

Lexical Semantics and Word Senses


Hongning Wang

CS@UVa

Today's lecture

1. Lexical semantics
 - Meaning of words
 - Relation between different meanings
2. WordNet
 - An ontology structure of word senses
 - Similarity between words
3. Distributional semantics
 - Similarity between words
 - Word sense disambiguation

What is the meaning of a word?

- Most words have many different senses
 - dog = animal or sausage?
 - lie = to be in a horizontal position or a false statement made with deliberate intent
- What are the relations of different words in terms of meaning?
 - Specific relations between senses
 - Animal is more general than dog
 - Semantic fields  *“a set of words grouped, referring to a specific subject ... not necessarily synonymous, but are all used to talk about the same general phenomenon”* - wiki
 - Money is related to bank

Word senses

- What does ‘bank’ mean?
 - A financial institution
 - E.g., “US bank has raised interest rates.”
 - A particular branch of a financial institution
 - E.g., “The bank on Main Street closes at 5pm.”
 - The sloping side of any hollow in the ground, especially when bordering a river
 - E.g., “In 1927, the bank of the Mississippi flooded.”
 - A ‘repository’
 - E.g., “I donate blood to a blood bank.”

Lexicon entries

lemma

¹bank *noun* \ˈbɑŋk\
Definition of BANK [Cite!](#) [g+1](#) [Like](#)

- 1** : a mound, pile, or ridge raised above the surrounding level: as
 - a** : a piled-up mass of cloud or fog
 - b** : an undersea elevation rising especially from the continental shelf
- 2** : the rising ground bordering a lake, river, or sea or forming the edge of a cut or hollow
- 3 a** : a steep slope (as of a hill)
- b** : the lateral inward tilt of a surface along a curve or of a vehicle (as an airplane) when turning
- 4** : a protective or cushioning rim or piece

²bank *verb*
Definition of BANK [Cite!](#) [g+1](#) [Like](#)

transitive verb

- 1 a** : to raise a bank about
- b** : to cover (as a fire) with fresh fuel and adjust the draft of air so as to keep in an inactive state
- c** : to build (a curve) with the roadbed or track inclined laterally upward from the inside edge
- 2** : to heap or pile in a bank
- 3 a** : to drive (a ball in billiards) into a cushion
- b** : to bounce (a ball or shot) off a surface (as a backboard) into or toward a goal <*bank* in a rebound>
- 4** : to form or group in a tier

senses

Some terminologies

- **Word forms:** runs, ran, running; good, better, best
 - Any, possibly inflected, form of a word
- **Lemma** (citation/dictionary form): run; good
 - A basic word form (e.g. infinitive or singular nominative noun) that is used to represent all forms of the same word
- **Lexeme:** RUN(V), GOOD(A), BANK¹(N), BANK²(N)
 - An abstract representation of a word (and all its forms), with a part-of-speech and a set of related word senses
 - Often just written (or referred to) as the lemma, perhaps in a different FONT
- **Lexicon**
 - A (finite) list of lexemes

Make sense of word senses

- Polysemy
 - A lexeme is polysemous if it has different related senses



bank = financial institution

or

a building

Make sense of word senses

- Homonyms
 - Two lexemes are homonyms if their senses are unrelated, but they happen to have the same spelling and pronunciation

Bank of America



bank = financial institution

or

river bank



Relations between senses

- Symmetric relations
 - Synonyms: couch/sofa
 - Two lemmas with the same sense
 - Antonyms: cold/hot, rise/fall, in/out
 - Two lemmas with the opposite sense
- Hierarchical relations:
 - Hypernyms and hyponyms: pet/dog
 - The hyponym (dog) is more specific than the hypernym (pet)
 - Holonyms and meronyms: car/wheel
 - The meronym (wheel) is a part of the holonym (car)

WordNet

*George Miller, Cognitive
Science Laboratory of Princeton
University, 1985*

- A very large lexical database of English:
 - 117K nouns, 11K verbs, 22K adjectives, 4.5K adverbs
- Word senses grouped into synonym sets (“synsets”) linked into a conceptual-semantic hierarchy
 - 82K noun synsets, 13K verb synsets, 18K adjectives synsets, 3.6K adverb synsets
 - Avg. # of senses: 1.23/noun, 2.16/verb, 1.41/adj, 1.24/adverb
- Conceptual-semantic relations
 - hypernym/hyponym

