CMSC 723: Computational Linguistics I — Session #10

Semantic Distance

Jimmy Lin
The iSchool
University of Maryland

Wednesday, November 4, 2009

Material drawn from slides by Saif Mohammad and Bonnie Dorr

Progression of the Course

- Words
  - Finite-state morphology
  - Part-of-speech tagging (TBL + HMM)
- Structure
  - CFGs + parsing (CKY, Earley)
  - N-gram language models
- Meaning!

Today's Agenda

- Lexical semantic relations
- WordNet
- Computational approaches to word similarity

Lexical Semantic Relations

What's meaning?

- Let's start at the word level…
- How do you define the meaning of a word?
- Look it up in the dictionary!

right	*adj.* located nearer the right hand esp. being on the right when facing the same direction as the observer.
left	*adj.* located nearer to this side of the body than the right.
red	*n.* the color of blood or a ruby.
blood	*n.* the red liquid that circulates in the heart, arteries and veins of animals.

Well, that really doesn't help…

Approaches to meaning

- Truth conditional
- Semantic network

Word Senses

- “Word sense” = distinct meaning of a word
- Same word, different senses
  - Homonyms (homonymy): unrelated senses; identical orthographic form is coincidental
    - Example: “financial institution” vs. “side of river” for bank
  - Polysemes (polysemy): related, but distinct senses
    - Example: “financial institution” vs. “sperm bank”
  - Metonyms (metonymy): “stand in”, technically, a sub-case of polysemy
    - Examples: author for works or author, building for organization, capital city for government
- Different word, same sense
  - Synonyms (synonymy)

Just to confuse you…

- Homophones: same pronunciation, different orthography, different meaning
  - Examples: would/wood, to/too/two
- Homographs: distinct senses, same orthographic form, different pronunciation
  - Examples: bass (fish) vs. bass (instrument)

Relationship Between Senses

- IS-A relationships
  - From specific to general (up): hypernym (hypernymy)
    - Example: bird is a hypernym of robin
  - From general to specific (down): hyponym (hyponymy)
    - Example: robin is a hyponym of bird
- Part-Whole relationships
  - wheel is a meronym of car (meronymy)
  - car is a holonym of wheel (holonymy)

WordNet Tour

Material drawn from slides by Christiane Fellbaum

What is WordNet?

- A large lexical database developed and maintained at Princeton University
- Includes most English nouns, verbs, adjectives, adverbs
- Electronic format makes it amenable to automatic manipulation: used in many NLP applications
- “WordNets” generically refers to similar resources in other languages

WordNet: History

- Research in artificial intelligence:
  - How do humans store and access knowledge about concept?
  - Hypothesis: concepts are interconnected via meaningful relations
  - Useful for reasoning
- The WordNet project started in 1986
  - Can most (all?) of the words in a language be represented as a semantic network where words are interlinked by meaning?
  - If so, the result would be a **large** semantic network…

Synonymy in WordNet

- WordNet is organized in terms of “synsets”
  - Unordered set of (roughly) synonymous “words” (or multi-word phrases)
- Each synset expresses a distinct meaning/concept

