

Algorithms for Natural Language Processing

Lecture 7: Lexical Semantics

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

$$\begin{bmatrix} +\text{human} \\ -\text{female} \\ -\text{adult} \end{bmatrix}$$

girl

$$\begin{bmatrix} +\text{human} \\ +\text{female} \\ -\text{adult} \end{bmatrix}$$

man

$$\begin{bmatrix} +\text{human} \\ -\text{female} \\ +\text{adult} \end{bmatrix}$$

woman

$$\begin{bmatrix} +\text{human} \\ +\text{female} \\ +\text{adult} \end{bmatrix}$$

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

Semantic Relations

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

synonymy, antonymy

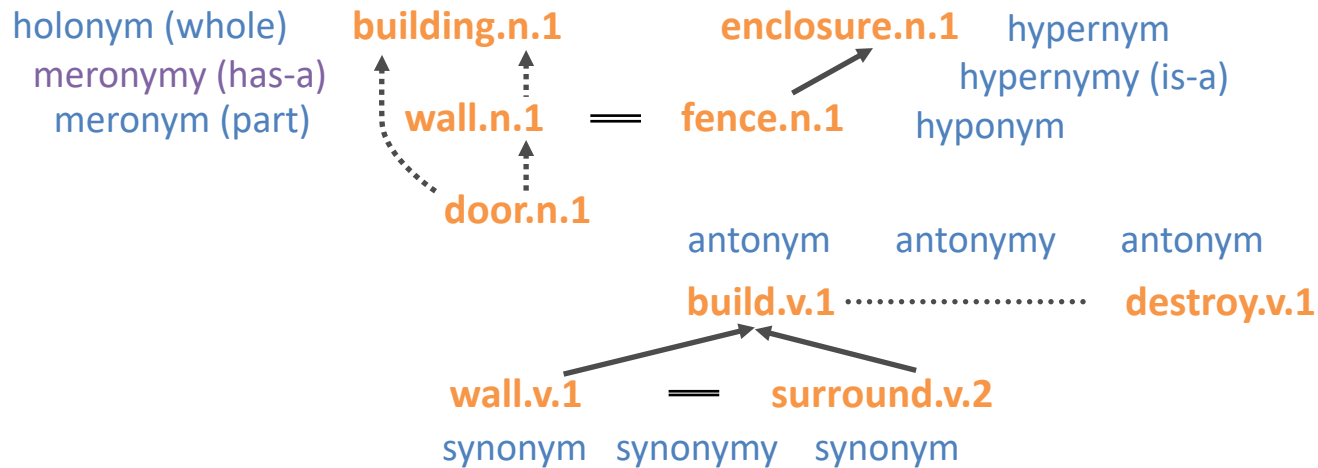
- Synonymy and antonymy are relations between words. They are not alone:

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



Roget's
THE SAURUS
of
ENGLISH WORDS
AND PHRASES

Roget's
THE SAURUS
of
ENGLISH WORDS
AND PHRASES

GROSSET
& DUNLAP

WordNet is a
glorified
electronic
thesaurus

Synsets for *dog* (n)

