Lecture 13: Classical Lexical Semantics

USC VSoE CSCI 544: Applied Natural Language Processing

Jonathan May -- 梅約納

October 11, 2017

slides from Dan Jurafsky and Nathan Schneider

Question Answering

- **Q: What is a good way to remove wine stains?**
- A1: Salt is a great way to eliminate wine stains
- A2: How to get rid of wine stains...
- A3: How to get red wine out of clothes...
- A4: Oxalic acid is infallible in removing iron-rust and ink stains.

Lots of synonyms!

Document Similarity

- Given a movie script, recommend similar movies

ALAN TURING (V.O.)
Are you paying attention?

INT. ALAN TURING'S HOUSE - DAY - 1951

A HALF-DOZEN POLICE OFFICERS swarm the Manchester home of mathematics professor Alan Turing.

ALAN TURING (V.O.)
Good. This is going to go very quickly now. If you are not listening carefully, you will miss things. Important things. You're writing some of this down? That's good.

BERTIE
Like mad King George the Third, there'll be King George the stammerer, who let his people down so badly in their hour of need!

Lionel sits down on the chair of Edward the Confessor. Leaning against it is the great two-handed sword of St. George.

BERTIE
What're you doing? Get up! You can't sit there!

LIONEL
Why not? It's a chair.

Intuition of Semantic Similarity

- Semantically Close
  - bank-money
  - apple-fruit
  - tree-forest
  - bank-river
  - pen-paper
  - run-walk
  - mistake-error
  - car-wheel
- Semantically Distant
  - doctor-beer
  - painting-January
  - money-river
  - apple-penguin
  - nurse-fruit
  - pen-river
  - clown-tramway
  - car-algebra

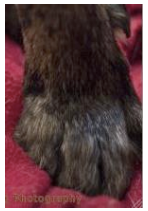
Why Are Two Words Similar?

- Meaning
  - Two concepts are close in terms of meaning (want-desire)
- World knowledge
  - Two concepts have similar properties, often occur together, or occur in similar contexts (pencil-pen, pen-ink, dog-cat)
- Psychology
  - We often think of the two concepts together (death-taxes)

Two Types of Relations

- Synonymy: two words are (roughly) interchangeable)


- Semantic similarity (distance): two words are somehow "related"
  - Sometimes there's an explicit lexical semantic relationship, often not

Validity of Semantic Similarity

- Is semantic distance a valid linguistic phenomenon?
- Experiment (Rubenstein and Goodenough, 1965)
  - Compiled a list of word pairs
  - Subjects asked to judge semantic distance (from 0 to 4) for each pair
- Results
  - Rank correlation between subjects is ~ 0.9
  - People are consistent!

Why Do This

- Task: automatically compute semantic similarity between words
- Can be useful for many applications:
  - Detecting paraphrases (i.e. automatic essay grading, plagiarism detection)
  - Information Retrieval
  - Machine Translation
- Why? because similarity gives us a way to generalize beyond word identities

Word similarity for plagiarism detection

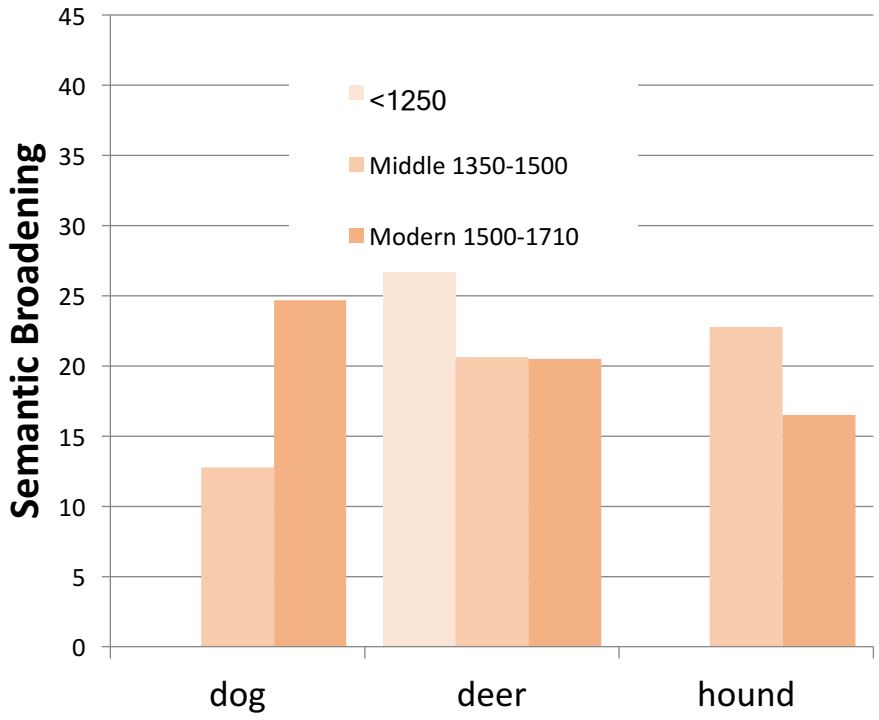
MAINFRAMES

Mainframes are primarily referred to large computers with rapid, advanced processing capabilities that can execute and perform tasks equivalent to many Personal Computers (PCs) machines networked together. It is characterized with high quantity Random Access Memory (RAM), very large secondary storage devices, and high-speed processors to cater for the needs of the computers under its service.