WordNet: Example

Noun

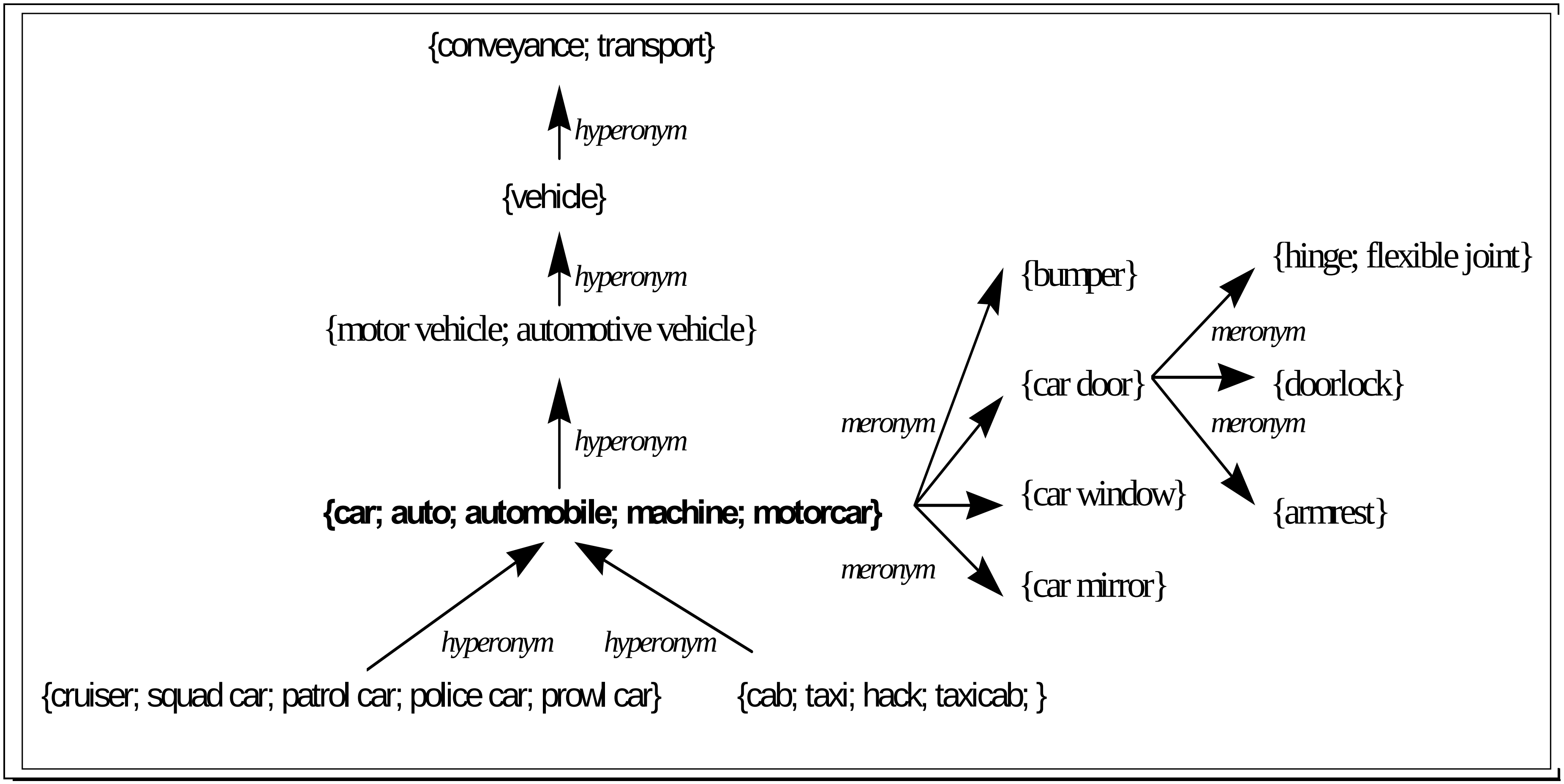
{pipe, tobacco pipe} (a tube with a small bowl at one end; used for smoking tobacco)
{pipe, pipage, piping} (a long tube made of metal or plastic that is used to carry water or oil or gas etc.)
{pipe, tube} (a hollow cylindrical shape)
{pipe} (a tubular wind instrument)
{organ pipe, pipe, pipework} (the flues and stops on a pipe organ)

Verb

{shriek, shrill, pipe up, pipe} (utter a shrill cry)
{pipe} (transport by pipeline) “pipe oil, water, and gas into the desert”
{pipe} (play on a pipe) “pipe a tune”
{pipe} (trim with piping) “pipe the skirt”

Observations about sense granularity?

The “Net” Part of WordNet



WordNet: Size

Part of speech	Word form	Synsets
Noun	117,798	82,115
Verb	11,529	13,767
Adjective	21,479	18,156
Adverb	4,481	3,621
Total	155,287	117,659

http://wordnet.princeton.edu/

MeSH

- Medical Subject Headings: another example of a theasuri
  - http://www.nlm.nih.gov/mesh/MBrowser.html
- Thesauri, ontologies, taxonomies, etc.

