Word Senses and WordNet

(Thesaurus-based)

Word Senses and Relations

- Homonymy, Polysemy, Synonymy, and more
- Online Resources
- Thesaurus methods for word similarity

Lexical Semantics

- Focus on word meanings:
  - Relations of meaning among words
    - Similarities & differences of meaning in similar context
  - Internal meaning structure of words
    - Basic internal units combine for meaning

Word Definitions

- What's a word?
  - Definitions so far: Types, tokens, stems, roots, inflected 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

I. Possible Word Relations

- Homonymy
- Polysemy
- Synonymy
- Antonymy
- Hypernomy
- Hyponomy
- Meronomy

Homonymy

- Lexemes share a form
  - Phonological, orthographic or both
  - But have unrelated, distinct meanings
- Clear examples
  - Bat (wooden stick-like thing) vs. bat (flying scary mammal thing)
  - Bank (financial institution) versus bank (riverside)
- Can be homophones, homographs:
  - Homophones:
    - Write/right, piece/peace, to/too/two
  - Homographs:
    - Desert/desert
    - Bass/bass

Issues for NLP Applications

- Text-to-Speech
  - Same orthographic form but different phonological form
    - bass vs. bass
- Information retrieval
  - Different meanings same orthographic form
    - QUERY: bat care
- Machine Translation (English -> Spanish)
  - `bat:` murciélago (animal) or bate (for baseball)

Polysemy

- The **bank** is constructed from red brick
  I withdrew the money from the **bank**
  - Are these the same sense? Different?
- Or consider the following WSJ example
  - While some banks furnish sperm only to married women, others are less restrictive
  - Which sense of bank is this?
    - Is it distinct from the river bank sense?
    - The savings bank sense?

Polysemy

- A single lexeme with multiple related meanings (bank the building, bank the financial institution)
- Most non-rare words have multiple meanings
  - Number of meanings related to word frequency
  - Verbs tend more to polysemy
  - Distinguishing polysemy from homonymy isn't always easy (or necessary)

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`)

Metaphor vs. Metonymy

- Metaphor: two different meaning domains are related
  - Citibank claimed it was misrepresented.
  - Corporation as person
- Metonymy: use of one aspect of a concept to refer to other aspects of entity or to entity itself
  - The Citibank is on the corner of Main and State.
  - Building stands for organization

How Do We Identify Words with Multiple Senses?

- ATIS examples
  - Which flights *serve* breakfast?
  - Does America West *serve* Philadelphia?
- The "zeugma" test: conjoin two potentially similar/dissimilar senses
  - ?Does United serve breakfast and San Jose?
  - Does United serve breakfast and lunch?

Synonymy

- 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 successfully substituted for each other in all situations
  - If so they have the same **propositional meaning**

Few 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.
- E.g, water and H_20, coffee and java

Terminology

- Lemmas and wordforms
  - 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
  - Specific surface forms carpets, sung are called wordforms
- The lemma bank has two senses:
  - Instead, a bank can hold the investments in a custodial account in the client's name.
  - 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 Relates Senses not Words

- Consider *big* and *large*
- Are they synonyms?
  - How **big** is that plane?
  - Would I be flying on a **large** or a small plane?
- How about:
  - 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

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 an attribute at opposite ends of a scale (*long/short, fast/slow*)
  - Be **reversives**: *rise/fall, up/down*

Hyponyms

- A 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

Hypernymy Defined

- Extensional
  - The class denoted by the superordinate
  - Extensionally includes 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

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*.

20

II. WordNet

- A hierarchically organized lexical database
- On-line thesaurus + aspects of a dictionary
  - Versions for other languages are under development

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

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...`

22

WordNet Entries

The noun "bass" has 8 senses in WordNet.
1. bass^{1} - (the lowest part of the musical range)
2. bass^{2}, bass part^{1} - (the lowest part in polyphonic music)
3. bass^{3}, basso^{1} - (an adult male singer with the lowest voice)
4. sea bass^{1}, bass^{4} - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass^{1}, bass^{5} - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass^{6}, bass voice^{1}, basso^{2} - (the lowest adult male singing voice)
7. bass^{7} - (the member with the lowest range of a family of musical instruments)
8. bass^{8} - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

The adjective "bass" has 1 sense in WordNet.
1. bass^{1}, deep^{6} - (having or denoting a low vocal or instrumental range)
"a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"

