Natural Language Processing

Lecture 16: Lexical Semantics
The Story Thus Far

• So far we have talked about...
  – Information extraction
  – Morphology
  – Language modelling
  – Classification
  – Syntax and syntactic parsing
The Path Forward

• Now we are going to talk about something that matters
The Path Forward

• Semantics (and pragmatics) are the glue that connect language to the real world
• In a sense, the other things we have talked about are only meaningful once semantics is taken into account at some level
• We will talk about...
  – Lexical semantics (the meanings of word)—this lecture
  – Word embeddings (a clever way of getting at lexical semantics)
  – Model-theoretic semantic representations for sentences
  – Semantic parsing and semantic role labelling
Three Ways of Looking at Word Meaning

• **Decompositional**
  – What the “components” of meaning “in” a word are

• **Ontological**
  – How the meaning of the word relates to the meanings of other words

• **Distributional**
  – What contexts the word is found in, relative to other words
Decompositional Semantics

boy
- human
- female
- adult

man
- human
- female
+ adult

girl
- human
+ female
- adult

woman
+ human
+ female
+ adult
Limitations of Decompositional Semantics

• Where do the features come from?
  – How do you divide semantic space into features like this?
  – How do you settle on a final list?

• How do you assign features to words in a principled fashion?

• How do you link these features to the real world?

• For these reasons, decompositional semantics is the least computationally useful approach to semantics
Ontological Approaches to Semantics
Semantic Relations

• In grammar school, or in preparation for standardized tests, you may have learned the following terms:

  synonymy, antonymy

• Synonymy and antonymy are relations between words. They are not alone:

  hyponymy, hypernymy, meronymy, holonymy
Semantic Relations

- **Synonymy**—equivalence
  - <small, little>
- **Antonymy**—opposition
  - <small, large>
- **Hyponymy**—subset; is-a relation
  - <dog, mammal>
- **Hypernymy**—superset
  - <mammal, dog>
- **Meronymy**—part-of relation
  - <liver, body>
- **Holonymy**—has-a relation
  - <body, liver>
Lexical Mini-Ontology

holonym (whole)  
meronymy (has-a)  
meronym (part)  

building.n.1  
wall.n.1  
wall.v.1  

enclosure.n.1  
fence.n.1  
surround.v.2  

hypernym  
hyponym  

build.v.1  
destroy.v.1  

antonym  
synonym  

WordNet

- WordNet is a lexical resource that organizes words according to their semantic relations.
Words have different senses
Each of those senses is associated with a synset (a set of words that are roughly synonymous for a particular sense)
These synsets are associated with one another through relations like antonymy, hyponymy, and meronymy
WordNet is a glorified electronic thesaurus
Synsets for *dog (n)*

- **S: (n)** dog, domestic **dog**, *Canis familiaris* (a member of the genus *Canis* (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) "the dog barked all night"
- **S: (n)** frump, **dog** (a dull unattractive unpleasant girl or woman) "she got a reputation as a frump"; "she's a real dog"
- **S: (n)** dog (informal term for a man) "you lucky dog"
- **S: (n)** cad, bounder, blackguard, **dog**, hound, heel (someone who is morally reprehensible) "you dirty dog"
- **S: (n)** frank, frankfurter, hotdog, hot dog, **dog**, wiener, wienerwurst, weenie (a smooth-textured sausage of minced beef or pork usually smoked; often served on a bread roll)
- **S: (n)** pawl, detent, click, **dog** (a hinged catch that fits into a notch of a ratchet to move a wheel forward or prevent it from moving backward)
- **S: (n)** andiron, firedog, **dog**, dog-iron (metal supports for logs in a fireplace) "the andirons were too hot to touch"
What’s a *Fish*?
(According to WordNet)

- **fish** (any of various mostly cold-blooded aquatic vertebrates usually having scales and breathing through gills)
- **aquatic vertebrate** (animal living wholly or chiefly in or on water)
- **vertebrate, craniate** (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
- **chordate** (any animal of the phylum Chordata having a notochord or spinal column)
- **animal, animate being, beast, brute, creature, fauna** (a living organism characterized by voluntary movement)
- **organism, being** (a living thing that has (or can develop) the ability to act or function independently)
- **living thing, animate thing** (a living (or once living) entity)
- **whole, unit** (an assemblage of parts that is regarded as a single entity)
- **object, physical object** (a tangible and visible entity; an entity that can cast a shadow)
- **entity** (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
Thesaurus-based Word Similarity

