Word Sense Disambiguation (WSD)

- Given
  - A word in context,
  - A fixed inventory of potential word senses

- Decide which sense of the word this is

- What set of senses?
  - English-to-Spanish MT: set of Spanish translations
  - Speech Synthesis: homographs like bass and bow
  - In general: the senses in a thesaurus like WordNet
TheWordNet entry for the noun *bat* has the following distinct senses.

Cluster these senses by using the definitions of homonymy and polysemy.

- bat#1: nocturnal mouselike mammal
- bat#2: (baseball) a turn trying to get a hit
- bat#3: a small racket . . . for playing squash
- bat#4: the club used in playing cricket
- bat#5: a club used for hitting a ball in various games
Two Variants of WSD

- **Lexical Sample task**
  - Small pre-selected set of target words
  - And inventory of senses for each word
  - Typically supervised ML: classifier per 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
  - ~Like part-of-speech tagging
    - Except each lemma has its own tagset
    - Less human agreement so upper bound is lower

Approaches

- Supervised machine learning
- “Unsupervised”
  - Thesaurus/Dictionary-based techniques
  - Selectional Association
- Lightly supervised
  - Bootstrapping
  - Preferred Selectional Association
Supervised Machine Learning Approaches

- Supervised machine learning approach
  - Training corpus depends on task
  - Train a classifier that can tag words in new text
  - Just as we saw for part-of-speech tagging
- What do we need?
  - Tag set ("sense inventory")
  - Training corpus (words tagged in context with sense)
  - Set of features extracted from the training corpus
  - A classifier

Supervised WSD: 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).
Bass in WordNet

- The noun bass has 8 senses in WordNet:
  - bass - (the lowest part of the musical range)
  - bass, bass part - (the lowest part in polyphonic music)
  - bass, basso - (an adult male singer with the lowest voice)
  - sea bass, bass - (flesh of lean-fleshed saltwater fish of the family Serranidae)
  - freshwater bass, bass - (any of various North American lean-fleshed freshwater fishes especially of the genus Micropterus)
  - bass, bass voice, basso - (the lowest adult male singing voice)
  - bass - (the member with the lowest range of a family of musical instruments)
  - bass -(nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

<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>bass¹</td>
<td>lubina</td>
<td>FISH/INSECT</td>
<td>...fish as Pacific salmon and striped bass and...</td>
</tr>
<tr>
<td>bass²</td>
<td>lubina</td>
<td>FISH/INSECT</td>
<td>...produce filets of smoked bass or sturgeon...</td>
</tr>
<tr>
<td>bass³</td>
<td>bajo</td>
<td>MUSIC</td>
<td>...exciting jazz bass player since Ray Brown...</td>
</tr>
<tr>
<td>bass⁴</td>
<td>bajo</td>
<td>MUSIC</td>
<td>...play bass because he doesn’t have to solo...</td>
</tr>
</tbody>
</table>
What kind of Corpora?

- 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

SemCor

```
<wf pos=PRP>He</wf>
<wf pos=VB lemma=recognize wnsn=4 lexsn=2:31:00::>recognized</wf>
<wf pos=DT>the</wf>
<wf pos=NN lemma=gesture wnsn=1 lexsn=1:04:00::>gesture</wf>
<punc>.</punc>
```
What Kind of Features?

- 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?’”

Frequency-based WSD

- WordNet first sense heuristic, about 60-70% accuracy
- To improve, need context
  - Selectional restrictions
  - “Topic”
dishes
  • washing dishes he
  • simple dishes including
  • convenient dishes to
  • of dishes and

bass
  • free bass with
  • pound bass of
  • and bass player
  • his bass while

•S: (n) dish (a piece of dishware normally used as a container for holding or serving food) "we gave them a set of dishes for a wedding present"

•S: (n) dish (a particular item of prepared food) "she prepared a special dish for dinner"

....
dishes
- includes washing *dishes* he says
- several simple *dishes* including braised
- and convenient *dishes* to fix
- variety of *dishes* and regional

bass
- the free *bass* with ease
- 52 pound *bass* of a
- guitar and *bass* player stand
- caught his *bass* while fishing

- “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?*
What sort of Features?