Consisting of advanced components, mainframes have the capability of running multiple large applications required by many and most enterprises and organizations. This is one of its advantages. Mainframes are also suitable to cater for those applications (programs) or files that are of very high demand by its users (clients). Examples of such organizations and enterprises using mainframes are online shopping websites such as Ebay, Amazon, and computing-giant

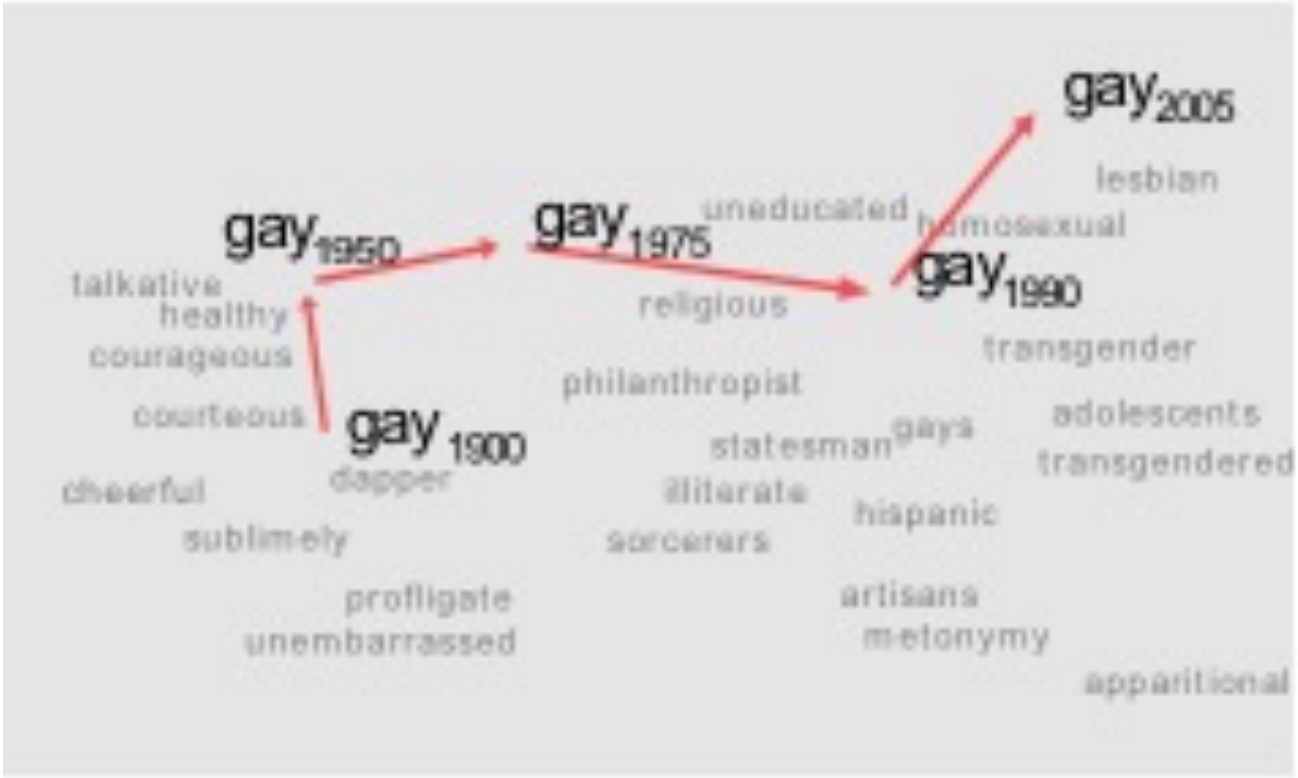
Word similarity for historical linguistics: semantic change over time

Sagi, Kaufmann, Clark 2013



how general are terms over time?

Kulkarni, Al-Rfou, Perozzi, Skiena 2015



Task Evaluation: Correlation with Humans

- Ask humans and computer to rank pairs of words in order of semantic distance
- Measure correlation

Evaluation: Word Choice Problems

- Identify the alternative that is closest in meaning to the target
- accidental
  - wheedle
  - ferment
  - inadvertent
  - abominate
- imprison
  - incarcerate
  - writhe
  - meander
  - inhibit

Evaluation: Malapropisms

- Find the word that is most likely wrong
  - Jack withdrew money from the ATM next to the band

Two classes of similarity algorithms

- Thesaurus-based algorithms
  - Are words “nearby” in a thesaurus hierarchy?
  - Do words have similar glosses (definitions)?
- Distributional algorithms
  - Do words behave similarly in real-world usage?