A WordNet example

- <http://wordnet.princeton.edu/>

WordNet Search - 3.1
- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- [S:](#) (n) **bank** (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"*
- [S:](#) (n) [depository financial institution](#), **bank**, [banking concern](#), [banking company](#) (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"*
- [S:](#) (n) **bank** (a long ridge or pile) *"a huge bank of earth"*
- [S:](#) (n) **bank** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- [S:](#) (n) **bank** (a supply or stock held in reserve for future use (especially in emergencies))
- [S:](#) (n) **bank** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- [S:](#) (n) **bank**, [cant](#), [camber](#) (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)

Hierarchical synset relations: nouns


- **Hypernym/hyponym** (between concepts)
 - The more general ‘meal’ is a hypernym of the more specific ‘breakfast’
- **Instance hypernym/hyponym** (between concepts and instances)
 - Austen is an instance hyponym of author *Jane Austen, 1775–1817, English novelist*
- **Member holonym/meronym** (groups and members)
 - professor is a member meronym of (a university’s) faculty
- **Part holonym/meronym** (wholes and parts)
 - wheel is a part meronym of (is a part of) car.
- **Substance meronym/holonym** (substances and components)
 - flour is a substance meronym of (is made of) bread

WordNet hypernyms & hyponyms

- S: (n) **bank** (sloping land (especially the slope beside a body of water))
 - direct hyponym / full hyponym
 - S: (n) riverbank, riverside (the bank of a river)
 - S: (n) waterside (land bordering a body of water)
 - direct hypernym / inherited hypernym / sister term
 - S: (n) slope, incline, side (an elevated geological formation)
 - derivationally related form
- S: (n) depository financial institution, **bank**, banking concern, banking company (a financial institution that accepts deposits and channels the money into lending activities)
 - direct hyponym / full hyponym
 - S: (n) credit union (a cooperative depository financial institution whose members can obtain loans from their combined savings)
 - direct hypernym / inherited hypernym / sister term
 - S: (n) depository financial institution, **bank**, banking concern, banking company (a financial institution that accepts deposits and channels the money into lending activities)
 - S: (n) Federal Reserve Bank, reserve bank (one of 12 regional banks that monitor and act as depositories for banks in their region)
 - S: (n) agent bank (a bank that acts as an agent for a foreign bank)
 - S: (n) commercial bank, full service bank (a financial institution that accepts demand deposits and makes loans and provides other services for the public)

Hierarchical synset relations: verbs

*the presence of a 'manner'
relation between two lexemes*



- Hypernym/troponym (between events)
 - travel/fly, walk/stroll
 - Flying is a troponym of traveling: it denotes a specific manner of traveling
- Entailment (between events):
 - snore/sleep
 - Snoring entails (presupposes) sleeping

WordNet similarity

- Path based similarity measure between words
 - Shortest path between two concepts (Leacock & Chodorow 1998)
 - $\text{sim} = 1/|\text{shortest path}|$
 - Path length to the root node from the least common subsumer (LCS) of the two concepts (Wu & Palmer 1994)
 - $\text{sim} = 2 * \text{depth}(\text{LCS}) / (\text{depth}(w_1) + \text{depth}(w_2))$
- <http://wn-similarity.sourceforge.net/>

the most specific concept which is an ancestor of both A and B.