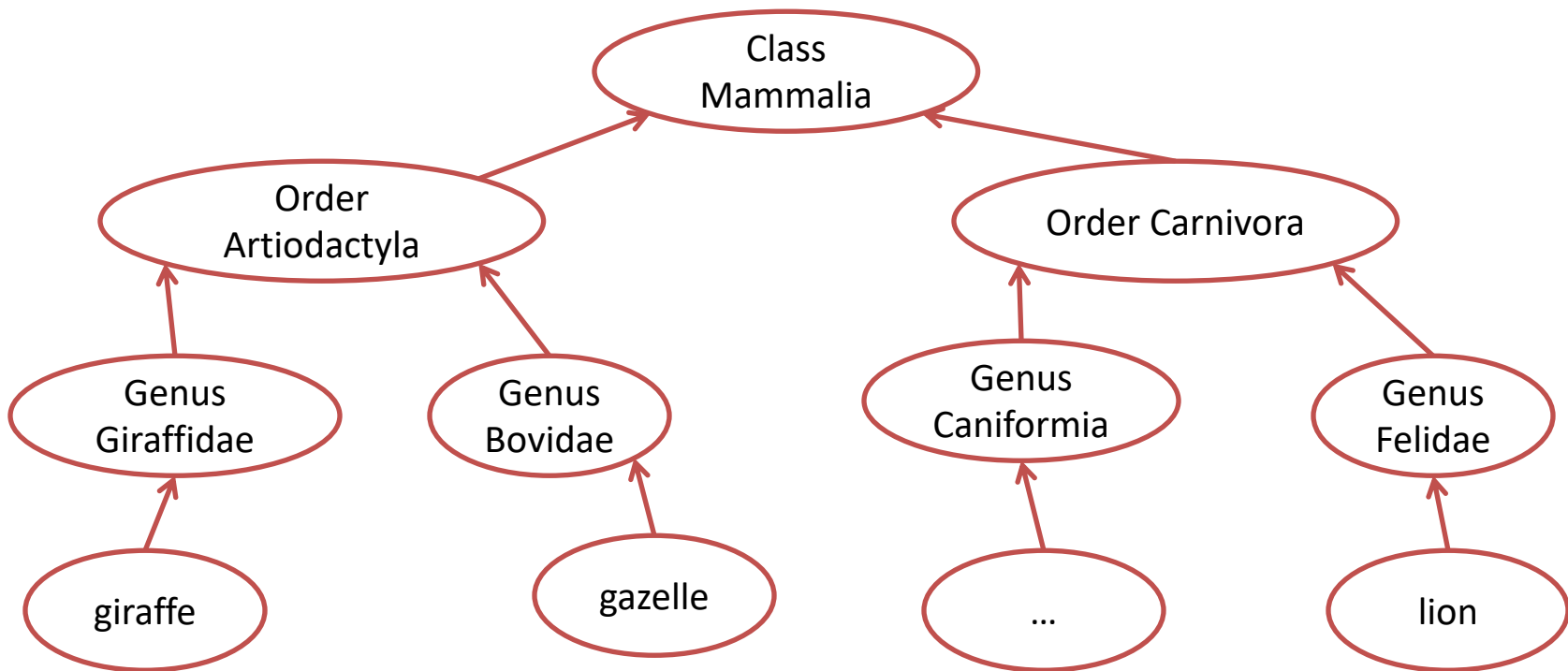
- S: (n) **dog**, **domestic dog**, **Canis familiaris** (a member of the genus *Canis* (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) "the dog barked all night"
- S: (n) **frump**, **dog** (a dull unattractive unpleasant girl 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)

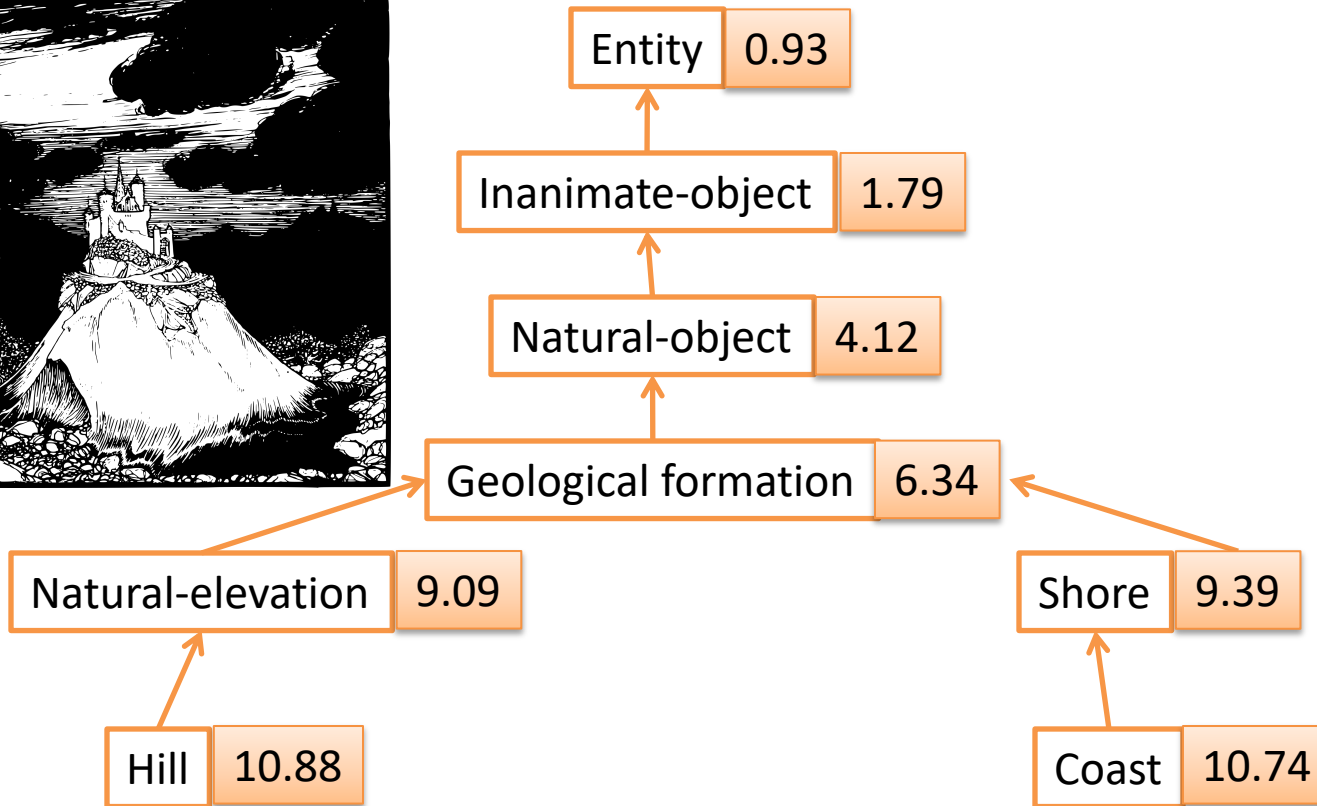
- **fish** (any of various mostly cold-blooded aquatic vertebrates usually having scales and breathing through gills)
- aquatic vertebrate (animal living wholly or chiefly in or on water)
- vertebrate, craniate (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large 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