WordNet 3.1

- A hierarchically organized lexical database of English
- On-line thesaurus + aspects of a dictionary
    - globalwordnet.org: wordnets in 50+ languages
        - quality/coverage/availability varies
- Let's talk about how words can semantically relate to each other first

Category	Unique Strings
Noun	117,798
Verb	11,529
Adjective	22,479
Adverb	4,481

Terminology: lemma and wordform

- A **lemma** or **citation form**
  - Representation of all forms with the same stem, part of speech, rough semantics
- A **wordform**
  - The inflected word as it appears in text

Wordform	Lemma
banks	bank
sung	sing
duermes	dormir

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

Homonymy: multi-sense as an artifact

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

- $bank_1$: financial institution, $bank_2$: sloping land
- bat_1: club for hitting a ball, bat_2: nocturnal flying mammal

1. Homographs (bank/bank, bat/bat)
2. Homophones:
    1. Write and right
    2. Piece and peace

Homonymy causes problems for NLP applications

- Information retrieval
  - "`bat care`"
- Machine Translation
  - `bat:` murciélago (animal) or bate (for baseball)
- Text-to-Speech
  - `bass` (stringed instrument) vs. `bass` (fish)

Polysemy: related multi-sense

- 1. The **bank** was constructed in 1875 out of local red brick.
- 2. I withdrew the money from the **bank**
- Are those 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 meanings

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?

- The “zeugma” test: Two senses of `serve`?
  - `Which flights **serve** breakfast?`
  - `Does Lufthansa **serve** Philadelphia?`
  - ?Does Lufthansa serve breakfast and San Jose?
- Since this conjunction sounds weird,
  - we say that these are **two different senses of “serve”**

Synonyms

- Word that have the same meaning in some or all contexts.
  - filbert / hazelnut
  - couch / sofa
  - big / large
  - automobile / car
  - vomit / throw up
  - Water / H_20
- Two lexemes are synonyms
  - if they can be substituted for each other in all situations
  - If so they have the same **propositional meaning**

Synonyms

- But there are 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.
- Example:
  - Water/H_20
  - 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/superordinate** ("hyper is super")
  - *vehicle* is a **hypernym** of *car*
  - *fruit* is a hypernym of *mango*

Superordinate/hyper	vehicle	fruit	furniture
Subordinate/hyponym	car	mango	chair

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)
- Another name: the **IS-A hierarchy**
  - A IS-A B (or A ISA B)
  - B **subsumes** A

Hyponyms and Instances

- WordNet has both **classes** and **instances**.
- An **instance** is an individual, a proper noun that is a unique entity
  - `San Francisco` is an **instance** of `city`
  - But `city` is a class
    - `city` is a **hyponym** of `municipality...location...`

Meronymy

- The part-whole relation
  - A *leg* is part of a *chair*; a *wheel* is part of a *car*.
- *Wheel* is a **meronym** of *car*, and *car* is a **holonym** of *wheel*.

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?

- **The synset (synonym set),** the set of near-synonyms, instantiates a sense or concept, with a gloss
- Example: chump as a noun with the 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)

WordNet Hypernym Hierarchy for “bass”

- S: (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) musician, instrumentalist, player (someone who plays a musical instrument (as a profession))
        - 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 for one 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) *"how big is that part compared to the whole?"; "the team is a unit"*
                    - S: (n) object, physical object (a tangible and visible entity; an entity that can cast 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 its own distinct existence (living or nonliving))