Word Similarity

Intuition of Semantic Similarity

Semantically close

- bank–money
- apple–fruit
- tree–forest
- bank–river
- pen–paper
- run–walk
- mistake–error
- car–wheel

Semantically distant

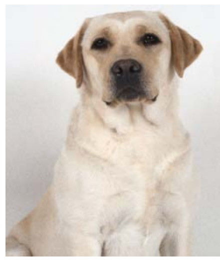
- doctor–beer
- painting–January
- money–river
- apple–penguin
- nurse–fruit
- pen–river
- clown–tramway
- car–algebra

Why?

- Meaning
  - The two concepts are close in terms of their meaning
- World knowledge
  - The two concepts have similar properties, often occur together, or occur in similar contexts
- Psychology
  - We often think of the two concepts together

Two Types of Relations

- Synonymy: two words are (roughly) interchangeable

- Semantic similarity (distance): somehow “related”
  - Sometimes, 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 of the word pairs
- Results:
  - Rank correlation between subjects is ~0.9
  - People are consistent!

Why do this?

- Task: automatically compute semantic similarity between words
- Theoretically useful for many applications:
  - Detecting paraphrases (i.e., automatic essay grading, plagiarism detection)
  - Information retrieval
  - Machine translation
  - …
- Solution in search of a problem?

Types of Evaluations

- Intrinsic
  - Internal to the task itself
  - With respect to some pre-defined criteria
- Extrinsic
  - Impact on end-to-end task

Analogy with cooking…

Evaluation: Correlation with Humans

- Ask automatic method to rank word pairs in order of semantic distance
- Compare this ranking with human-created ranking
- Measure correlation

Evaluation: Word-Choice Problems

Identify that alternative which is closest in meaning to the target:

accidental	**imprison**
wheedle	incarcerate
ferment	writhe
inadvertent	meander
abominate	inhibit

Evaluation: Malapropisms

Jack withdrew money from the ATM next to the band.

band is unrelated to all of the other words in its context…

Evaluation: Malapropisms

Jack withdrew money from the ATM next to the bank.

Wait, you mean bank?

Evaluation: Malapropisms

- Actually, semantic distance is a poor technique…
- What’s a simple, better solution?
- Even still, task can be used for a fair comparison

Word Similarity: Two Approaches

- Thesaurus-based
  - We’ve invested in all these resources… let’s exploit them!
- Distributional
  - Count words in context

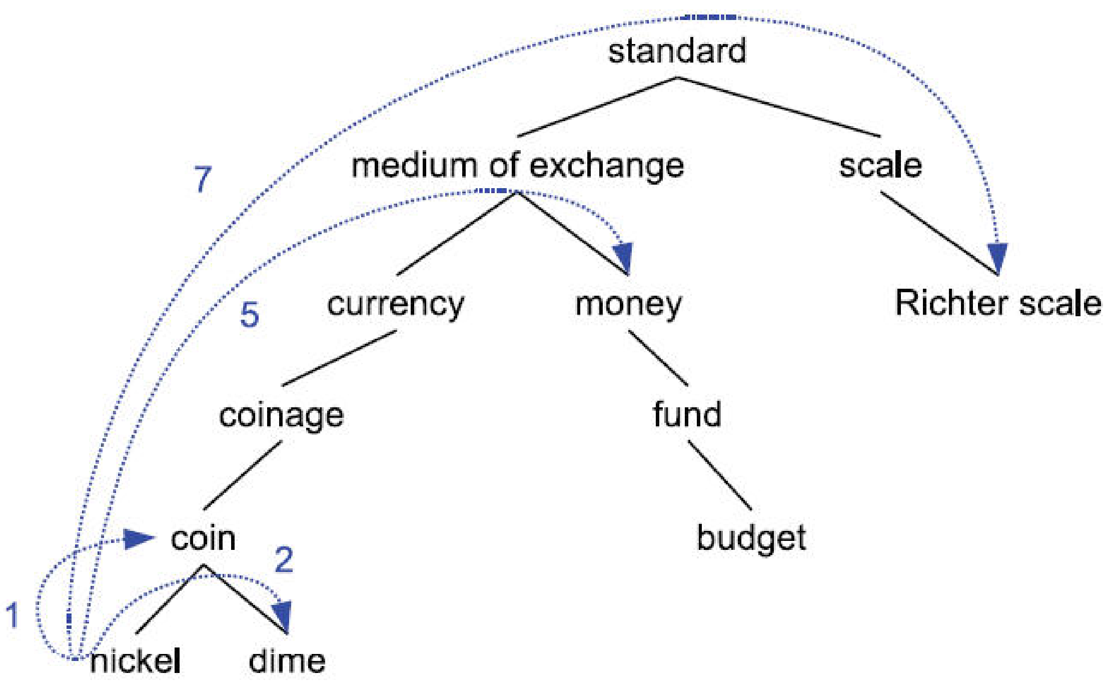
Word Similarity:
Thesaurus-Based Approaches

