Lecture 3:
Word Sense and Lexical Semantics
• All materials are posted on the course website:
  ttic.uchicago.edu/~kgimpel/teaching/31190-s18/index.html

• Assignment 1 due 6:00 pm on Wed., April 11th
Roadmap

• words, morphology, lexical semantics
• text classification
• simple neural methods for NLP
• language modeling and word embeddings
• recurrent/recursive/convolutional networks in NLP
• sequence labeling, HMMs, dynamic programming
• syntax and syntactic parsing
• semantics, compositionality, semantic parsing
• machine translation and other NLP tasks
Words

• types and tokens
• morphology
• distributional word vectors
• word sense and lexical semantics
Counting Context Words

sugar, a sliced lemon, a tablespoonful of their enjoyment. Cautiously she sampled her first well suited to programming on the digital for the purpose of gathering data and apricot pineapple computer. information preserve or jam, a pinch each of, and another fruit whose taste she likened In finding the optimal R-stage policy from necessary for the study authorized in the

<table>
<thead>
<tr>
<th>aardvark</th>
<th>computer</th>
<th>data</th>
<th>pinch</th>
<th>result</th>
<th>sugar</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>...</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>...</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>...</td>
</tr>
</tbody>
</table>

Figure 19.2 Co-occurrence vectors for four words, computed from the Brown corpus, showing only six of the dimensions (hand-picked for pedagogical purposes). Note that a real vector would be vastly more sparse.

The shading in Fig. 17.2 makes clear the intuition that the two words apricot and pineapple are more similar (both pinch and sugar tend to occur in their window) while digital and information are more similar. Note that |V|, the length of the vector, is generally the size of the vocabulary, usually between 10,000 and 50,000 words (using the most frequent words in the J&M/SLP3
Handling Sentence Boundaries

<s> these symptoms do not imply sickness . </s>

- pad sentences with <s> at the beginning and </s> at the end
- these are context words just like any others
- when context window goes outside the sentence, you can use a single padding symbol or enough to fill the window
  - (I used a single padding symbol at each end)
Cosine Similarity

- dot product divided by vector length product

\[
\frac{u^\top v}{\|u\| \|v\|}
\]

- cosine of angle between the vectors
<table>
<thead>
<tr>
<th>Context words of “cooked” with largest counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>123 ,</td>
</tr>
<tr>
<td>92 and</td>
</tr>
<tr>
<td>79 the</td>
</tr>
<tr>
<td>71 .</td>
</tr>
<tr>
<td>68 &lt;s&gt;</td>
</tr>
<tr>
<td>66 &lt;/s&gt;</td>
</tr>
<tr>
<td>53 in</td>
</tr>
<tr>
<td>39 a</td>
</tr>
<tr>
<td>38 is</td>
</tr>
<tr>
<td>35 of</td>
</tr>
<tr>
<td>30 with</td>
</tr>
<tr>
<td>28 are</td>
</tr>
<tr>
<td>25 to</td>
</tr>
<tr>
<td>23 or</td>
</tr>
<tr>
<td>23 it</td>
</tr>
<tr>
<td>20 (</td>
</tr>
<tr>
<td>19 be</td>
</tr>
<tr>
<td>15 )</td>
</tr>
<tr>
<td>14 &quot;</td>
</tr>
</tbody>
</table>
Pointwise Mutual Information (PMI)

• do two events $x$ and $y$ co-occur more often than if they were independent?