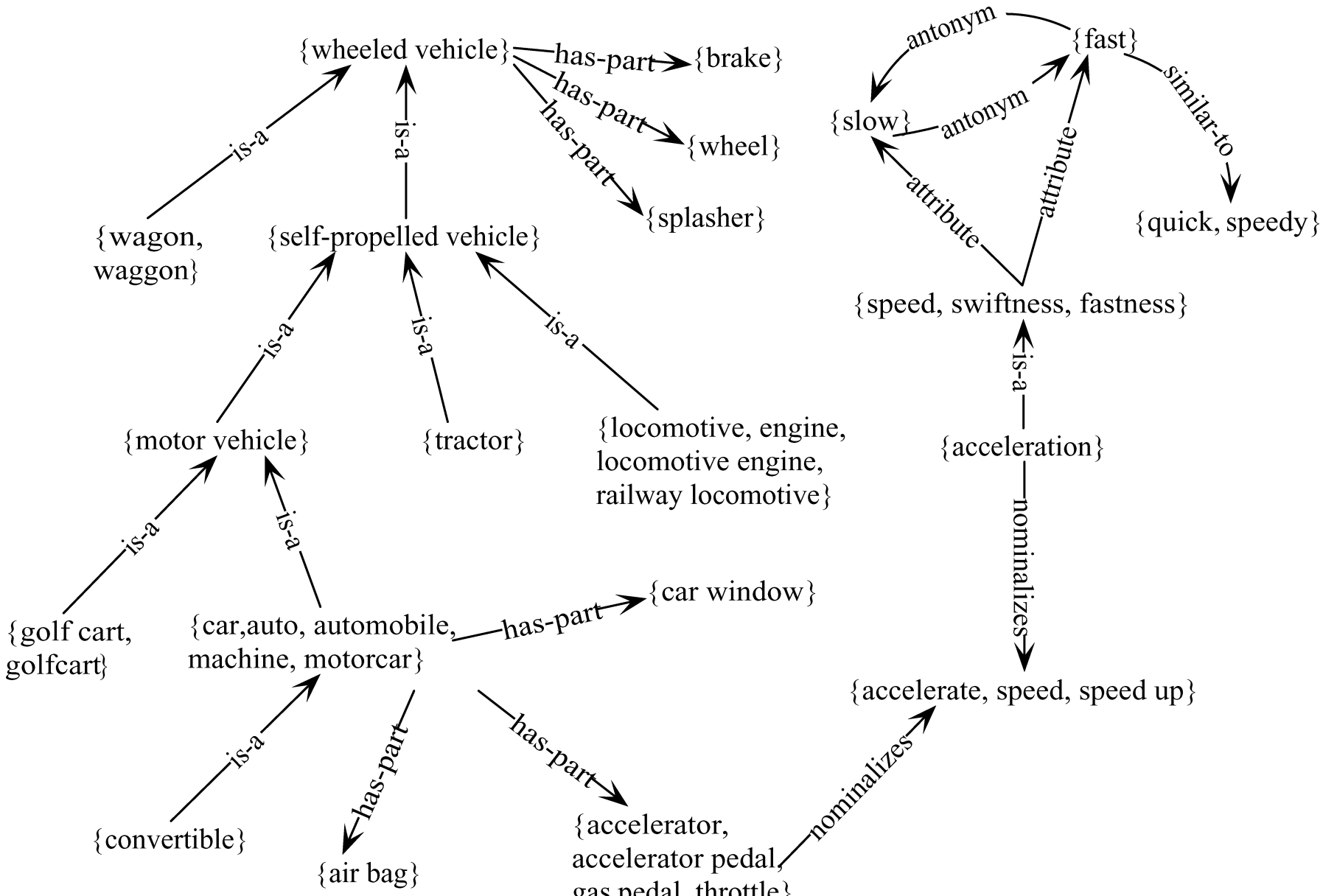
WordNet Noun Relations

Relation	Also Called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$breakfast^1 \rightarrow meal^1$
Hyponym	Subordinate	From concepts to subtypes	$meal^1 \rightarrow lunch^1$
Instance Hypernym	Instance	From instances to their concepts	$Austen^1 \rightarrow author^1$
Instance Hyponym	Has-Instance	From concepts to concept instances	$composer^1 \rightarrow Bach^1$
Member Meronym	Has-Member	From groups to their members	$faculty^2 \rightarrow professor^1$
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	$course^7 \rightarrow meal^1$
Substance Meronym		From substances to their subparts	$water^1 \rightarrow oxygen^1$
Substance Holonym		From parts of substances to wholes	$gin^1 \rightarrow martini^1$
Antonym		Semantic opposition between lemmas	$leader^1 \Longleftrightarrow follower^1$
Derivationally Related Form		Lemmas w/same morphological root	$destruction^1 \Longleftrightarrow destroy^1$

WordNet VerbRelations

Relation	Definition	Example
Hypernym	From events to superordinate events	*fly*^{9} $\rightarrow$ *travel*^{5}
Troponym	From events to subordinate event (often via specific manner)	*walk*^{1} $\rightarrow$ *stroll*^{1}
Entails	From verbs (events) to the verbs (events) they entail	*snore*^{1} $\rightarrow$ *sleep*^{1}
Antonym	Semantic opposition between lemmas	*increase*^{1} $\Longleftrightarrow$ *decrease*^{1}
Derivationally Related Form	Lemmas with same morphological root	*destroy*^{1} $\Longleftrightarrow$ *destruction*^{1}

WordNet: Viewed as a graph



WordNet 3.1

- Where it is:
  - http://wordnetweb.princeton.edu/perl/webwn
- Libraries
  - Python: WordNet from NLTK
    - http://www.nltk.org/Home
  - Java:
    - JWNL, extJWNL on sourceforge

Word Similarity

- **Synonymy**: a binary relation
  - Two words are either synonymous or not
- **Similarity** (or **distance**): a looser metric
  - Two words are more similar if they share more features of meaning
- Similarity is properly a relation between **senses**
  - The word "`bank`" is not similar to the word "`slope`"
  - Bank^{1} is similar to fund^{3}
  - Bank^{2} is similar to slope^{5}
- But we'll compute similarity over both words and senses

Why word similarity

- A practical component in lots of NLP tasks
  - Question answering
  - Natural language generation
  - Automatic essay grading
  - Plagiarism detection
- A theoretical component in many linguistic and cognitive tasks
  - Historical semantics
  - Models of human word learning
  - Morphology and grammar induction