Note: In theory, applicable to any hierarchically-arranged lexical semantic resource, but most commonly applied to WordNet

Path-Length Similarity

- Similarity based on length of path between concepts:

$$\mathrm{sim}_{\mathrm{path}}(c_1, c_2) = -\log \mathrm{pathlen}(c_1, c_2)$$



Concepts vs. Words

- Similarity based on length of path between concepts

$$\text{sim}_{\text{path}}(c_1, c_2) = -\log \text{pathlen}(c_1, c_2)$$

- But which sense?
- Pick closest pair:

$$\text{sim}(w_1, w_2) = \max_{\substack{c_1 \in \text{senses}(w_1) \\ c_2 \in \text{senses}(w_2)}} \text{sim}(c_1, c_2)$$

- Similar techniques applied to all concept-based metrics

Wu-Palmer Method

- Similarity based on depth of nodes:

$$\text{sim}_{\text{Wu-Palmer}}(c_1, c_2) = \frac{2 \times \text{depth}(\text{LCS}(c_1, c_2))}{\text{depth}(c_1) + \text{depth}(c_2)}$$

  - LCS is the lowest common subsumer
  - depth(*c*) is the depth of node *c* in the hierarchy
- Explain the behavior of this similarity metric…
  - What if the LCS is close? Far?
  - What if c_1 and c_2 are at different levels in the hierarchy?

Edge-Counting Methods: Discussion

- Advantages
  - Simple, intuitive
  - Easy to implement
- Major disadvantage:
  - Assumes each edge has same semantic distance… not the case?

Resnik Method

- Probability that a randomly selected word in a corpus is an instance of concept *c*:

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

  - words(*c*) is the set of words subsumed by concept *c*
  - *N* is total number of words in corpus also in thesaurus

- Define "information content":

