSEVAL-III (2004)
• SENSEVAL-IV -> SEMEVAL (2007)
• SEMEVAL (2010)

SLIDE ADAPTED FROM CHRIS MANNING
WSD Performance

• Varies widely depending on how difficult the disambiguation task is
• Accuracies of over 90% are commonly reported on some of the classic, often fairly easy, WSD tasks (pike, star, interest)
• Senseval brought careful evaluation of difficult WSD (many senses, different POS)
• Senseval 1: more fine grained senses, wider range of types:
  • Overall: about 75% accuracy
  • Nouns: about 80% accuracy
  • Verbs: about 70% accuracy
Summary

• Lexical Semantics
  • Homonymy, Polysemy, Synonymy
  • Thematic roles

• Computational resource for lexical semantics
  • WordNet

• Task
  • Word sense disambiguation

• Next: semantic parsing