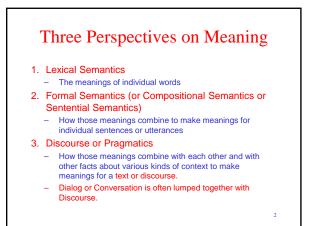
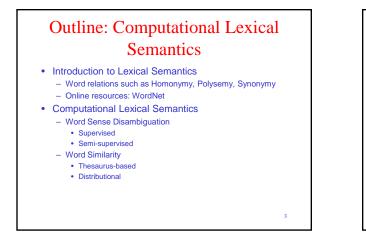
Lexical Semantics Chapter 19

Lecture #13

November 2012 Largely based on Kathy Mckeown's lecture which was based on Dan Jurafsky, Jim Martin, and Chris Manning



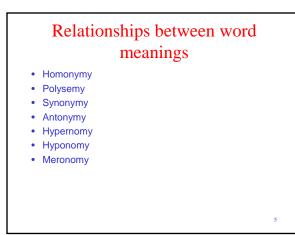


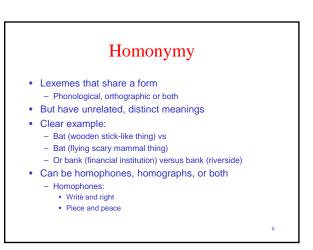
• What's a word?

Definitions we've used over the class: Types, tokens, stems, roots, uninflected forms, etc...

- Lexeme: An entry in a lexicon consisting of a pairing of a form with a single meaning representation
- Lexicon: A collection of lexemes
- Lemma citation form uninflected form (used to represent a lexeme). Need to do morphological parsing to get from wordform to lemma (lemmatization)

Lemma is part-of-speech specific (e.g., table N and V)





Homonymy causes problems for NLP applications

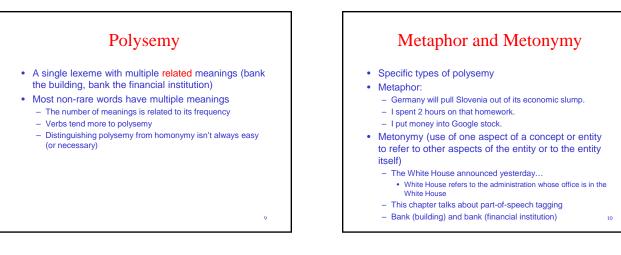
• Text-to-Speech

- Same orthographic form but different phonological form
 - Bass vs bass
 - Bow vs bow
 - Record vs record
- Information retrieval
 - Different meanings same orthographic formQUERY: bat care
- Machine Translation
- Speech recognition

Polysemy

- The bank is constructed from red brick I withdrew the money from the bank
- Are those the same sense?
- What about river bank?
- What about: The food **bank** is having a donation drive next week.
- Different senses but some more related than others...
- When two senses are related semantically we call it polysemy (rather than homonymy)

8

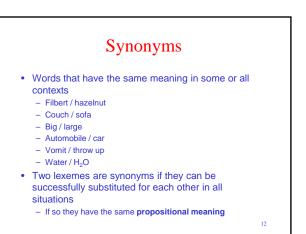


How do we know when a word has more than one sense?

• ATIS examples

- Which flights serve breakfast?
- Does America West serve Philadelphia?
- The "zeugma" test:

- ?Does United serve breakfast and San Jose?



Synonyms

- But there are few (or no) examples of perfect synonym
 - 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
 - Large coke versus *big coke

Some more terminology Lemmas and word forms 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 Corpus is the lemma for corpora Specific surface forms carpets, sung, corpora are called wordforms

- The lemma bank has two senses:
 - Instead, a bank can hold the investments in...
 - 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

15

13

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 are at opposite ends of a scale (long/short, fast/slow)
 - Be reversives (describe a change of movement in opposite directions): *rise/fall, up/down*

Hyponym

- 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

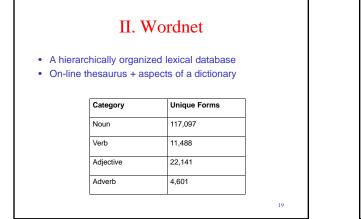
Superordinate	Vehicle	Fruit	Furniture	mammal]
Hyponym	Car	Mango	Chair	Dog	

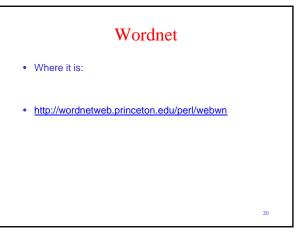
Hyponymy more formally

• 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)





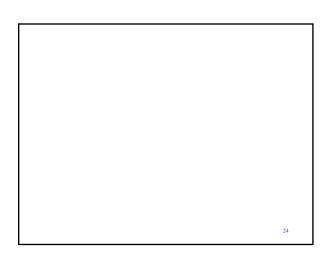
Format of WordNet Entries

- The noun bass has 8 senses in wordnet:
- 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) •
- $\underline{S}_{:}$ (n) <u>sea bass</u>, **bass** (the lean flesh of a saltwater fish of the family Serranidae)
- <u>S:</u> (n) <u>freshwater bass</u>, 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) $\underline{S:}$ (n) bass (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes) •
- And 1 Adjective Sense:
- 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" • 21