WordNet Noun Relations

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

WordNet Verb Relations

Relation	Definition	Example
Hypernym	From events to superordinate events	$fly^9 \rightarrow travel^5$
Troponym	From a verb (event) to a specific manner elaboration of that verb	$walk^1 \rightarrow stroll^1$
Entails	From verbs (events) to the verbs (events) they entail	$snore^1 \rightarrow sleep^1$
Antonym	Opposites	$increase^1 \Longleftrightarrow decrease^1$

WordNet Hierarchies

```
Sense 3
bass, basso --
(an adult male singer with the lowest voice)
=> singer, vocalist, vocalizer, vocaliser
   => musician, instrumentalist, player
      => performer, performing artist
         => entertainer
            => person, individual, someone...
               => organism, being
                  => living 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 the lowest range of a family of
musical instruments)
=> musical instrument, instrument
   => device
      => instrumentality, instrumentation
         => artifact, artefact
            => whole, unit
               => object, physical object
                  => physical entity
                     => entity
```

How is ‘Sense’ Defined in WordNet?

- The set of near-synonyms for a WordNet sense is called a synset (synonym set); their version of a sense or a concept
- 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], sucker[1], soft touch[1], mug[2]}
- Each of these senses share this same gloss (but not every sense)
- For WordNet, the meaning of this sense of chump is this list.

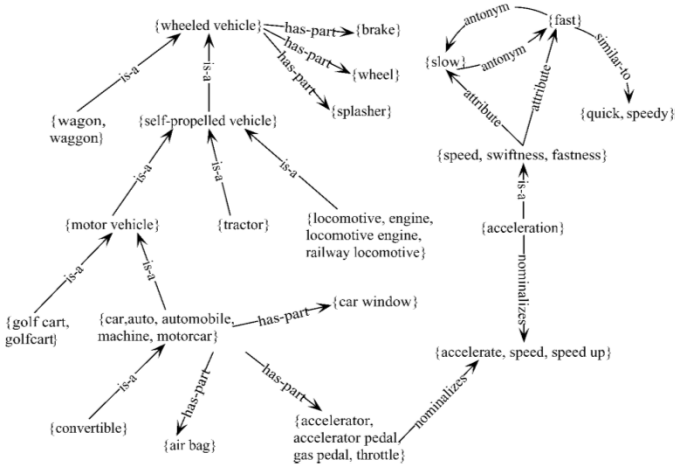
S: (n) **field** (a piece of land cleared of trees and usually enclosed)
S: (n) battlefield, battleground, field of battle, field of honor, **field** (a region where a battle is being (or has been) fought)
S: (n) **field** (somewhere (away from a studio or office or library or laboratory) where practical work is done or data is collected)
S: (n) discipline, subject, subject area, subject field, **field**, field of study, study, bailiwick (a branch of knowledge)
S: (n) **field**, field of force, force field (the space around a radiating body within which its electromagnetic oscillations can exert force on another similar body not in contact with it)
S: (n) **field**, field of operation, line of business (a particular kind of commercial enterprise)
S: (n) sphere, domain, area, orbit, **field**, arena (a particular environment or walk of life)
S: (n) playing field, athletic field, playing area, **field** (a piece of land prepared for playing a game)

Time flies.

(thanks to Dr. Wiebe for Allegheny Cemetery photos)

S: (v) fly, wing (travel through the air; be airborne) "Man cannot fly"
S: (v) fly (move quickly or suddenly) "He flew about the place"
S: (v) fly, aviate, pilot (operate an airplane) "The pilot flew to Cuba"
S: (v) fly (transport by aeroplane) "We fly flowers from the Caribbean to North America"
S: (v) fly (cause to fly or float) "fly a kite"
S: (v) fly (be dispersed or disseminated) "Rumors and accusations are flying"
S: (v) fly (change quickly from one emotional state to another) "fly into a rage"
S: (v) fly, fell, vanish (pass away rapidly) "Time flies like an arrow"; "Time fleeing beneath him"

WordNet: Viewed as a graph



32

"Supersenses"
The top level hypernyms in the hierarchy

A word's supersense can be a useful coarse-grained representation of word meaning for NLP tasks (counts from Schneider and Smith 2013's Streusel corpus)