WordNet::Similarity

Measure	Word 1	Word 2	Score	Trace
path	apple#n#1	pizza#n#1	0.0909	<p>HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 matter#n#3 solid#n#1 food#n#2 produce#n#1 edible_fruit#n#1 apple#n#1</p> <p>HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 object#n#1 whole#n#2 natural_object#n#1 plant_part#n#1 plant_organ#n#1 reproductive_structure#n#1 fruit#n#1 edible_fruit#n#1 apple#n#1</p> <p>HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 object#n#1 whole#n#2 natural_object#n#1 plant_part#n#1 plant_organ#n#1 reproductive_structure#n#1 fruit#n#1 pome#n#1 apple#n#1</p> <p>HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 matter#n#3 substance#n#7 food#n#1 nutriment#n#1 dish#n#2 pizza#n#1</p> <p><u>Shortest path: apple#n#1 edible_fruit#n#1 produce#n#1 food#n#2 solid#n#1 matter#n#3 substance#n#7 food#n#1 nutriment#n#1 dish#n#2 pizza#n#1</u></p> <p><u>Path length = 11</u></p>
path	apple#n#2	pizza#n#1	0.0526	<p>HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 object#n#1 whole#n#2 living_thing#n#1 organism#n#1 plant#n#2 vascular_plant#n#1 woody_plant#n#1 tree#n#1 angiospermous_tree#n#1 fruit_tree#n#1 apple_tree#n#1 apple#n#2</p> <p>HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 matter#n#3 substance#n#7 food#n#1 nutriment#n#1 dish#n#2 pizza#n#1</p> <p>Shortest path: apple#n#2 apple_tree#n#1 fruit_tree#n#1 angiospermous_tree#n#1 tree#n#1 woody_plant#n#1 vascular_plant#n#1 plant#n#2 organism#n#1 living_thing#n#1 whole#n#2 object#n#1 physical_entity#n#1 matter#n#3 substance#n#7 food#n#1 nutriment#n#1 dish#n#2 <u>pizza#n#1</u></p> <p><u>Path length = 19</u></p>

WordNet::Similarity

Measure	Word 1	Word 2	Score	Trace
wup	apple#n#1	pizza#n#1	0.4444	HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 matter#n#3 solid#n#1 food#n#2 produce#n#1 edible_fruit#n#1 apple#n#1 HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 object#n#1 whole#n#2 natural_object#n#1 plant_part#n#1 plant_organ#n#1 reproductive_structure#n#1 fruit#n#1 edible_fruit#n#1 apple#n#1 HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 object#n#1 whole#n#2 natural_object#n#1 plant_part#n#1 plant_organ#n#1 reproductive_structure#n#1 fruit#n#1 pome#n#1 apple#n#1 HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 matter#n#3 <u>substance#n#7 food#n#1 nutriment#n#1 dish#n#2 pizza#n#1</u> <u>Lowest Common Subsumers: matter#n#3 (Depth=4) !</u> Depth(apple#n#1) = 9 Depth(pizza#n#1) = 9
wup	apple#n#2	pizza#n#1	0.25	HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 object#n#1 whole#n#2 living_thing#n#1 organism#n#1 plant#n#2 vascular_plant#n#1 woody_plant#n#1 tree#n#1 angiospermous_tree#n#1 fruit_tree#n#1 apple_tree#n#1 apple#n#2 HyperTree: *Root*#n#1 entity#n#1 physical_entity#n#1 matter#n#3 <u>substance#n#7 food#n#1 nutriment#n#1 dish#n#2 pizza#n#1</u> <u>Lowest Common Subsumers: physical_entity#n#1 (Depth=3) !</u> Depth(apple#n#2) = 15 Depth(pizza#n#1) = 9

Distributional hypothesis

- What is **tezgüino**?
 - A bottle of **tezgüino** is on the table.
 - Everybody likes **tezgüino**.
 - **Tezgüino** makes you drunk.
 - We make **tezgüino** out of corn.
- The contexts in which a word appears tell us a lot about what it means

Distributional semantics

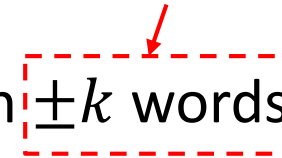
- Use the contexts in which words appear to measure their similarity
 - Assumption: similar contexts => similar meanings
 - Approach: represent each word w as a vector of its contexts c
 - Vector space representation
 - Each dimension corresponds to a particular context c_n
 - Each element in the vector of w captures the degree to which the word w is associated with the context c_n
 - Similarity metric
 - Cosine similarity