$$\text{pmi}(x; y) = \log_2 \frac{p(x, y)}{p(x)p(y)}$$

• replace raw counts with pmi scores
<table>
<thead>
<tr>
<th>PMI</th>
<th>Word</th>
<th>PMI</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.30533</td>
<td>beef</td>
<td>7.66406</td>
<td>chili</td>
</tr>
<tr>
<td>8.88418</td>
<td>shrimp</td>
<td>7.56264</td>
<td>rice</td>
</tr>
<tr>
<td>8.63397</td>
<td>potatoes</td>
<td>7.56167</td>
<td>soup</td>
</tr>
<tr>
<td>8.61946</td>
<td>ate</td>
<td>7.45315</td>
<td>flour</td>
</tr>
<tr>
<td>8.56584</td>
<td>dishes</td>
<td>7.43874</td>
<td>steamed</td>
</tr>
<tr>
<td>8.50945</td>
<td>eaten</td>
<td>7.43715</td>
<td>crushed</td>
</tr>
<tr>
<td>8.4931</td>
<td>beans</td>
<td>7.41193</td>
<td>meals</td>
</tr>
<tr>
<td>8.33137</td>
<td>texture</td>
<td>7.39793</td>
<td>digest</td>
</tr>
<tr>
<td>8.29489</td>
<td>vegetables</td>
<td>7.39175</td>
<td>rockies</td>
</tr>
<tr>
<td>8.25088</td>
<td>soda</td>
<td>7.34773</td>
<td>ramsay</td>
</tr>
<tr>
<td>8.20831</td>
<td>meat</td>
<td>7.33211</td>
<td>honey</td>
</tr>
<tr>
<td>8.15708</td>
<td>sauce</td>
<td>7.32253</td>
<td>toxicity</td>
</tr>
<tr>
<td>8.08345</td>
<td>consuming</td>
<td>7.29057</td>
<td>cared</td>
</tr>
<tr>
<td>7.9532</td>
<td>cuisine</td>
<td>7.28626</td>
<td>tomatoes</td>
</tr>
<tr>
<td>7.94043</td>
<td>raw</td>
<td>7.27912</td>
<td>boiling</td>
</tr>
<tr>
<td>7.78435</td>
<td>curry</td>
<td>7.27769</td>
<td>dal</td>
</tr>
<tr>
<td>7.7563</td>
<td>juice</td>
<td>7.27485</td>
<td>citrus</td>
</tr>
<tr>
<td>7.74444</td>
<td>vegetable</td>
<td>7.25649</td>
<td>doncaster</td>
</tr>
</tbody>
</table>
How should we evaluate word vectors?
### WordSim353
(Finkelstein et al., 2002)

<table>
<thead>
<tr>
<th>word pair</th>
<th>similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>journey</td>
<td>voyage</td>
</tr>
<tr>
<td>king</td>
<td>queen</td>
</tr>
<tr>
<td>computer</td>
<td>software</td>
</tr>
<tr>
<td>law</td>
<td>lawyer</td>
</tr>
<tr>
<td>forest</td>
<td>graveyard</td>
</tr>
<tr>
<td>rooster</td>
<td>voyage</td>
</tr>
</tbody>
</table>
WordSim353  
(Finkelstein et al., 2002)

**Instructions:**

Assign a numerical similarity score between 0 and 10 (0 = words are totally unrelated, 10 = words are VERY closely related).

<table>
<thead>
<tr>
<th>computer</th>
<th>software</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>law</td>
<td>lawyer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>forest</td>
<td>graveyard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rooster</td>
<td>voyage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Instructions:

Assign a numerical similarity score between 0 and 10 (0 = words are totally unrelated, 10 = words are VERY closely related).

When estimating similarity of antonyms, consider them "similar" (i.e., belonging to the same domain or representing features of the same concept), rather than "dissimilar".

<table>
<thead>
<tr>
<th>forest</th>
<th>graveyard</th>
</tr>
</thead>
<tbody>
<tr>
<td>rooster</td>
<td>voyage</td>
</tr>
<tr>
<td>word pair</td>
<td>similarity</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------</td>
</tr>
<tr>
<td>journey voyage</td>
<td>9.3</td>
</tr>
<tr>
<td>king queen</td>
<td>8.6</td>
</tr>
<tr>
<td>computer software</td>
<td>8.5</td>
</tr>
<tr>
<td>law lawyer</td>
<td>8.4</td>
</tr>
<tr>
<td>forest graveyard</td>
<td>1.9</td>
</tr>
<tr>
<td>rooster voyage</td>
<td>0.6</td>
</tr>
</tbody>
</table>
SimLex-999  
(Hill et al., 2014)

<table>
<thead>
<tr>
<th>word pair</th>
<th>similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>insane</td>
<td>9.6</td>
</tr>
<tr>
<td>attorney</td>
<td>9.4</td>
</tr>
<tr>
<td>author</td>
<td>8.0</td>
</tr>
<tr>
<td>diet</td>
<td>1.2</td>
</tr>
<tr>
<td>new</td>
<td>0.2</td>
</tr>
</tbody>
</table>

measures **paraphrastic** similarity:  
two words are “similar” if they have similar meanings
there are many word similarity datasets
some focus on topical relatedness, others focus on similarity in meaning
in assignment 1, you will evaluate your word vectors using MEN (relatedness) and SimLex-999 (meaning)
Evaluation Metrics for Word Similarity

• Spearman rank correlation coefficient

measures correlation between two variables:
  – variable 1: human-annotated similarities for word pairs
  – variable 2: cosine similarities computed with your word vectors for the same word pairs
Sparse versus dense vectors

• so far, our vectors are
  – long (length = 25,000)
  – sparse (mostly zero)

• why might we want to reduce vector dimensionality?
Why reduce dimensionality?