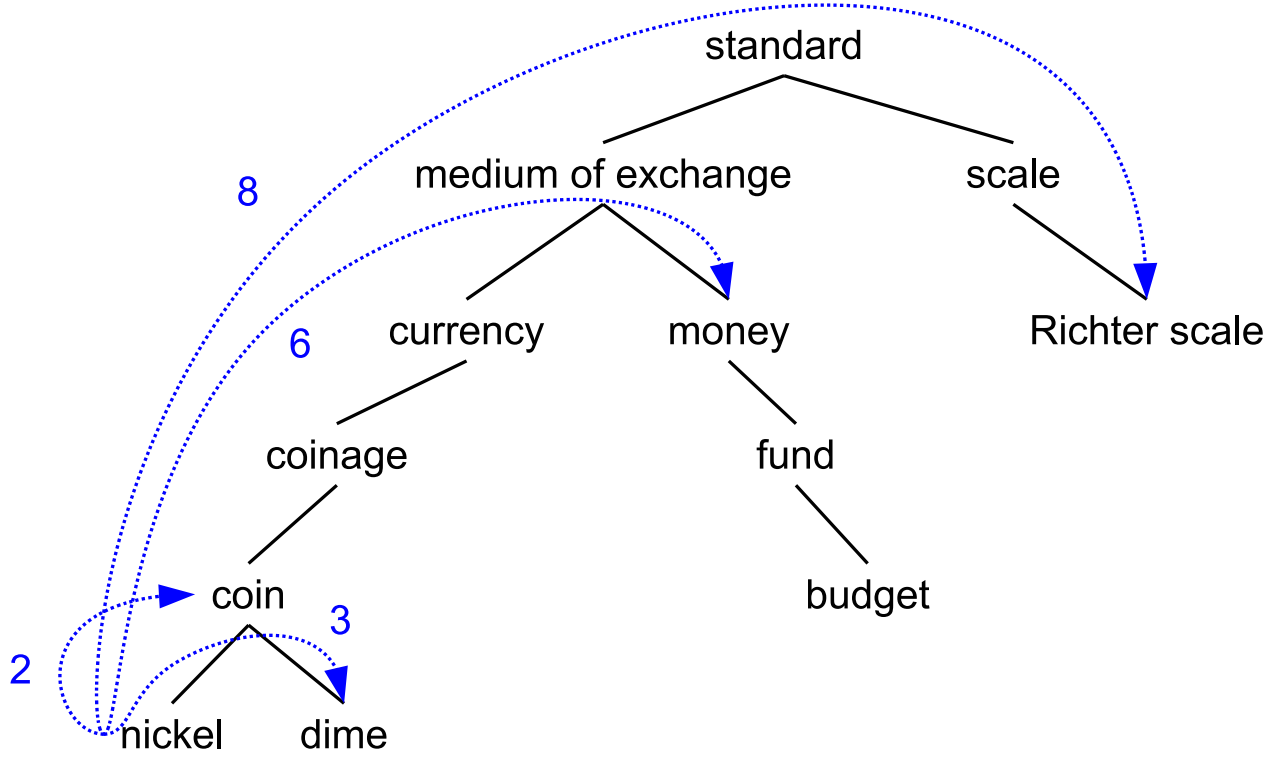
Word similarity and word relatedness

- We often distinguish **word similarity** from **word relatedness**
  - **Similar words**: near-synonyms
  - **Related words**: can be related any way
    - `car, bicycle`: **similar**
    - `car, gasoline`: **related**, not similar

Thesaurus-based similarity algorithms

- **Thesaurus-based algorithms**
  - **Are words "nearby" in hypernym hierarchy?**
  - **Do words have similar glosses (definitions)?**
- Distributional algorithms
  - Do words have similar distributional contexts?