Information Content

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



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
- 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

Beef

1 fertility. Organ meats such as beef and chicken liver, tongue and hear
2 controlling scours. HOW TO FEED: BEEF AND DAIRY CALVES - 0.2 gram Dy
3 ing process discolors the treated beef and liquid accumulates in prepackag
4 say. He did say she could get her beef and vegetables in cans this summer
5 and feed efficiency of fattening beef animals. HOW TO FEED: At the
6 steaks, chops, chicken and prime beef as well as Tom's favorite dish, stu
7 ross from him was surmounted by a beef barrel with ends knocked out. In t
8 counter of boards laid across two beef barrels. There was, of course, no
9 Because Holstein cattle weren't a beef breed, they were rarely seen on a
10 2-5 grams of phenothiazine daily; beef calves- .5 to 1.5 grams daily depe
11 ties of this drug. HOW TO FEED: BEEF CATTLE (FINISHING RATION) - To
12 dairy cows and lesser amounts to beef cattle and poultry. About 90 percent
13 raises enough poultry, pigs, and beef cattle for most of their needs. Lo
14 on of liver abscesses in feed-lot beef cattle. Prevention of bacterial pne
15 pal feed bunk types for dairy and beef cattle: (1) Fence-line bunks- catt
16 es feed efficiency. HOW TO FEED: BEEF CATTLE - 10 milligrams of diet
17 the rations you are feeding your beef, dairy cattle, and sheep are adequa
18 itive business more profitable for beef, dairy, and sheep men. The tar
19 o bear. She was ready to kill the beef, dress it out, and with vegetables
20 . She had raised a calf, grown it beef-fat. She had, with her own work-wea
21 with feeding low-moisture corn in beef-feeding programs. Several firms ar
22 he shelf life (at 35 F) of fresh beef from 5 days to 5 or 6 weeks. Howeve
23 canned pork products. Tests with beef have been largely unsuccessful beca
24 for eggs, pigs to eat garbage, a beef herd and wastes of all kinds. Separ
25 their money's worth. A good many beef-hungry settlers were accepting the

Chicken

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

Context Vectors

	arts	boil	data	function	large	sugar	summarized	water
apricot	0	1	0	0	1	1	0	1
pineapple	0	1	0	0	1	1	0	1
digital	0	0	1	1	1	0	1	0
information	0	0	1	1	1	0	1	0

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

	Modified-by-ferocious(adj)	Subject-of-devour(v)	Object-of-pet(v)	Modified-by-African(adj)	Modified-by-big(adj)
Lion	15	5	0	6	15
Dog	7	3	8	0	12
Cat	1	1	6	1	9
Elephant	0	0	0	10	15
...					

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

Rum

vodka
cognac
brandy
whisky
liquor
detergent
cola
gin
lemonade
cocoa
chocolate
scotch
noodle
tequila
juice

Write

read
speak
present
receive
call
release
sign
offer
know
accept
decide
issue
prepare
consider
publish

Ancient

old
modern
traditional
medieval
historic
famous
original
entire
main
indian
various
single
african
japanese
giant

Mathematics

physics
biology
geology
sociology
psychology
anthropology
astronomy
arithmetic
geography
theology
hebrew
economics
chemistry
scripture
biotechnology

Questions?