Noun						Verb		
GROUP	1469	*place*	BODY	87	*hair*	STATIVE	2922	*is*
PERSON	1202	*people*	STATE	56	*pain*	COGNITION	1093	*know*
ARTIFACT	971	*car*	NATURAL OBJ.	54	*flower*	COMMUNIC.*	974	*recommend*
COGNITION	771	*way*	RELATION	35	*portion*	SOCIAL	944	*use*
FOOD	766	*food*	SUBSTANCE	34	*oil*	MOTION	602	*go*
ACT	700	*service*	FEELING	34	*discomfort*	POSSESSION	309	*pay*
LOCATION	638	*area*	PROCESS	28	*process*	CHANGE	274	*fix*
TIME	530	*day*	MOTIVE	25	*reason*	EMOTION	249	*love*
EVENT	431	*experience*	PHENOMENON	23	*result*	PERCEPTION	143	*see*
COMMUNIC.*	417	*review*	SHAPE	6	*square*	CONSUMPTION	93	*have*
POSSESSION	339	*price*	PLANT	5	*tree*	BODY	82	*get...done*
ATTRIBUTE	205	*quality*	OTHER	2	*stuff*	CREATION	64	*cook*
QUANTITY	102	*amount*				CONTACT	46	*put*
ANIMAL	88	*dog*				COMPETITION	11	*win*
						WEATHER	0	—

33

WordNet 3.1

- Where it is:
  - http://wordnetweb.princeton.edu/perl/webwn
- Libraries
  - Python: WordNet from NLTK
    - http://www.nltk.org/search.html?q=wordnet
  - Java:
    - JWNL, extJWNL on sourceforge

Other (domain specific) thesauri

MeSH: Medical Subject Headings
thesaurus from the National Library of Medicine

- **MeSH (Medical Subject Headings)**
  - 177,000 entry terms that correspond to 26,142 biomedical "headings"
- **Hemoglobins**

  Entry Terms: Eryhem, Ferrous Hemoglobin, Hemoglobin

  Synset

  Definition: The oxygen-carrying proteins of ERYTHROCYTES. They are found in all vertebrates and some invertebrates. The number of globin subunits in the hemoglobin quaternary structure differs between species. Structures range from monomeric to a variety of multimeric arrangements

The MeSH Hierarchy

1. + Anatomy [A]
2. + Organisms [B]
3. + Diseases [C]
4. − Chemicals and Drugs [D]
   - Inorganic Chemicals [D01] +
   - Organic Chemicals [D02] +
   - Heterocyclic Compounds [D03] +
   - Polycyclic Compounds [D04] +
   - Macromolecular Substances [D05] +
   - Hormones, Hormone Substitutes, an
   - Enzymes and Coenzymes [D08] +
   - Carbohydrates [D09] +
   - Lipids [D10] +
   - Amino Acids, Peptides, and Proteins
   - Nucleic Acids, Nucleotides, and Nucl
   - Complex Mixtures [D20] +
   - Biological Factors [D23] +
   - Biomedical and Dental Materials [D25] +
   - Pharmaceutical Preparations [D26] +
   - Chemical Actions and Uses [D27] +
5. + Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. + Psychiatry and Psychology [F]
7. + Phenomena and Processes [G]

Amino Acids, Peptides, and Proteins [D12]
Proteins [D12.776]
Blood Proteins [D12.776.124]
Acute-Phase Proteins [D12.776.124.050] +
Anion Exchange Protein 1, Erythrocyte [D12.776.124.078
Ankyrins [D12.776.124.080]
beta 2-Glycoprotein I [D12.776.124.117]
Blood Coagulation Factors [D12.776.124.125] +
Cholesterol Ester Transfer Proteins [D12.776.124.197]
Fibrin [D12.776.124.270] +
Glycophorin [D12.776.124.300]
Hemocyanin [D12.776.124.337]
➤ Hemoglobins [D12.776.124.400]
Carboxyhemoglobin [D12.776.124.400.141]
Erythrocruorins [D12.776.124.400.220]

37

Uses of the MeSH Ontology

- Provide synonyms ("entry terms")
  - E.g., glucose and dextrose
- Provide hypernyms (from the hierarchy)
  - E.g., glucose ISA monosaccharide
- Indexing in MEDLINE/PubMED database
  - NLM's bibliographic database:
    - 20 million journal articles
    - Each article hand-assigned 10-20 MeSH terms

III. 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

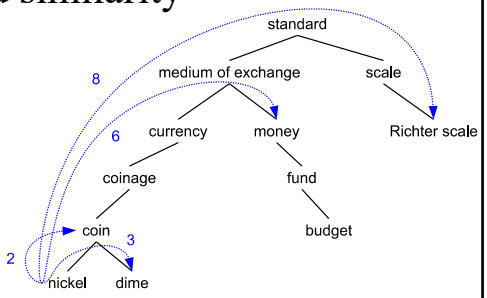
Word similarity and word relatedness

- 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

Two classes of 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?
  - Distributional (Vector) semantics (prior classes)