$$\text{IC}(c) = -\log P(c)$$

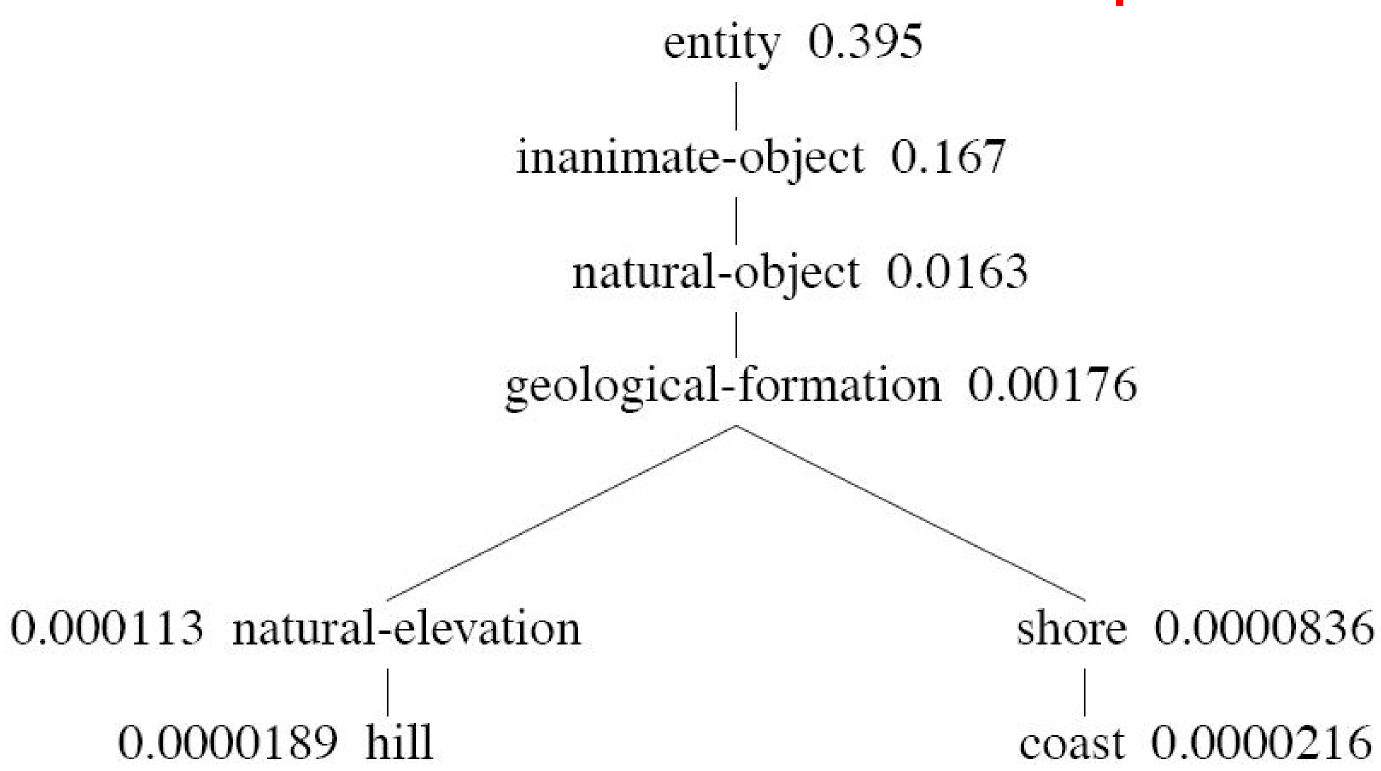
- Define similarity:

$$\text{sim}_{\text{Resnik}}(c_1, c_2) = -\log P(\text{LCS}(c_1, c_2))$$

Resnik Method: Example

$$\text{sim}_{\text{Resnik}}(c_1, c_2) = -\log P(\text{LCS}(c_1, c_2))$$

Explain its behavior…



Jiang-Conrath Distance

- Can we do better than the Resnik method?
- Intuition (duh?)
  - Commonality: the more A and B have in common, more similar they are
  - Difference: the more differences between A and B, the less similar they are
- Jiang-Conrath Distance:

$$\text{dist}_{\text{JC}}(c_1, c_2) = 2 \times \log P(\text{LCS}(c_1, c_2)) - (\log P(c_1) + \log P(c_2))$$

Explain its behavior…

  - Note: distance, not similarity!
  - Generally works well

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

Quick Aside: Thesauri Induction

- Building thesauri automatically?
- Pattern-based techniques work really well!

$NP\{, NP\} * \{,\}$ (and|or) other NP_H	... temples, treasuries, and other important civic buildings.
NP_H such as $\{NP,\}^*$ (or|and) NP	red algae such as Gelidium
such NP_H as $\{NP,\}^*$ (or|and) NP	works by such authors as Herrick, Goldsmith, and Shakespeare
NP_H $\{,\}$ including $\{NP,\}^*$ (or|and) NP	All common-law countries, including Canada and England
NP_H $\{,\}$ especially $\{NP,\}^*$ (or|and) NP	... most European countries, especially France, England, and Spain

  - Co-training between patterns and relations
  - Useful for augmenting/adapting existing resources

Word Similarity:
Distributional Approaches

Distributional Approaches: Intuition

- “You shall know a word by the company it keeps!” (Firth, 1957)
- Intuition:
  - If two words appear in the same context, then they must be similar
  - Watch out for antonymy!
- Basic idea: represent a word *w* as a feature vector

$$\vec{w} = (f_1, f_2, f_3, \ldots f_N)$$

  - Features represent the context…
- So what’s the context?