• short vectors may be easier to use as features (fewer weights to tune)
• reducing dimensionality may better handle variability in natural language due to synonymy:
  – *car* and *automobile* are synonyms, but are distinct dimensions
  – fails to capture similarity between a word with *car* as a neighbor and one with *automobile* as a neighbor
Dimensionality Reduction: Intuition

• approximate an $N$-dimensional dataset using fewer dimensions:
  – rotate axes into a new space
  – in which first dimension captures most variance in original dataset

• many such (related) methods:
  – principal component analysis (PCA)
  – factor analysis
  – singular value decomposition (SVD)
Dimensionality reduction

PCA dimension 1

PCA dimension 2
SVD embeddings versus sparse vectors

• dense SVD embeddings sometimes work better than sparse PMI vectors at tasks (like word similarity)
  – denoising: low-order dimensions may represent unimportant information
  – truncation may help the models generalize better to unseen data
  – smaller number of dimensions may make it easier for classifiers to effectively assign weights to dimensions for the task
  – dense models may do better at capturing higher order co-occurrence
Words

• types and tokens
• morphology
• distributional word vectors
• word sense and lexical semantics
Word Sense Ambiguity

• many words have multiple meanings
Word Sense Ambiguity
Word Sense Ambiguity

credit: A. Zwicky
Terminology: lemma and wordform

• **lemma**
  – words with same lemma have same stem, part of speech, rough semantics

• **wordform**
  – 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*$_1$ can hold the investments in a custodial account
  sense 2: ...as agriculture burgeons on the east *bank*$_2$ 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
• two ways to categorize the patterns of multiple meanings of words:
  – homonymy: the multiple meanings are unrelated (coincidental?)
  – polysemy: the multiple meanings are related
Homonymy

**homonyms**: words that share a form but have unrelated, distinct meanings:

- $\textit{bank}_1$: financial institution  \quad $\textit{bank}_2$: sloping land
- $\textit{bat}_1$: club for hitting a ball  \quad $\textit{bat}_2$: nocturnal flying mammal

**homographs**: same spelling, different meanings

$\textit{bank}/\textit{bank}$, $\textit{bat}/\textit{bat}$

**homophones**: same pronunciation, different meanings

$\textit{write}/\textit{right}$, $\textit{piece}/\textit{peace}$
Homonymy causes problems for NLP

• information retrieval
  – query: *bat care*

• machine translation
  – *bat*: *murciélago* (animal) or *bate* (for baseball)

• text-to-speech
  – *bass* (stringed instrument) vs. *bass* (fish)
Polysemy

1: *The bank* was constructed in 1875 out of local red brick.

2: *I withdrew the money from the bank.*

- are these the same sense?
  - sense 2: “a financial institution”
  - sense 1: “the building belonging to a financial institution”

- a **polysemous** word has related meanings
  - most non-rare words have multiple related meanings
Homonymy or Polysemy?

an edge tool with a heavy bladed head mounted across a handle

a fixed reference line for the measurement of coordinates

axes
Homonymy or Polysemy?

an edge tool with a heavy bladed head mounted across a handle

a fixed reference line for the measurement of coordinates
Homonymy or Polysemy?

an imaginary line about which a body rotates

a fixed reference line for the measurement of coordinates
Homonymy or Polysemy?

axes

an imaginary line about which a body rotates

a fixed reference line for the measurement of coordinates
Homonymy or Polysemy?

in an inactive or inoperative state

being or moving lower in position or less in some value

down
Homonymy or Polysemy?

down

in an inactive or inoperative state

being or moving lower in position or less in some value
Homonymy or Polysemy?

soft fine feathers

being or moving lower in position or less in some value

down

down
Homonymy or Polysemy?

soft fine feathers

being or moving lower in position or less in some value

donward
Homonymy or Polysemy?

- **down**
  - in an inactive or inoperative state
  - being or moving lower in position or less in some value
  - unhappy

What is the meaning of life?
I don't know. The computers are down.
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?