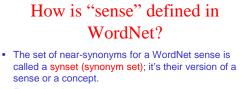
WordNet Noun Relations

Relation	Also called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$breakfast^1 \rightarrow meal^1$
Hyponym	Subordinate	From concepts to subtypes	$meal^1 \rightarrow lunch^1$
Member Meronym	Has-Member	From groups to their members	$faculty^2 \rightarrow professor$
Has-Instance		From concepts to instances of the concept	$composer^1 \rightarrow Bach^1$
Instance		From instances to their concepts	$Austen^1 \rightarrow author^1$
Member Holonym	Member-Of	From members to their groups	$copilot^1 \rightarrow crew^1$
Part Meronym	Has-Part	From wholes to parts	$table^2 \rightarrow leg^3$
Part Holonym	Part-Of	From parts to wholes	$course^7 \rightarrow meal^1$
Antonym		Opposites	$leader^1 \rightarrow follower^1$
			22

	WordNet Verb Relation	ons
Relation Hypernym Troponym Entails	From verbs (events) to the verbs (events) they entail	$snore^1 \rightarrow sleep^1$
Antonym	Opposites	increase ¹ ⇔ decrease
		23



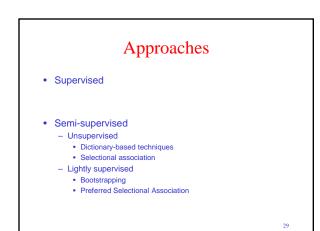
WordNe	t Hierarchies	
	e meraremes	
Sense 3 bass, basso		
	per with the lowest voice)	
	t, vocalizer, vocaliser	
	strumentalist, player	
	performing artist	
=> enterta		
-> pers	son, individual, someone	
=> c	organism, being	
	Iiving thing, animate thing,	
	=> whole, unit	
	=> object, physical object	
	=> physical entity => entity	
-	causal agent, cause, causal agency	
	-> physical entity	
	=> entity	
Sense 7		
bass		
(the member with th musical instruments	he lowest range of a family of	
=> musical instrume		
-> device	ne, institute	
-> instrument	tality, instrumentation	
	t, artefact	
	le, unit	
-> 0	object, physical object	25
	> physical entity	25
	-> entity	



- 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, fall guy#1
- sucker#1, soft touch#1, mug#2 (a person who is gullible and easy to take advantage of)
- Each of these senses share this same gloss
- Thus, for WordNet, the meaning of this sense of chump <u>is</u> this list.





Supervised Machine Learning Approaches

- Supervised machine learning approach:
 - A training corpus of ?
 - Used to train a classifier that can tag words in text
 - Just as in 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



WordNet Bass • The noun "bass" has 8 senses in WordNet • <u>S</u>: (n) bass#1 (the lowest part of the musical range) • <u>S</u>: (n) bass#2, bass part#1 (the lowest part in polyphonic music) • <u>S</u>: (n) bass#3, basso#1 (an adult male singer with the lowest voice) • <u>S</u>: (n) sea bass#1, bass#4 (the lean flesh of a saltwater fish of the family Serranidae) • <u>S</u>: (n) freshwater bass#1, bass#5 (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))

- <u>S:</u> (n) bass#6, bass voice#1, basso#2 (the lowest adult male singing voice)
- <u>S:</u> (n) bass#7 (the member with the lowest range of a family of musical instruments)
- <u>S:</u> (n) bass#8 (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Inventory of sense tags for bass				
WordNet	Spanish	Roget		
Sense	Translation	Category	Target Word in Context	
bass ⁴	lubina	FISH/INSECT	fish as Pacific salmon and striped bass and	
bass ⁴	lubina	FISH/INSECT	produce filets of smoked bass or sturgeon	
bass ⁷	bajo	MUSIC	exciting jazz bass player since Ray Brown	
bass ⁷	bajo	MUSIC	play bass because he doesn't have to solo	
			33	

Supervised WSD 2: Get a corpus

- Lexical sample task:
 - Line-hard-serve corpus -4000 examples of each
 - Interestcorpus -2369 sense-tagged examples

• All words:

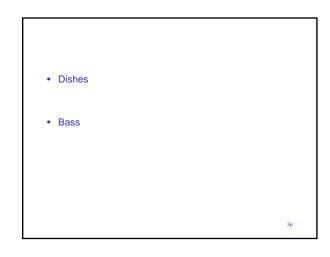
- Semantic concordance: a corpus in which each open-class word is labeled with a sense from a specific dictionary/thesaurus.
 - SemCor: 234,000 words from Brown Corpus, manually tagged with WordNet senses
 - SENSEVAL-3 competition corpora -2081 tagged word tokens

34

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?"



• washing dishes .

- simple dishes including
- convenient dishes to
- of dishes and
- free bass with
- pound bass of
- and bass player
- his bass while

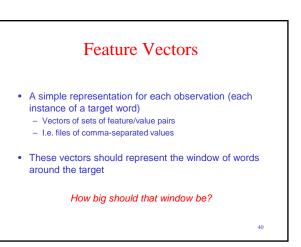
37

• "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 chcken livers with green peppers.
- Post quick and convenient *dishes* to fix when your in a hurry.
- Japanese cuisine offers a great variety of dishes and regional specialties

38

- 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.



Two kinds of features in the vectors

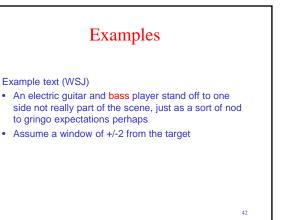
• Collocational features and bag-of-words features

- 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

41



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

Collocational

- Position-specific information about the words in the window
- guitar and bass player stand
- [guitar, NN, and, CC, player, NN, stand, VB]
- Wordn-2,POSn-2,wordn-1,POSn-1,Wordn+1POSn+1...
- · In other words, a vector consisting of
- [position n word, position n part-of-speech...]

44

