Semantic Relatedness using Lexical Resources

Carlo Strapparava
FBK-Irst Istituto per la ricerca scientifica e tecnologica
I-38050 Povo, Trento, ITALY
strappa@fbk.eu

Outline

- A brief introduction of WordNet
- WordNet Domains, a lexical resource in which each sense is annotated with a domain label
- The problem to determine the semantic distance between two lexical expressed concepts
- The problem has a long history in philosophy, psychology, artificial intelligence, ...

⇒ Here we deal with the perspective of semantic relatedness of two lexemes in a lexical resource
⇒ E.g. “doctor” and “hospital” how much are similar?
WordNet: an on-line lexical database

- WordNet is an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory
- It was developed at Princeton University by George Miller’s team
- WordNet is a public domain resource http://wordnet.princeton.edu/

- English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept
- An attempt to organize lexical information in terms of word meanings, rather than word forms
- Different relations link the synonym sets
Psycholinguistic foundation

- Psycholinguistics is an interdisciplinary field of research concerned with the cognitive bases of linguistic competence
- Beginning with word association studies at the turn of the century and continuing down to the experiments of the past twenty years, psycholinguists have discovered many synchronic properties of the mental lexicon

Lexical semantics

- Lexical semantics is the conventional association between a lexicalized concept and an utterance that plays a syntactic role
- *Word form* refers to the physical utterance or inscription
- *Word meaning* refers to a lexicalized concept
## Lexical matrix

<table>
<thead>
<tr>
<th>Word Meanings</th>
<th>F₁</th>
<th>F₂</th>
<th>F₃</th>
<th>…</th>
<th>Fₙ</th>
</tr>
</thead>
<tbody>
<tr>
<td>M₁</td>
<td>E₁₁</td>
<td>E₁₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₂</td>
<td>E₂₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₃</td>
<td></td>
<td>E₃₃</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…</td>
<td></td>
<td></td>
<td></td>
<td>…</td>
<td></td>
</tr>
<tr>
<td>Mₘ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eₘₙ</td>
</tr>
</tbody>
</table>

- Mapping between word forms and meanings are many:many
- F₁ and F₂ are synonyms
- F₂ is polysemous
- Polysemy and synonymy are problems gaining access to information in the mental lexicon:
  - a listener/reader who recognizes a form must cope with polysemy
  - a speaker/writer who express a meaning must decide between synonyms

## The synsets

- How are word meanings represented in WordNet?
- Simulating the lexical matrix, M₁ can be represented by simply listing the word forms
  - M₁ = \{F₁, F₂, …\}
- Ex. \{board, plank\} and \{board, committe\}
- Given a word form, its senses are sorted (most frequent)
The synsets (2)

- Synsets do not explain what the concepts are; they merely signify that the concepts exist.
- Native (English) speakers are assumed to have already acquired the concepts, and are expected to recognize them from the list of words in the synsets.

=> extensional approach from a computer science point of view.

Different relations

- WordNet is organized by lexical and semantic relations.
- Lexical relation => between word forms.
- Semantics relation => between word meanings (synsets).
Lexical relations

- **Synonymy**
  - Leibniz: two expressions are synonymous if the substitution of one for the other does not alter the truth value of the sentence
  - => need to partition WordNet into nouns, verbs, adjectives, and adverbs

- **Antonymy** ex. [rich/poor] [rise/fall]
  - The antonym of a word \( x \) is sometimes \( \text{not-}x \), but not always: \( \text{not rich} \neq \text{poor} \)
  - Central organization principle for the adjectives

---

**Bipolar Adjective Structure**
Semantic relations (1)

- **Hyponymy/Hyperonymy** (the ISA relation)
  - A synset \( \{x_1, x_2, \ldots \} \) is an hyponym of the synset \( \{y_1, y_2, \ldots \} \) if native speakers accept sentences such as *An \( x \) is a (kind of) \( y \)*
  - Hyponymy is transitive and asymmetrical
  - In WordNet there is *normally* a single hyperonym (but there are some cases of multiple hyperonyms)
  - Hypo/Hyperonymy is the central organization principle of nouns

Semantic relations (2)