- **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

Example

- 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
Collocations

- Position-specific information about the words in the window
- **guitar and bass player stand**
  - [guitar, NN, and, CC, player, NN, stand, VB]
  - Word_{n-2}, POS_{n-2}, word_{n-1}, POS_{n-1}, Word_{n+1}, POS_{n+1}…
  - In other words, a vector consisting of
  - [position n word, position n part-of-speech…]

Bag of Words

- Information about what words 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 in “bass” sentences:

\[
[\text{fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band}]
\]

• The vector for:

\[
\text{guitar and bass player stand}
\]

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

- Once we cast the WSD problem as a classification problem, many techniques possible
  - Naïve Bayes
  - Decision lists
  - Decision trees
  - Neural nets
  - Support vector machines
  - Nearest neighbor methods…

Classifiers

- 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
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 \( C = \{ c_1, c_2, \ldots, c_J \} \)
  - A training set of \( m \) hand-labeled text windows again called “documents” \( (d_1, c_1), \ldots, (d_m, c_m) \)

- **Output:**
  - a learned classifier \( \gamma: d \rightarrow c \)

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 the maximum likelihood estimate from a sense-tagged corpus (count($s_i$, $w_j$)/count($w_j$)) – how likely is bank to mean ‘financial institution’ over all instances of bank?

- $P(v_j | s)$ is the maximum likelihood of each feature given a candidate sense (count($v_j$, $s$)/count($s$)) – how likely is the previous word to be ‘river’ when the sense of bank is ‘financial institution’?

- Calculate $\hat{s} = \arg\max_{s} p(s) \prod_{j=1}^{n} p(v_j | s)$ for each possible sense and take the highest scoring sense as the most likely choice.
Naïve Bayes Evaluation

- On a corpus of examples of uses of the word line, naïve Bayes achieved about 73% correct
- Is this good?

\[
\hat{P}(w | c) = \frac{\text{count}(w, c) + 1}{\text{count}(c) + |V|}
\]

Priors:
- \( P(f) = \frac{3}{4} \)
- \( P(g) = \frac{1}{4} \)

Conditional Probabilities:
- \( P(\text{line} | f) = \frac{1+1}{8+6} = 2/14 \)
- \( P(\text{guitar} | f) = \frac{0+1}{8+6} = 1/14 \)
- \( P(\text{jazz} | f) = \frac{0+1}{8+6} = 1/14 \)
- \( P(\text{line} | g) = \frac{1+1}{3+6} = 2/9 \)
- \( P(\text{guitar} | g) = \frac{1+1}{3+6} = 2/9 \)
- \( P(\text{jazz} | g) = \frac{1+1}{3+6} = 2/9 \)

Choosing a class:
- \( P(f | d_5) = \frac{3}{4} * \frac{2}{14} * \left(\frac{1}{14}\right)^2 * \frac{1}{14} = 0.00003 \)
- \( P(g | d_5) = \frac{1}{4} * \frac{2}{9} * \left(\frac{2}{9}\right)^2 * \frac{2}{9} = 0.00006 \)
Decision Lists

- Can be treated as a case statement.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Sense</th>
</tr>
</thead>
<tbody>
<tr>
<td>fish within window</td>
<td>bass¹</td>
</tr>
<tr>
<td>striped bass</td>
<td>bass¹</td>
</tr>
<tr>
<td>guitar within window</td>
<td>bass²</td>
</tr>
<tr>
<td>bass player</td>
<td>bass²</td>
</tr>
<tr>
<td>piano within window</td>
<td>bass²</td>
</tr>
<tr>
<td>tenor within window</td>
<td>bass²</td>
</tr>
<tr>
<td>sea bass</td>
<td>bass¹</td>
</tr>
<tr>
<td>play/N bass</td>
<td>bass²</td>
</tr>
<tr>
<td>river within window</td>
<td>bass¹</td>
</tr>
<tr>
<td>violin within window</td>
<td>bass²</td>
</tr>
<tr>
<td>salmon within window</td>
<td>bass¹</td>
</tr>
<tr>
<td>on bass</td>
<td>bass²</td>
</tr>
<tr>
<td>bass are</td>
<td>bass¹</td>
</tr>
</tbody>
</table>

Learning Decision Lists

- Restrict lists to rules that test a single feature
- Evaluate each possible test and rank them based on how well they work
- Order the top-N tests as the decision list
Yarowsky’s Metric

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

$$\log \left( \frac{P(\text{Sense}_1 \mid \text{Feature})}{P(\text{Sense}_2 \mid \text{Feature})} \right)$$

- This gives about 95% on this test…

WSD Evaluations and Baselines

- *In vivo* (intrinsic) versus *in vitro* (extrinsic) evaluation
- In vitro evaluation 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, Lesk algorithm
Most Frequent Sense

- Wordnet senses are ordered in frequency order
- So “most frequent sense” in WordNet = “take the first sense”
- Sense frequencies come from SemCor

<table>
<thead>
<tr>
<th>Freq</th>
<th>Synset</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>338</td>
<td>plant¹, works, industrial plant</td>
<td>buildings for carrying on industrial labor</td>
</tr>
<tr>
<td>207</td>
<td>plant², flora, plant life</td>
<td>a living organism lacking the power of locomotion</td>
</tr>
<tr>
<td>2</td>
<td>plant³</td>
<td>something planted secretly for discovery by another</td>
</tr>
<tr>
<td>0</td>
<td>plant⁴</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: Dictionary/Thesaurus Methods