Context Features

- Word co-occurrence within a window:

	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

- Grammatical relations:

	subj-of, absorb	*subj-of*, adapt	*subj-of*, behave	...	*pobj-of*, inside	*pobj-of*, into	...	*nmod-of*, abnormality	*nmod-of*, anemia	*nmod-of*, architecture	...	*obj-of*, attack	*obj-of*, call	*obj-of*, come from	*obj-of*, decorate	...	*nmod*, bacteria	*nmod*, body	*nmod*, bone marrow
cell	1	1	1		16	30		3	8	1		6	11	3	2		3	2	2

Context Features

- Feature values
  - Boolean
  - Raw counts
  - Some other weighting scheme (e.g., *idf, tf.idf*)
  - Association values (next slide)
- Does anything from last week applicable here?

Association Metric

- Commonly-used metric: Pointwise Mutual Information

$$\text{association}_{\text{PMI}}(w, f) = \log_2 \frac{P(w, f)}{P(w)P(f)}$$

  - What's the interpretation?
- Can be used as a feature value or by itself

Cosine Distance

- Semantic similarity boils down to computing some measure on context vectors
- Cosine distance: borrowed from information retrieval

$$\mathrm{sim}_{\mathrm{cosine}}(\vec{v},\vec{w}) = \frac{\vec{v}\cdot\vec{w}}{|\vec{v}||\vec{w}|} = \frac{\sum_{i=1}^{N} v_i \times w_i}{\sqrt{\sum_{i=1}^{N} v_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}}$$

  - Interpretation?

Jaccard and Dice

- Jaccard

$$\text{sim}_{\text{Jaccard}}(\vec{v}, \vec{w}) = \frac{\sum_{i=1}^{N} \min(v_i, w_i)}{\sum_{i=1}^{N} \max(v_i, w_i)}$$

- Dice

$$\text{sim}_{\text{Dice}}(\vec{v}, \vec{w}) = \frac{2 \times \sum_{i=1}^{N} \min(v_i, w_i)}{\sum_{i=1}^{N} (v_i + w_i)}$$

Information-Theoretic Measures

- Kullback-Leibler divergence (aka relative entropy)

$$D(P \| Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$$

  - See any issues?
  - Note: asymmetric

- Jenson-Shannon divergence

$$JS(P \| Q) = D(P \| \frac{P+Q}{2}) + D(Q \| \frac{P+Q}{2})$$

Distributional Approaches: Evaluation

- Same as thesaurus-based approaches
- One additional method: use thesaurus as ground truth!

Distributional Approaches: Discussion

- No thesauri needed: data driven
- Can be applied to any pair of words
- Can be adapted to different domains

Distributional Profiles: Example

DP of *star*	**DP of *fusion***
space 0.21	*heat* 0.16
movie 0.16	*hydrogen* 0.16
famous 0.15	*energy* 0.13
light 0.12	*hot* 0.09
rich 0.11	*light* 0.09
heat 0.08	*space* 0.04
planet 0.07	*gravity* 0.03
hydrogen 0.07	*pressure* 0.03

Distributional Profiles: Example

DP of *star*	**DP of *fusion***
space 0.21	*heat* 0.16
movie 0.16	*hydrogen* 0.16
famous 0.15	*energy* 0.13
light 0.12	*hot* 0.09
rich 0.11	*light* 0.09
heat 0.08	*space* 0.04
planet 0.07	*gravity* 0.03
hydrogen 0.07	*pressure* 0.03

What's the problem?

DP of *star*	DP of *fusion*
space 0.21	*heat* 0.16
movie 0.16	*hydrogen* 0.16
famous 0.15	*energy* 0.13
light 0.12	*hot* 0.09
rich 0.11	*light* 0.09
heat 0.08	*space* 0.04
planet 0.07	*gravity* 0.03
hydrogen 0.07	*pressure* 0.03