- **Meronymy/Holonymy** (the HASA relation)
  - A synset \( \{x_1, x_2, \ldots \} \) is a meronym of the synset \( \{y_1, y_2, \ldots \} \) if native speakers accept sentences such as *An \( x \) is a part of \( y \) or A \( y \) has an \( x \) (as a part)*
  - Meronymy is transitive and asymmetric and can be used to construct a *part-of* hierarchy
Semantic relations (3)

- Particular semantic relations in the verb hierarchy
  - **Troponym**: a verb expressing a specific manner elaboration of another verb
  - X is a *troponym* of Y if to X is to Y in some manner or *Y is a particular way to X*
  - **Entailment**: a verb X *entails* Y if X cannot be done unless Y is or has been done

A summary of WordNet relations

<table>
<thead>
<tr>
<th>Relations for nouns are:</th>
<th>Relations for verbs are:</th>
<th>Relations for adjectives are:</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="Antonym" alt="Antonym" /></td>
<td><img src="Antonym" alt="Antonym" /></td>
<td><img src="Antonym" alt="Antonym" /></td>
</tr>
<tr>
<td>@ <img src="Hyponym" alt="Hyponym" /></td>
<td>@ <img src="Hypernym" alt="Hypernym" /></td>
<td>&amp; ![Similar to](Similar to)</td>
</tr>
<tr>
<td>~ <img src="Hyponym" alt="Hyponym" /></td>
<td>~ <img src="Hyponym" alt="Hyponym" /></td>
<td>&lt; ![Participle of verb](Participle of verb)</td>
</tr>
<tr>
<td>#m ![Member meronym](Member meronym)</td>
<td># <img src="Entailment" alt="Entailment" /></td>
<td>\ <img src="Pertainym" alt="Pertainym (pertains to noun)" title="pertains to noun" /></td>
</tr>
<tr>
<td>#s ![Substance meronym](Substance meronym)</td>
<td>&gt; <img src="Cause" alt="Cause" /></td>
<td>^ ![Also see](Also see)</td>
</tr>
<tr>
<td>#p ![Part meronym](Part meronym)</td>
<td>^ ![Also see](Also see)</td>
<td>^ ![Also see](Also see)</td>
</tr>
<tr>
<td>%m ![Member holonym](Member holonym)</td>
<td>$ ![Verb Group](Verb Group)</td>
<td>$ ![Verb Group](Verb Group)</td>
</tr>
<tr>
<td>%s ![Substance holonym](Substance holonym)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>%p ![Part holonym](Part holonym)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>= <img src="Attribute" alt="Attribute" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Ex. read, snore, eat*
Unique Beginners (nouns)

- [entity, something, (anything having existence (living or nonliving))]  
- [psychological_feature, (a feature of the mental life of a living organism)]  
- [abstraction, (a general concept formed by extracting common features from specific examples)]  
- [state, (the way something is with respect to its main attributes; "the current state of knowledge"; "his state of health"; "in a weak financial state")  
- [event, (something that happens at a given place and time)]  
- [act, human_action, human_activity, (something that people do or cause to happen)]  
- [group, grouping, (any number of entities (members) considered as a unit)]  
- [possession, (anything owned or possessed)]  
- [phenomenon, (any state or process known through the senses rather than by intuition or reasoning)]

WordNet 1.6: some figures

- # total lemmata 129.509
  - Nouns 94.474
  - Verbs 10.319
  - Adjectives 20.170
  - Adverbs 4.546
- # total synsets 99.642
  - Nouns 66.025
  - Verbs 12.127
  - Adjectives 17.915
  - Adverbs 3.575
Other theories of lexical meaning

- Lexical decomposition
- Meaning postulates
- Prototypes
- Semantic networks

Lexical decomposition

- Word meaning = composition of meaning primitives
  - Ex. to buy (Jackendoff 1983)

\[
\text{GO} ([\text{FROM}[l], \text{TO}[l]])
\]
\[
[\text{EXCH}([\text{GO} ([\text{MONEY} [\text{FROM}[l], \text{TO}[l]])])]
\]
Meaning postulates