• “zeugma” test: two senses of *serve*?
  – *Which flights serve breakfast?*
  – *Does Lufthansa serve Philadelphia?*
  – *Does Lufthansa serve breakfast and Philadelphia?*

• since this conjunction sounds weird, we say that these are **two different senses of serve**
Synonyms

• words with same meaning in some or all contexts:
  – filbert / hazelnut
  – couch / sofa
  – big / large
  – water / $H_2O$

• two lexemes are synonyms if they can be substituted for each other in all situations
Synonyms

• 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.

• examples:
  – 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** ("hyper is super")
  – *vehicle* is a hypernym of *car*
  – *fruit* is a hypernym of *mango*
Meronymy/Holonymy

• part-whole relation
  – *wheel* is a **meronym** of *car*
  – *car* is a **holonym** of *wheel*
WordNet 3.0

- hierarchically organized lexical database
- on-line thesaurus + aspects of a dictionary
  - some languages available or under development: Arabic, Finnish, German, Portuguese...

<table>
<thead>
<tr>
<th>Category</th>
<th>Unique Strings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>117,798</td>
</tr>
<tr>
<td>Verb</td>
<td>11,529</td>
</tr>
<tr>
<td>Adjective</td>
<td>22,479</td>
</tr>
<tr>
<td>Adverb</td>
<td>4,481</td>
</tr>
</tbody>
</table>
Senses of *bass* in WordNet

**Noun**

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

**Adjective**

- **S:** (adj) *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?

• **synset (synonym set):** set of near-synonyms; instantiates a sense or concept, with a gloss

• example: *chump* as a noun with gloss:
  “a person who is gullible and easy to take advantage of”

• this sense of *chump* is shared by 9 words:
  *chump*¹, *fool*², *gull*¹, *mark*⁹, *patsy*¹, *fall guy*¹, *sucker*¹, *soft touch*¹, *mug*²

• each of **these** senses have this same gloss
  – (not *every* sense; sense 2 of *gull* is the aquatic bird)
Noun

- **S**: (n) fool, sap, saphead, muggins, tomfool (a person who lacks good judgment)
- **S**: (n) chump, fool, gull, mark, patsy, fall guy, sucker, soft touch, mug (a person who is gullible and easy to take advantage of)
- **S**: (n) jester, fool, motley fool (a professional clown employed to entertain a king or nobleman in the Middle Ages)

**ambiguity**

- one form, multiple meanings → split form
  - the three senses of *fool* belong to different synsets

**variability**

- multiple forms, one meaning → merge forms
  - each synset contains senses of several different words
WordNet Hypernym Hierarchy for bass

(n) bass, basso (an adult male singer with the lowest voice)

- **direct hypernym** / **inherited hypernym** / **sister term**
  - S: (n) singer, vocalist, vocalizer, vocaliser (a person who sings)
    - S: (n) performer, performing artist (an entertainer who performs a dramatic or musical work for an audience)
    - S: (n) entertainer (a person who tries to please or amuse)
      - S: (n) person, individual, someone, somebody, mortal, soul (a human being) "there was too much person to do"
      - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
  - S: (n) living thing, animate thing (a living (or once living) entity)
    - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) "part compared to the whole?"; "the team is a unit"
      - S: (n) object, physical object (a tangible and visible entity; an entity that casts a shadow) "it was full of rackets, balls and other objects"
    - S: (n) physical entity (an entity that has physical existence)
      - S: (n) entity (that which is perceived or known or inferred to have distinct existence (living or nonliving))
### Supersenses: top level hypernyms in hierarchy

(counts from Schneider & Smith’s Streusel corpus)