Distributional Profiles of Concepts

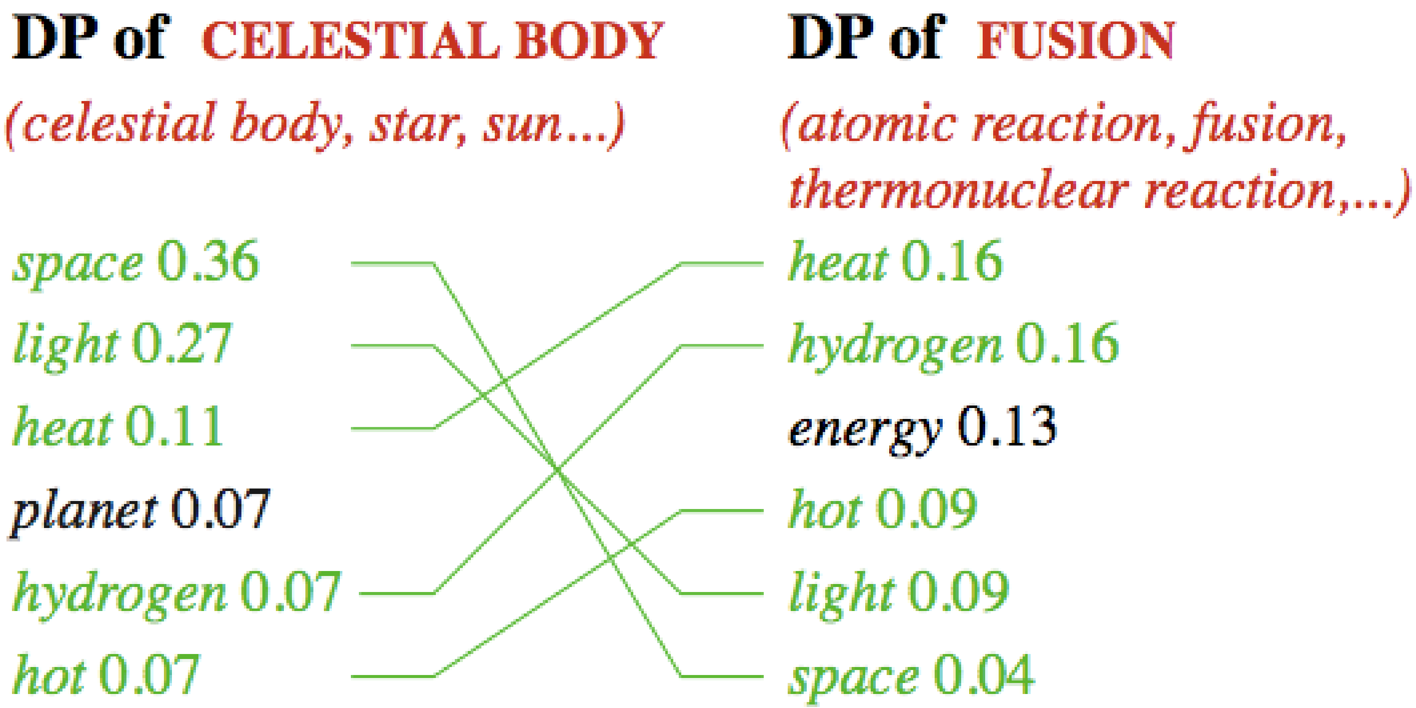
DP of CELESTIAL BODY	**DP of CELEBRITY**
(celestial body, star, sun,...)	*(celebrity, hero, star,...)*
space 0.36	*famous* 0.24
light 0.27	*movie* 0.14
heat 0.11	*rich* 0.14
planet 0.07	*fan* 0.10
hydrogen 0.06	*hot* 0.04
hot 0.01	*fashion* 0.01

Semantic Similarity: “Celebrity”

DP of CELEBRITY	**DP of FUSION**
(celebrity, hero, star,...)	*(atomic reaction, fusion, thermonuclear reaction,...)*
famous 0.24	*heat* 0.16
movie 0.14	*hydrogen* 0.16
rich 0.14	*energy* 0.13
fan 0.10	*hot* 0.09
hot 0.04	*light* 0.09
fashion 0.01	*space* 0.04

Semantically distant…

Semantic Similarity: “Celestial body”



Semantically close!

Solution?

- We need word sense disambiguation!
- Stay tuned for next week…

Recap: Today's Agenda

- Lexical semantic relations
- WordNet
- Computational approaches to word similarity