- The Lesk Algorithm
- Selectional Restrictions

Simplified Lesk

- Match dictionary entry of sense that best matches context

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>bank¹</th>
<th>Gloss:</th>
<th>a financial institution that accepts deposits and channels the money into lending activities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Examples:</td>
<td>“he cashed a check at the bank”, “that bank holds the mortgage on my home”</td>
</tr>
<tr>
<td>bank²</td>
<td>Gloss:</td>
<td>sloping land (especially the slope beside a body of water)</td>
</tr>
<tr>
<td></td>
<td>Examples:</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>
Simplified Lesk

- Match dictionary entry of sense that best matches context: bank1 (deposits, mortgage)

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>bank¹</th>
<th>Gloss:</th>
<th>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”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Examples:</td>
<td>“he cashed a check at the bank”, “that bank holds the mortgage on my home”</td>
</tr>
<tr>
<td>bank²</td>
<td>Gloss:</td>
<td>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”</td>
</tr>
<tr>
<td></td>
<td>Examples:</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

- Compare entries for each context word for overlap

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
Original Lesk: pine cone

- Compare entries for each context word for overlap
- Cone3 selected: evergreen, tree

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

“Time flies like an arrow” – what are the correct senses?

- time#n#5 (the continuum of experience in which events pass from the future through the present to the past)
- time#v#1 (measure the time or duration of an event or action or the person who performs an action in a certain period of time) “he clocked the runners”
- flies#n#1 (two-winged insects characterized by active flight)
- flies#v#8 (pass away rapidly) “Time flies like an arrow”; “Time fleeing beneath him”
- like#v#4 (feel about or towards; consider, evaluate, or regard) “How did you like the President’s speech last night?”
- like#a#1 (resembling or similar; having the same or some of the same characteristics; often used in combination) “suits of like design”; “a limited circle of likeminds”; “members of the cat family have like dispositions”; “as like as two peas in a pod”; “doglike devotion”; “a dreamlike quality”
Try the original algorithm on "Time flies like an arrow" using WordNet senses below. Assume that the words are to be disambiguated one at a time, from left to right, and that the results from earlier decisions are used later in the process.

<table>
<thead>
<tr>
<th>Sense</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>time#n#5</td>
<td>the continuum of experience in which events pass from the future through the present to the past</td>
</tr>
<tr>
<td>time#v#1</td>
<td>measure the time or duration of an event or action or the person who performs an action in a certain period of time</td>
</tr>
<tr>
<td>flies#n#1</td>
<td>two-winged insects characterized by active flight</td>
</tr>
<tr>
<td>flies#v#8</td>
<td>(pass away rapidly) &quot;Time flies like an arrow&quot;; &quot;Time fleeing beneath him&quot;</td>
</tr>
<tr>
<td>like#v#4</td>
<td>(feel about or towards; consider, evaluate, or regard) &quot;How did you like the President’s speech last night?&quot;</td>
</tr>
<tr>
<td>like#a#1</td>
<td>(resembling or similar; having the same or some of the same characteristics; often used in combination) “suits of like design”; “a limited circle of likeminds”; “members of the cat family have like dispositions”; “as like as two peas in a pod”; “doglike devotion”; “a dreamlike quality&quot;</td>
</tr>
</tbody>
</table>

"Time flies like an arrow"

<table>
<thead>
<tr>
<th>Sense</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>time#n#5</td>
<td>(the continuum of experience in which events pass from the future through the present to the past)</td>
</tr>
<tr>
<td>time#v#1</td>
<td>(measure the time or duration of an event or action or the person who performs an action in a certain period of time) &quot;he clocked the runners&quot;</td>
</tr>
<tr>
<td>flies#n#1</td>
<td>(two-winged insects characterized by active flight)</td>
</tr>
<tr>
<td>flies#v#8</td>
<td>(pass away rapidly) &quot;Time flies like an arrow&quot;; &quot;Time fleeing beneath him&quot;</td>
</tr>
<tr>
<td>like#v#4</td>
<td>(feel about or towards; consider, evaluate, or regard) &quot;How did you like the President’s speech last night?&quot;</td>
</tr>
<tr>
<td>like#a#1</td>
<td>(resembling or similar; having the same or some of the same characteristics; often used in combination) “suits of like design”; “a limited circle of likeminds”; “members of the cat family have like dispositions”; “as like as two peas in a pod”; “doglike devotion”; “a dreamlike quality&quot;</td>
</tr>
</tbody>
</table>