Path based similarity



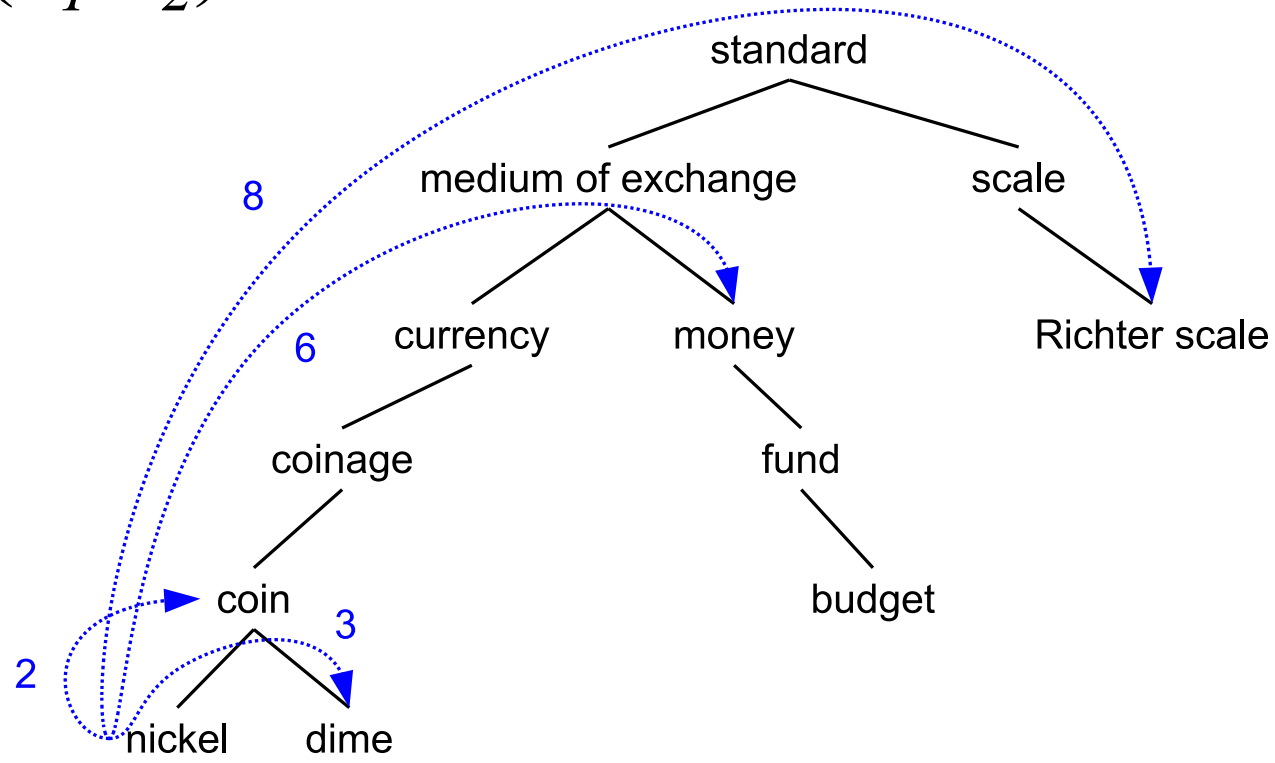
- Two concepts (senses/synsets) are similar if they are near each other in the thesaurus hierarchy
  - =have a short path between them
  - concepts have path 1 to themselves

Refinements to path-based similarity

- pathlen(c_1,c_2) = 1 + number of edges in the shortest path in the hypernym graph between sense nodes c_1 and c_2
- $\text{simpath}(c_1,c_2) = \frac{1}{\text{pathlen}(c_1,c_2)}$
  - ranges from 0 to 1
- $\text{wordsim}(w_1,w_2) = \max_{\substack{c_1 \in \text{senses}(w_1),\\ c_2 \in \text{senses}(w_2)}} \text{simpath}(c_1,c_2)$

Example: path-based similarity

$\text{simpath}(c_1,c_2) = 1/\text{pathlen}(c_1,c_2)$



simpath(*nickel,coin*) = 1/2 = .5

simpath(*fund,budget*) = 1/2 = .5

simpath(*nickel,currency*) = 1/4 = .25

simpath(*nickel,money*) = 1/6 = .17

simpath(*nickel,standard*) = 1/6 = .17

Problem with basic path-based similarity

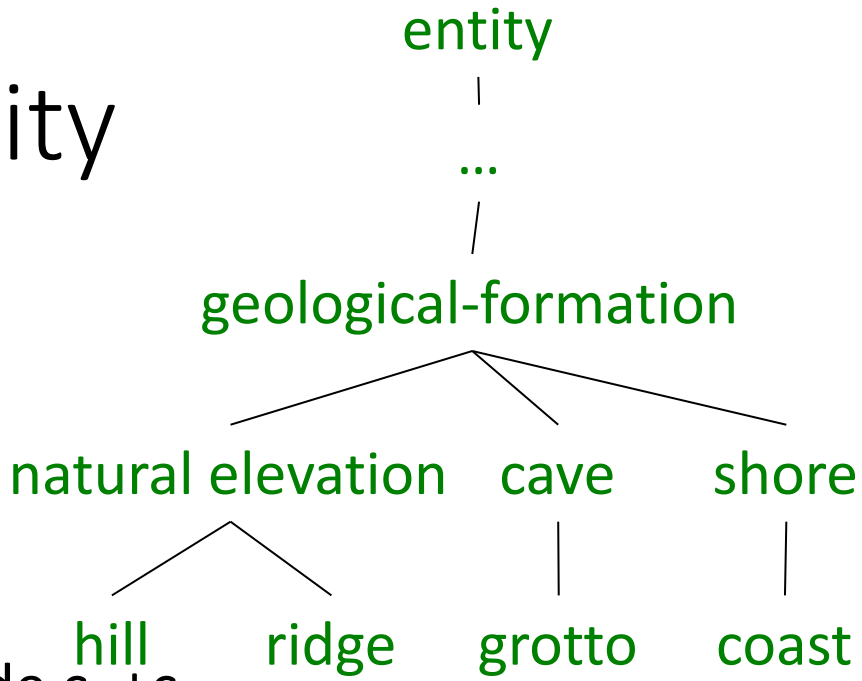
- Assumes each link represents a uniform distance
  - But *nickel* to *money* seems to us to be closer than *nickel* to *standard*
  - Nodes high in the hierarchy are very abstract
- We instead want a metric that
  - Represents the cost of each edge independently
  - Words connected only through abstract nodes are less similar