Class Mammalia

Order Artiodactyla
  - Genus Giraffidae
    - giraffe
  - Genus Bovidae
    - gazelle

Order Carnivora
  - Genus Caniformia
  - Genus Felidae
    - lion
Information Content

$IC(c) = -\log \frac{\text{# words that are equivalent to or are hyponyms of } c}{\text{# words in corpus}}$

- Entity 0.93
- Inanimate-object 1.79
  - Natural-object 4.12
    - Geological formation 6.34
      - Natural-elevation 9.09
        - Hill 10.88
      - Shore 9.39
        - Coast 10.74

(Adapted from Lin. 1998. An information Theoretic Definition of Similarity. ICML.)
WordNet Interfaces

• Various interfaces to WordNet are available
  – Many languages listed at https://wordnet.princeton.edu/related-projects
  – NLTK (Python)
    >>> from nltk.corpus import wordnet as wn
    >>> wn.synsets('dog')
    (returns list of Synset objects)
    http://www.nltk.org/howto/wordnet.html
Limitations of WordNet and Ontological Semantics

• WordNet is a useful resource that many of you will use in your projects
• There are intrinsic limits to this type of resource, however:
  – It requires many years of manual effort by skilled lexicographers
  – In the case of WordNet, some of the lexicographers were not that skilled, and this has led to inconsistencies
  – The ontology is only as good as the ontologist(s); it is not driven by data
• We will now look at an approach to lexical semantics that is data driven and does not rely on lexicographers
fertility. Organ meats such as beef and chicken liver, tongue and hear controlling scour. HOW TO FEED: BEEF and DAIRY CALVES - 0.2 gram Dy ing process discolors the treated beef and liquid accumulates in prepackag say. He did say she could get her beef and vegetables in cans this summer and feed efficiency of fattening beef animals. HOW TO FEED: At the steaks, chops, chicken and prime beef as well as Tom's favorite dish, stu ross from him was surmounted by a beef barrel with ends knocked out. In t counter of boards laid across two beef barrels. There was, of course, no Because Holstein cattle weren't a beef breed, they were rarely seen on a 2-5 grams of phenothiazine daily; beef calves - .5 to 1.5 grams daily de ties of this drug. HOW TO FEED: BEEF CATTLE (FINISHING RATION) - To dairy cows and lesser amounts to beef cattle and poultry. About 90 percen raises enough poultry, pigs, and beef cattle for most of their needs. Lo on of liver abscesses in feed-lot beef cattle. Prevention of bacterial pne pal feed bunk types for dairy and beef cattle: (1) Fence-line bunks - catt es feed efficiency. HOW TO FEED: BEEF CATTLE - 10 milligrams of diet the rations you are feeding your beef, dairy cattle, and sheep are adequa itive business more profitable for beef, dairy, and sheep men. The tar o bear. She was ready to kill the beef, dress it out, and with vegetables . She had raised a calf, grown it beef-fat. She had, with her own work-wea with feeding low-moisture corn in beef-feeding programs. Several firms ar he shelf life (at 35 F) of fresh beef from 5 days to 5 or 6 weeks. Howeve canned pork products. Tests with beef have been largely unsuccessful beca for eggs, pigs to eat garbage, a beef herd and wastes of all kinds. Separ their money's worth. A good many beef-hungry settlers were accepting the
Chicken

y the irradiated and refrigerated chicken. Acceptance of radiopasteurization
toehouse". Glendora dropped a chicken and a flurry of feathers, and went
will specialize in steaks, chops, chicken and prime beef as well as Tom's fa
ard as the one concerned with the chicken and the egg. Which came first? Is
he millions of buffalo and prairie chicken and the endless seas of grass that
"! "Come on, there's some cold chicken and we'll see what else". They wen
ves to extend the storage life of chicken at a low cost of about 0.5 cent per
CHASEN CADILLAC# Use one 6-ounce chicken breast for each guest. Salt and pe
ion juice, to about half cover the chicken breasts. Bake slowly at least one-
d, in butter. Sprinkle over top of chicken breasts. Serve each breast on a th
around, they had a hard time". #CHASEN CADILLAC# Use one 6-ounce chicken
successful, and the shelf life of chicken can be extended to a month or more
ay from making a cake, building a chicken coop, or producing a book, to found
, they decided, but a deck full of chicken coops and piggens was hardly suita
im. "Johnny insisted on cooking a chicken dinner in my honor- he's always bee
utes. Kid Ory, the trombonist chicken farmer, is also one of the solid a
y Johnson reaching around the wire chicken fencing, which half covered the tr
yes glittering behind dull silver chicken fencing. "That was Tee-wah I was t
wine in the pot roast or that the chicken had been marinated in brandy, and
yed this same game and called it "Chicken". He could not go through the f
if the Mexicans hiding in a little chicken house had passed through his head,
I'll never forget him cleaning the chicken in the tub". A story, no doubt
. Organ meats such as beef and chicken liver, tongue and heart are planne
p. "Miss Sarah, I can't cut up no chicken. Miss Maude say she won't". Aga
pot. "What is it"? he asked. "Chicken", Mose said, and theatrically licke
im"? Adam shook his head. "Chicken", Mose said. She was a child too m
## Context Vectors