Path based similarity



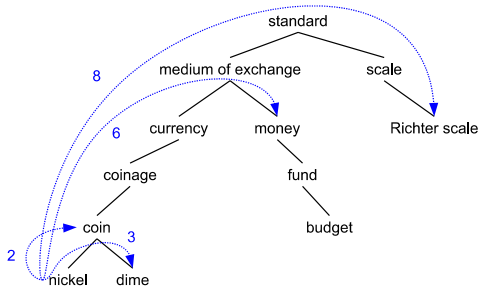
- Two concepts (senses/synsets) are similar if they are near each other in the thesaurus hierarchy
  - =have a short path between them
  - concepts have path 1 to themselves

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
- simpath(c_1,c_2) = $\frac{1}{\text{pathlen}(c_1, c_2)}$
- wordsim(w_1,w_2) = $\max_{c_1 \in \text{senses}(w_1), c_2 \in \text{senses}(w_2)} \text{sim}(c_1,c_2)$

Example: path-based similarity

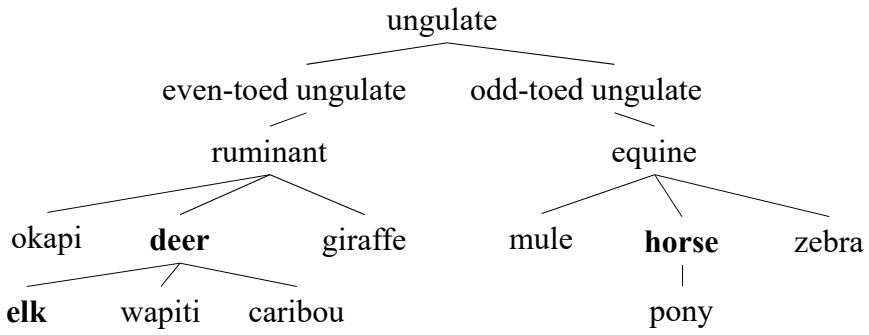
simpath(c_1,c_2) = 1/pathlen(c_1,c_2)



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(*coinage*,*Richter scale*) = 1/6 = .17

Which pair of words exhibits the greatest similarity?

1. Deer-elk
2. Deer-horse



Problem with basic path-based similarity

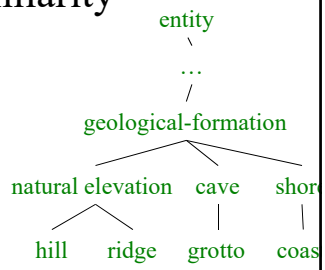
- Assumes each link represents a uniform distance
  - But *nickel* to *money* seems to us to be closer than *nickel* to *standard*
  - Nodes high in the hierarchy are very abstract
- We instead want a metric that
  - Represents the cost of each edge independently
  - Words connected only through abstract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



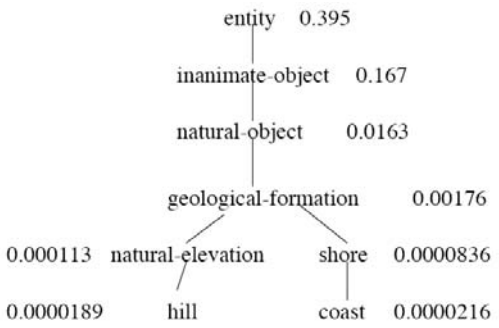
- Train by counting in a corpus
  - Each instance of `hill` counts toward frequency of *natural elevation*, *geological formation*, *entity*, etc
  - Let words(c) be the set of all words that are children of node c
    - words("geo-formation") = {hill,ridge,grotto,coast,cave,shore,natural elevation}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

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

Information content similarity

- WordNet hierarchy augmented with probabilities P(c)

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



Information content and probability

- The **self-information** of an event, also called its **surprisal:**
  - how surprised we are to know it; how much we learn by knowing it.
  - The more surprising something is, the more it tells us when it happens
  - We'll measure self-information in **bits**.