<table>
<thead>
<tr>
<th>Noun</th>
<th>Count</th>
<th>Examples</th>
<th>Count</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP</td>
<td>1469</td>
<td>place</td>
<td>87</td>
<td>hair</td>
</tr>
<tr>
<td>PERSON</td>
<td>1202</td>
<td>people</td>
<td>56</td>
<td>pain</td>
</tr>
<tr>
<td>ARTIFACT</td>
<td>971</td>
<td>car</td>
<td>54</td>
<td>flower</td>
</tr>
<tr>
<td>COGNITION</td>
<td>771</td>
<td>way</td>
<td>35</td>
<td>portion</td>
</tr>
<tr>
<td>FOOD</td>
<td>766</td>
<td>food</td>
<td>34</td>
<td>oil</td>
</tr>
<tr>
<td>ACT</td>
<td>700</td>
<td>service</td>
<td>34</td>
<td>discomfort</td>
</tr>
<tr>
<td>LOCATION</td>
<td>638</td>
<td>area</td>
<td>28</td>
<td>process</td>
</tr>
<tr>
<td>TIME</td>
<td>530</td>
<td>day</td>
<td>25</td>
<td>reason</td>
</tr>
<tr>
<td>EVENT</td>
<td>431</td>
<td>experience</td>
<td>23</td>
<td>result</td>
</tr>
<tr>
<td>COMMUNIC.*</td>
<td>417</td>
<td>review</td>
<td>6</td>
<td>square</td>
</tr>
<tr>
<td>POSSESSION</td>
<td>339</td>
<td>price</td>
<td>5</td>
<td>tree</td>
</tr>
<tr>
<td>ATTRIBUTE</td>
<td>205</td>
<td>quality</td>
<td>2</td>
<td>stuff</td>
</tr>
<tr>
<td>QUANTITY</td>
<td>102</td>
<td>amount</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANIMAL</td>
<td>88</td>
<td>dog</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Verb</th>
<th>Count</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>STATIVE</td>
<td>2922</td>
<td>is</td>
</tr>
<tr>
<td>COGNITION</td>
<td>1093</td>
<td>know</td>
</tr>
<tr>
<td>COMMUNIC.*</td>
<td>974</td>
<td>recommend</td>
</tr>
<tr>
<td>SOCIAL</td>
<td>944</td>
<td>use</td>
</tr>
<tr>
<td>MOTION</td>
<td>602</td>
<td>go</td>
</tr>
<tr>
<td>POSSESSION</td>
<td>309</td>
<td>pay</td>
</tr>
<tr>
<td>CHANGE</td>
<td>274</td>
<td>fix</td>
</tr>
<tr>
<td>EMOTION</td>
<td>249</td>
<td>love</td>
</tr>
<tr>
<td>PERCEPTION</td>
<td>143</td>
<td>see</td>
</tr>
<tr>
<td>CONSUMPTION</td>
<td>93</td>
<td>have</td>
</tr>
<tr>
<td>BODY</td>
<td>82</td>
<td>get...done</td>
</tr>
<tr>
<td>CREATION</td>
<td>64</td>
<td>cook</td>
</tr>
<tr>
<td>CONTACT</td>
<td>46</td>
<td>put</td>
</tr>
<tr>
<td>COMPETITION</td>
<td>11</td>
<td>win</td>
</tr>
<tr>
<td>WEATHER</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
WordNet: Viewed as a graph

We note that each word sense univocally identifies a single synset. For instance, given \textit{car}\textsubscript{1n} the corresponding synset \{\textit{car}\textsubscript{1n}, \textit{auto}\textsubscript{1n}, \textit{automobile}\textsubscript{1n}, \textit{machine}\textsubscript{4n}, \textit{motorcar}\textsubscript{1n}\} is univocally determined. In Figure 3 we report an excerpt of the WordNet semantic network containing the \textit{car}\textsubscript{1n} synset. For each synset, WordNet provides the following information:

- A gloss, that is, a textual definition of the synset possibly with a set of usage examples (e.g., the gloss of \textit{car}\textsubscript{1n} is “a 4-wheeled motor vehicle; usually propelled by an internal combustion engine; ‘he needs a car to get to work’ ”).

- Lexical and semantic relations, which connect pairs of word senses and synsets, respectively: while semantic relations apply to synsets in their entirety (i.e., to all members of a synset), lexical relations connect word senses included in the respective synsets. Among the latter we have the following:
  - Antonymy: \textit{X} is an antonym of \textit{Y} if it expresses the opposite concept (e.g., \textit{good}\textsubscript{1a} is the antonym of \textit{bad}\textsubscript{1a}). Antonymy holds for all parts of speech.
  - Pertainymy: \textit{X} is an adjective which can be defined as “of or pertaining to” a noun (or, rarely, another adjective) (e.g., \textit{dental}\textsubscript{1a} pertains to \textit{tooth}\textsubscript{1n}).
  - Nominalization: a noun \textit{X} nominalizes a verb \textit{Y} (e.g., \textit{service}\textsubscript{2n} nominalizes the verb \textit{serve}\textsubscript{4v}).