Information content similarity metrics

Resnik 1995

- Let's define $P(c)$ as:
  - The probability that a randomly selected word in a corpus is an instance of concept c
  - Formally: there is a distinct random variable, ranging over words, associated with each concept in the hierarchy
    - for a given concept, each observed noun is either
      - a member of that concept with probability $P(c)$
      - not a member of that concept with probability $1\text{-}P(c)$
  - All words are members of the root node (Entity)
    - $P(root)=1$
  - The lower a node in hierarchy, the lower its probability

Information content similarity



- Train by counting in a corpus
  - Each instance of `hill` counts toward frequency of *natural elevation*, *geological formation*, *entity*, etc
  - Let $\text{words}(c)$ be the set of all words that are children of node c, +c
    - words("geo-formation") = {hill,ridge,grotto,coast,cave,shore,natural elevation, geo-formation}
    - words("natural elevation") = {hill, ridge, natural elevation}

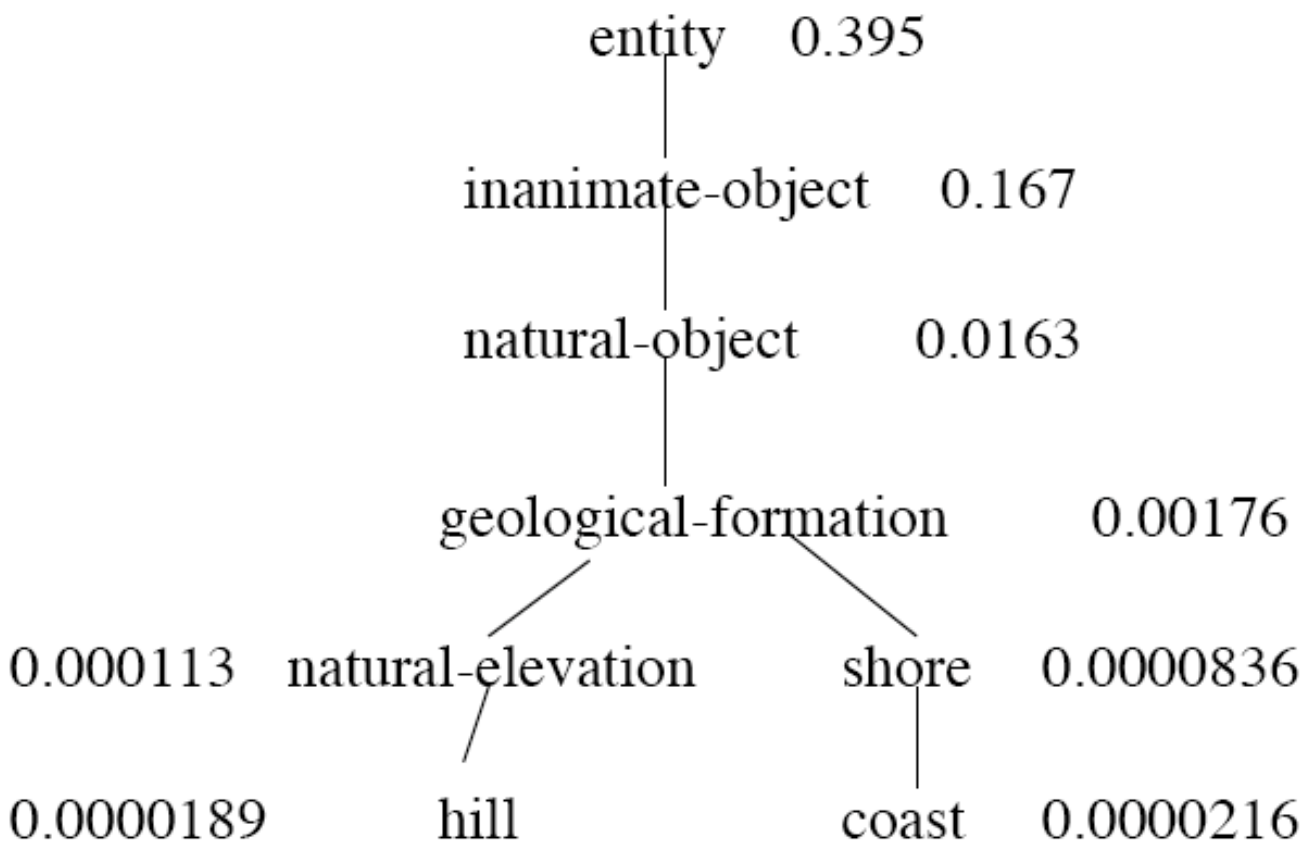
$$P(c) = \frac{\sum_{w \in words(c)} count(w)}{N}$$