How to define the contexts

- Nearby words

- w appears near c if c occurs within $\pm k$ words of w
 - It yields fairly broad thematic relations
- Decide on a fixed vocabulary of N context words $c_1 \dots c_N$
 - Prefer words occur frequently enough in the corpus but not too frequent (i.e., avoid stopwords)
- Co-occurrence count of word w and context c as the corresponding element in the vector
 - Pointwise Mutual Information (PMI)

within a sentence



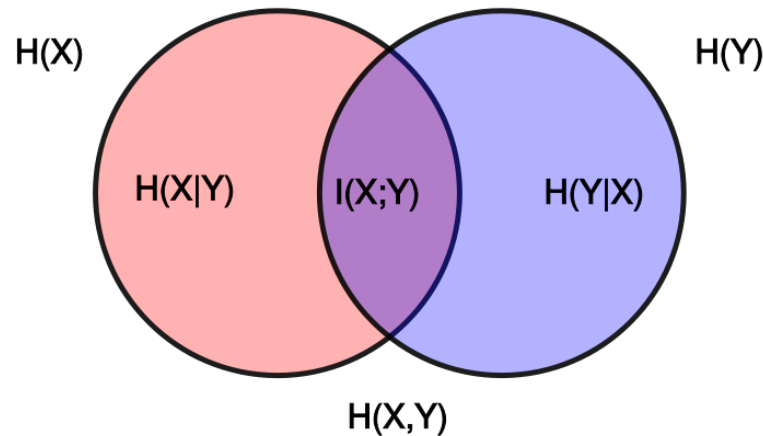
- Grammatical relations

- How often is w used as the subject of the verb c ?
- Fine-grained thematic relations

Mutual information

- Relatedness between two random variables

$$-I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right)$$



Pointwise mutual information

within a sentence

- PMI between w and c using a **fixed window** of $\pm k$ words

$$- PMI(w; c) = p(w, c) \log\left(\frac{p(w, c)}{p(w)p(c)}\right)$$

How often w and c co-occur inside a window

How often w occurs

How often c occurs

Word sense disambiguation

- What does this word mean?
 - This **plant** needs to be **watered** each day.
 - living plant
 - This **plant** **manufactures** 1000 widgets each day.
 - factory
- Word sense disambiguation (WSD)
 - Identify the sense of content words (noun, verb, adjective) in context (assuming a fixed inventory of word senses)

Dictionary-based methods

- A dictionary/thesaurus contains glosses and examples of a word

bank¹

Gloss: a financial institution that accepts deposits and channels the money into lending activities


Examples: *“he cashed the check at the bank”,
“that bank holds the mortgage on my home”*

bank²

Gloss: sloping land (especially the slope beside a body of water)

Examples: *“they pulled the canoe up on the bank”,
“he sat on the bank of the river and watched the current”*

Lesk algorithm

- Compare the context with the dictionary definition of the sense  context words
 - Construct the **signature** of a word in context by the signatures of its senses in the dictionary
 - **Signature** = set of context words (in examples/gloss or in context)
 - Assign the dictionary sense whose gloss and examples are the most **similar** to the context in which the word occurs
 - Similarity = size of intersection of context signature and sense signature

Sense signatures

bank¹

Gloss: a financial institution that accepts deposits and channels the money into lending activities

Examples: “he *cash*ed the *check* at the *bank*”,
“that *bank* *hold*s the *mortgage* on my *home*”

Signature(bank¹) = {*financial, institution, accept, deposit, channel, money, lend, activity, cash, check, hold, mortgage, home*}

bank²

Gloss: sloping land (especially the slope beside a body of water)

Examples: “they *pull*ed the *canoe* up on the *bank*”,
“he *sat* on the *bank* of the *river* and *watch*ed the *current*”

Signature(bank²) = {*slope, land, body, water, pull, canoe, sit, river, watch, current*}

Signature of target word

*“The **bank** refused to give me a loan.”*

- Simplified Lesk
 - Words in context
 - *Signature(bank) = {**refuse, give, loan**}*
- Original Lesk
 - Augmented signature of the target word
 - *Signature(bank) = {**refuse, reject, request,...** , **give, gift, donate,...** **loan, money, borrow,...**}*

Learning-based Methods

- Will be discussed in the lecture of “Text Categorization”
 - Basically treat each sense as an independent class label
 - Construct classifiers to assign each instance with context into the classes/senses

What you should know

- Lexical semantics
 - Relationship between words
 - WordNet
- Distributional semantics
 - Similarity between words
 - Word sense disambiguation

Today's reading

- Speech and Language Processing
 - Chapter 19: Lexical Semantics
 - Chapter 20: Computational Lexical Semantics