Among the semantic relations we have the following:

- Hypernymy (also called kind-of or is-a): \textit{Y} is a hypernym of \textit{X} if every \textit{X} is a (kind of) \textit{Y} (e.g., \textit{motor vehicle}\textsubscript{1n} is a hypernym of \textit{car}\textsubscript{1n}). Hypernymy holds between pairs of nominal or verbal synsets.

is a (hyponym/hypernym/meronym/holonym) of
is a (hyponym/hypernym/meronym/holonym) of
piano₁

is a

(hyponym/hypernym)

of

instrument₁
piano$_1$ is a (hyponym/hypernym) of instrument$_1$
Word Sense Disambiguation (WSD)

- given:
  - a word in context
  - a fixed inventory of potential word senses
- decide which sense of the word this is
- why? machine translation, question answering, sentiment analysis, text-to-speech
- what set of senses?
  - English-to-Spanish machine translation: set of Spanish translations
  - text-to-speech: homographs like \textit{bass} and \textit{bow}
  - in general: the senses in a thesaurus like WordNet
Two Variants of WSD Task

• lexical sample task
  – small pre-selected set of target words (*line*, *plant*, *bass*)
  – inventory of senses for each word
  – supervised 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
8 Senses of *bass* in WordNet

**Noun**

- **S: (n) bass** (the lowest part of the musical range)
- **S: (n) bass, bass part** (the lowest part in polyphonic music)
- **S: (n) bass, basso** (an adult male singer with the lowest voice)
- **S: (n) sea bass, bass** (the lean flesh of a saltwater fish of the family Serranidae)
- **S: (n) freshwater bass, bass** (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
- **S: (n) bass, bass voice, basso** (the lowest adult male singing voice)
- **S: (n) bass** (the member with the lowest range of a family of musical instruments)
- **S: (n) 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 bass and…</td>
</tr>
<tr>
<td>bass(^4)</td>
<td>lubina</td>
<td>FISH/INSECT</td>
<td>… produce filets of smoked bass or sturgeon…</td>
</tr>
<tr>
<td>bass(^7)</td>
<td>bajo</td>
<td>MUSIC</td>
<td>… exciting jazz bass player since Ray Brown…</td>
</tr>
<tr>
<td>bass(^7)</td>
<td>bajo</td>
<td>MUSIC</td>
<td>… play bass because he doesn’t have to solo…</td>
</tr>
</tbody>
</table>
WSD Evaluation and Baselines

• best evaluation: **extrinsic (“task-based”)**
  – embed WSD in a task and see if it helps!

• **intrinsic** evaluation often done for convenience

• strong baseline: most frequent sense
Most Frequent Sense

• WordNet senses are ordered by frequency
• most frequent is first
• sense frequencies come from *SemCor* corpus

<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>
Performance 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%
Training Data for WSD

• **semantic concordance**: 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
Features for WSD?
Features for WSD?
Intuition from Warren 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... 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...

‘What minimum value of N will... lead to the correct choice of meaning for the central word?’”
Example

- using a window of +/- 3 from the target:

  An *electric guitar and bass player stand off* to one side not really part of the scene
Semi-Supervised Learning

**problem:** supervised learning requires large hand-built resources

  what if you don’t have much training data?

**solution:** bootstrapping

  generalize from a very small hand-labeled seed set
Bootstrapping

• “one sense per collocation” heuristic:
  – a word reoccurring in collocation with the same word will almost surely have the same sense

• For *bass*:
  – *play* occurs with the music sense of *bass*
  – *fish* occurs with the fish sense of *bass*
We need more good teachers – right now, there are only a half a dozen who can **play** the free **bass** with ease.

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.

The researchers said the worms spend part of their life cycle in such **fish** as Pacific salmon and striped **bass** and Pacific rockfish or snapper.

And it all started when **fishermen** decided the striped **bass** in Lake Mead were too skinny.
Bootstrapping

• “one sense per collocation” heuristic:
  – a word reoccurring in collocation with the same word will almost surely have the same sense

• “one sense per discourse” heuristic:
  – sense of a word is highly consistent within a document (Yarowsky, 1995)
  – especially topic-specific words
Stages in Yarowsky bootstrapping algorithm for plant

(a)

(b)
Exceptions

• Exceptions to one sense per collocation?
  – wedding band

• Exceptions to one sense per discourse?
  – “I’m going to rest for the rest of the day.”
  – “Last year was his last year.”
  – “Those plants generate so much pollution that no plants grow within a hundred feet.”
Summary

• word sense disambiguation: choosing correct sense in context
• applications: MT, QA, etc.
• main intuition:
  – lots of information in a word’s context
  – simple algorithms based on word counts can be surprisingly good