N = words in the corpus

Information content similarity

- WordNet hierarchy augmented with probabilities P(c)

D. Lin. 1998. An Information-Theoretic Definition of Similarity. ICML 1998



Information content: definitions

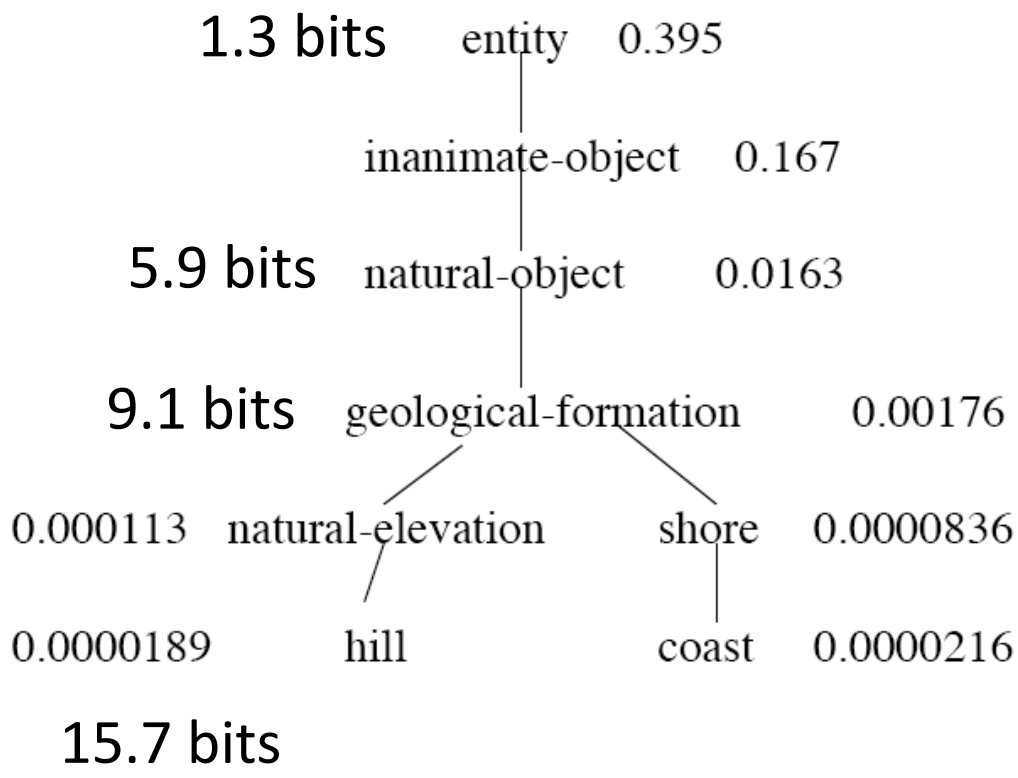
- Information content:

  $IC(c) = -\log P(c)$

- Most informative subsumer (Lowest common subsumer)

  $LCS(c_1,c_2) =$

  The most informative (lowest) node in the hierarchy subsuming both c_1 and c_2



Using information content for similarity: the Resnik method

Philip Resnik. 1995. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. IJCAI 1995.
Philip Resnik. 1999. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. JAIR 11, 95-130.

- The similarity between two words is related to their common information
- The more two words have in common, the more similar they are
- Resnik: measure common information as:
  - The information content of the most informative (lowest) subsumer (MIS/LCS) of the two nodes
  - $\text{sim}_{\text{resnik}}(c_1,c_2) = -\log P(\text{LCS}(c_1,c_2))$

Python Notebook Demo

- Files should be available to you on blackboard

Other thesaurus-based similarity metrics

$$\text{sim}_{\text{path}}(c_1, c_2) = \frac{1}{pathlen(c_1, c_2)}$$

$$\text{sim}_{\text{resnik}}(c_1, c_2) = -\log P(LCS(c_1, c_2)) \quad \text{sim}_{\text{lin}}(c_1, c_2) = \frac{2\log P(LCS(c_1, c_2))}{\log P(c_1) + \log P(c_2)}$$

$$\text{sim}_{\text{jiangconrath}}(c_1, c_2) = \frac{1}{\log P(c_1) + \log P(c_2) - 2\log P(LCS(c_1, c_2))}$$

$$\text{sim}_{eLesk}(c_1, c_2) = \sum_{r,q \in RELS} \text{overlap}(\text{gloss}(r(c_1)), \text{gloss}(q(c_2)))$$

Thesaurus Methods: Limitations

- Measure is only as good as the resource
- Limited in scope
  - Assumes IS-A relations
  - Works mostly for nouns
- Role of context not accounted for
- Not easily domain-adaptable
- Resources not available in many languages