**Time WSD**: tie, backoff to most frequent, but can’t because POS differ
“Time flies like an arrow”

- **time**
  - n#5 (the continuum of experience in which events pass from the future through the present to the past)
  - v#1 (measure the time or duration of an event or action or the person who performs an action in a certain period of time) “he clocked the runners”

- **flies**
  - n#1 (two-winged insects characterized by active flight)
  - v#8 (pass away rapidly) “Time flies like an arrow”; “Time fleeing beneath him”

- **like**
  - v#4 (feel about or towards; consider, evaluate, or regard) “How did you like the President’s speech last night?”
  - a#1 (resembling or similar; having the same or some of the same characteristics; often used in combination) “suits of like design”; “a limited circle of likeminds”; “members of the cat family have like dispositions”; “as like as two peas in a pod”; “doglike devotion”; “a dreamlike quality”

- Flies WSD: select verb

---

**Corpus Lesk**

- Add corpus examples to glosses and examples
- The best performing variant
The Corpus Lesk algorithm

- Assumes we have some sense-labeled data (like SemCor)
- Take all the sentences with the relevant word sense:
  These short, "streamlined" meetings usually are sponsored by local banks, Chambers of Commerce, trade associations, or other civic organizations.
- Now add these to the gloss + examples for each sense, call it the “signature” of a sense.
- Choose sense with most word overlap between context and signature.

Corpus Lesk: IDF weighting

- Instead of just removing function words
  - Weigh each word by its `promiscuity` across documents
  - Down-weights words that occur in every `document` (gloss, example, etc)
  - These are generally function words, but is a more fine-grained measure
- Weigh each overlapping word by inverse document frequency
Graph-based methods

- First, WordNet can be viewed as a graph
  - senses are nodes
  - relations (hypernymy, meronymy) are edges
  - Also add edge between word and unambiguous gloss words

How to use the graph for WSD

- Insert target word and words in its sentential context into the graph, with directed edges to their senses

“She drank some milk”

- Now choose the **most central sense**

Add some probability to “drink” and “milk” and compute node with highest “pagerank”
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 --> V NP
    - V → serve <theme> {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!*)
- Resnik 1988: 44% with traditional methods
- Can we take a statistical approach?
Semi-Supervised Bootstrapping

- What if you don’t have enough data or hand-built resources 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
  - Generalize from a small hand-labeled seed set

Bootstrapping

- For *bass*
  - Assume *play* occurs with the music sense and *fish* occurs with the fish sense
Sentences Extracts for **bass** and **player**

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>We need more good teachers – right now, there are only a half dozen who can <strong>play</strong> the free <strong>bass</strong> with ease.</td>
<td></td>
</tr>
<tr>
<td>An electric guitar and <strong>bass player</strong> stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.</td>
<td></td>
</tr>
<tr>
<td>When the New Jersey Jazz Society, in a fund-raiser for the American Jazz Hall of Fame, honors this historic night next Saturday, Harry Goodman, Mr. Goodman’s brother and <strong>bass player</strong> at the original concert, will be in the audience with other family members.</td>
<td></td>
</tr>
<tr>
<td>The researchers said the worms spend part of their life cycle in such <strong>fish</strong> as Pacific salmon and striped <strong>bass</strong> and Pacific rockfish or snapper.</td>
<td></td>
</tr>
<tr>
<td>And it all started when fishermen decided the striped <strong>bass</strong> in Lake Mead were too skinny. <strong>Though still a far cry from the lake’s record 52-pound</strong> <strong>bass</strong> of a decade ago, “you could fillet these <strong>fish</strong> again, and that made people very, very happy,” Mr. Paulson says.</td>
<td></td>
</tr>
</tbody>
</table>

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” not true at all once you move to fine-grained senses

3) **One sense per collocation:**
   - A word recurring in collocation with the same word will almost surely have the same sense
Stages in Yarowsky Bootstrapping Algorithm

Issues

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

History of Senseval

- ACL-SIGLEX workshop (1997)
- SENSEVAL-I (1998)
  - Lexical Sample for English, French, and Italian
- SENSEVAL-II (Toulouse, 2001)
  - Lexical Sample and All Words
- SENSEVAL-IV -> SEMEVAL (2007)
- Newer: SEM, First Joint Conference on Lexical and Computational Semantics
2012

• SEM: 1st Conf. on Lexical & Computational Semantics
• SemEval: International Workshop on Semantic Evaluations
  – 1. English Lexical Simplification
  – 2. Measuring Degrees of Relational Similarity
  – 3. Spatial Role Labeling
  – 4. Evaluating Chinese Word Similarity
  – 5. Chinese Semantic Dependency Parsing
  – 6. Semantic Textual Similarity
  – 7. COPA: Choice Of Plausible Alternatives
    An evaluation of common-sense causal reasoning
  – 8. Cross-lingual Textual Entailment for Content Synchronization

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

• 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