<table>
<thead>
<tr>
<th></th>
<th>arts</th>
<th>boil</th>
<th>data</th>
<th>function</th>
<th>large</th>
<th>sugar</th>
<th>summarized</th>
<th>water</th>
</tr>
</thead>
<tbody>
<tr>
<td>apricot</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>pineapple</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>digital</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>information</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure 20.9** Co-occurrence vectors for four words, computed from the Brown corpus, showing only 8 of the (binary) dimensions (hand-picked for pedagogical purposes to show discrimination). Note that *large* occurs in all the contexts and *arts* occurs in none; a real vector would be extremely sparse.
## Hypothetical Counts based on Syntactic Dependencies

<table>
<thead>
<tr>
<th></th>
<th>Modified-by-ferocious(adj)</th>
<th>Subject-of-devour(v)</th>
<th>Object-of-pet(v)</th>
<th>Modified-by-African(adj)</th>
<th>Modified-by-big(adj)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lion</strong></td>
<td>15</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td><strong>Dog</strong></td>
<td>7</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td><strong>Cat</strong></td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td><strong>Elephant</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td><strong>...</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A Problem

• Some words are going to occur together many times just because they are very frequent
• The English words *the* and *is* are likely to occur in the same window many times
• They may not have a lot to do with one another except for the fact that they are frequent
• How should we address this?
Pointwise Mutual Information

$$\text{PMI}(w, f) = \log_2 \frac{p(w, f)}{p(w) \times p(f)} = \log_2 \frac{N \times \text{count}(w, f)}{\text{count}(w) \times \text{count}(f)}$$
## Distributionally Similar Words

<table>
<thead>
<tr>
<th>Rum</th>
<th>Write</th>
<th>Ancient</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>vodka</td>
<td>read</td>
<td>old</td>
<td>physics</td>
</tr>
<tr>
<td>cognac</td>
<td>speak</td>
<td>modern</td>
<td>biology</td>
</tr>
<tr>
<td>brandy</td>
<td>present</td>
<td>traditional</td>
<td>geology</td>
</tr>
<tr>
<td>whisky</td>
<td>receive</td>
<td>medieval</td>
<td>sociology</td>
</tr>
<tr>
<td>liquor</td>
<td>call</td>
<td>historic</td>
<td>psychology</td>
</tr>
<tr>
<td>detergent</td>
<td>release</td>
<td>famous</td>
<td>anthropology</td>
</tr>
<tr>
<td>cola</td>
<td>sign</td>
<td>original</td>
<td>astronomy</td>
</tr>
<tr>
<td>gin</td>
<td>offer</td>
<td>entire</td>
<td>arithmetic</td>
</tr>
<tr>
<td>lemonade</td>
<td>know</td>
<td>main</td>
<td>geography</td>
</tr>
<tr>
<td>cocoa</td>
<td>accept</td>
<td>indian</td>
<td>theology</td>
</tr>
<tr>
<td>chocolate</td>
<td>decide</td>
<td>various</td>
<td>hebrew</td>
</tr>
<tr>
<td>scotch</td>
<td>issue</td>
<td>single</td>
<td>economics</td>
</tr>
<tr>
<td>noodle</td>
<td>prepare</td>
<td>african</td>
<td>chemistry</td>
</tr>
<tr>
<td>tequila</td>
<td>consider</td>
<td>japanese</td>
<td>scripture</td>
</tr>
<tr>
<td>juice</td>
<td>publish</td>
<td>giant</td>
<td>biotechnology</td>
</tr>
</tbody>
</table>

(from an implementation of the method described in Lin. 1998. Automatic Retrieval and Clustering of Similar Words. COLING-ACL. Trained on newswire text.)
Questions?