  I(w)= -log2 P(w)
- I flip a coin; P(heads)= 0.5
- How many bits of information do I learn by flipping it?
  - I(heads) = -log2(0.5) = -log2 (1/2) = log2 (2) = 1 bit
- I flip a biased coin: P(heads)= 0.8 I don't learn as much
  - I(heads) = -log2(0.8) = -log2(0.8) = .32 bits

51

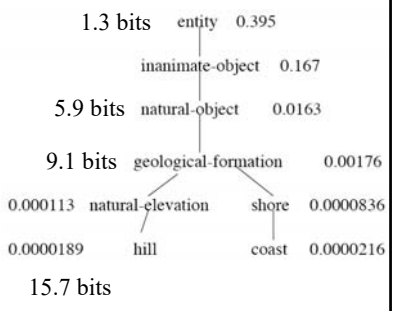
Information content: definitions

- 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
  - $sim_{resnik}(c_1,c_2) = -\log P(LCS(c_1,c_2))$

Dekang Lin method

Dekang Lin. 1998. An Information-Theoretic Definition of Similarity. ICM

- Intuition: Similarity between A and B is not just what they have in common
- The more **differences** between A and B, the less similar they are:
  - Commonality: the more A and B have in common, the more similar they are
  - Difference: the more differences between A and B, the less similar
- Commonality: IC(common(A,B))
- Difference: IC(description(A,B)-IC(common(A,B))

Dekang Lin similarity theorem

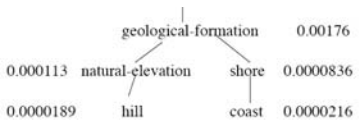
- The similarity between A and B is measured by the ratio between the amount of information needed to state the commonality of A and B and the information needed to fully describe what A and B are

$$sim_{Lin}(A,B) \propto \frac{IC(common(A,B))}{IC(description(A,B))}$$

- Lin (altering Resnik) defines IC(common(A,B)) as 2 x information of the LCS

$$sim_{Lin}(c_1,c_2) = \frac{2\log P(LCS(c_1,c_2))}{\log P(c_1)+\log P(c_2)}$$

Lin similarity



$$sim_{Lin}(A,B) = \frac{2\log P(LCS(c_1,c_2))}{\log P(c_1)+\log P(c_2)}$$

$$sim_{Lin}(\text{hill},\text{coast}) = \frac{2\log P(\text{geological-formation})}{\log P(\text{hill})+\log P(\text{coast})}$$

$$= \frac{2\ln 0.00176}{\ln 0.0000189 + \ln 0.0000216}$$

$$= .59$$

The (extended) Lesk Algorithm

- A thesaurus-based measure that looks at **glosses**
- Two concepts are similar if their glosses contain similar words
  - ***Drawing paper***: paper that is specially prepared for use in drafting
  - ***Decal***: the art of transferring designs from specially prepared paper to a wood or glass or metal surface

- For each *n*-word phrase that's in both glosses
  - Add a score of n^2
  - Paper and specially prepared for $1 + 2^2 = 5$
  - Compute overlap also for other relations
    - glosses of hypernyms and hyponyms

Summary: thesaurus-based similarity

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

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

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

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

Libraries for computing thesaurus-based similarity

- NLTK
  - http://nltk.github.com/api/nltk.corpus.reader.html?highlight=similarity - nltk.corpus.reader.WordNetCorpusReader.res_similarity
- WordNet::Similarity
  - http://wn-similarity.sourceforge.net/

59

Evaluating similarity

- Extrinsic (task-based, end-to-end) Evaluation:
  - Question Answering
  - Essay grading
- Intrinsic Evaluation:
  - Correlation between algorithm and human word similarity ratings
    - Wordsim353: 353 noun pairs rated 0-10. *sim(plane,car)=5.77*
  - Taking TOEFL multiple-choice vocabulary tests
    - `Levied is closest in meaning to:`
      `imposed, believed, requested, correlated`

Problems with thesaurus-based meaning

- Not every language has a thesaurus
- Even if we have a thesaurus, *recall* problems
  - Many words are missing
  - Most phrases are missing
  - Some connections between senses are missing
  - Adjectives and verbs have less structured hyponymy

Hyponymy and Other Relations

- Could we discover new relationhsips and add them to a taxonomy?
- Why – unknown word problem (at one time Microsoft or IBM, but not Google)

Hearst Approach

- Based on hand-built patterns
- E.g. *NP-0 such as NP-1* implies *hyponym (NP-1, NP-0)*

- Corpus-based pattern extraction (Snow, Jurafsky, Ng 2005)

Summary

- Lexical Semantics
  - Homonymy, Polysemy, Synonymy, etc.
- Computational resource for lexical semantics
  - WordNet, etc.
- Word Similarity
  - Thesaurus methods