- Mental representations of the meaning relations between words (Fodor 1970)
  - Ex. *to buy*
    - `buy(x,y,z) → get (x,y,z)`
    - `buy(x,y,z) → pay (x,y,z)`
    - `buy(x,y,z) → choose (x,y)`
    - `buy(x,y,z) → sell (z,y,x)`
  - Ex. *bachelor*
    - `bachelor(x) → man(x) ∧ ¬ married(x)`

Prototypes

- Word meaning = information holding of the most typical examples of a concept but not necessarily of all instances (Rosch 1975)
  - Ex. *tiger*

<table>
<thead>
<tr>
<th>Property</th>
<th>Possible values</th>
<th>Default values</th>
</tr>
</thead>
<tbody>
<tr>
<td>is-a</td>
<td>feline</td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td>&lt; 180 Kg</td>
<td>120 Kg</td>
</tr>
<tr>
<td>height</td>
<td>&lt; 106 cm</td>
<td>80 cm</td>
</tr>
<tr>
<td>length</td>
<td>&lt; 250 cm</td>
<td>180 cm</td>
</tr>
<tr>
<td>colour</td>
<td>yellow with black stripes, white, black, yellow</td>
<td>yellow with black stripes</td>
</tr>
<tr>
<td>habitat</td>
<td>jungle, river banks</td>
<td>jungle</td>
</tr>
</tbody>
</table>
Semantic networks

- Word meaning = set of relations with other meanings of the lexicon (Quillian 1968)
  - Ex. to buy

What is not present in WordNet?

- WordNet contains very little syntax
  - Syntax would be very useful for verb subcategorizations
- Almost no links among the principal part-of-speeches
- Sometimes the semantic categorization is not accurate (or too accurate)
SemCor

- The SemCor corpus is a subset of the English Brown corpus containing almost 700,000 words.
- In SemCor all the words are tagged by PoS, and more than 200,000 content words are also lemmatized and sense-tagged according to WordNet 1.6.
- More in detail, the SemCor corpus is composed of 352 texts. In 186 texts all the open class words (nouns, verbs, adjectives, and adverbs) are annotated with PoS, lemma and sense, while in the remaining 166 texts only verbs are annotated with lemma and sense.
- The "all-words" component of SemCor has 359,732 tokens among which 192,639 are semantically annotated, while the "only-verbs" component has 316,814 tokens among which 41,497 verb occurrences are semantically annotated.

Extensions of WordNet

- Many research groups are extending WordNet in various aspects
  - => Multilinguality
  - Ex. MultiWordNet (IRST)
    - It is based on the assumption that the meaning networks already defined for the original English version may, for the most part, be reused for other languages.
- WordNet-Domains
- WordNet-Affect
- ...
MultiWordNet: multilingual extension of WordNet

http://multiwordnet.fbk.eu

An example of a multilingual synset

\{English: \{registration, enrollment\} \mid Italian: \{iscrizione, registrazione\}\}

\textit{hierarchy info: “list-of-parents”, “list-of-children”, etc…}\}
MultiSemCor

- MultiSemCor is an English/Italian parallel corpus, aligned at the word level and annotated with PoS, lemma and word sense.
- The procedure followed for creating MultiSemCor consists in:
  1. Getting the Italian translations of the SemCor texts (thanks to professional translators)
  2. Automatically aligning Italian and English texts at the sentence and word level
  3. Automatically transferring the word sense annotations from English to the aligned Italian words
- The final result is an Italian corpus annotated with PoS, lemma and word sense, but also an aligned parallel corpus lexically annotated with a shared inventory of word senses.

http://multisemcor.fbk.eu

Automatic discovery of WN relations

- (Marti Hearst 98) uses Lexico-Syntactic Pattern Extraction to mine text for hyponym relationships.
- This is important to reduce the enormous workload in the strictly intellectual compilation of the large lexical databases needed for NLP and retrieval tasks.
- The following examples illustrate two sample patterns (the implied hyponym relationships should be obvious):
  - red algae, such as Gelidium
  - bruises, broken bones, or other injuries
- Patterns can be hand-coded and/or identified automatically.
Automatic discovery of WN relations (cont.)

- A pattern discovery method:
  - Decide on a lexical relation that is of interest (e.g. meronymy)
  - Pick a list of word pairs form WN in which this relation holds (e.g. house-porch)
  - Extract sentences from large corpus in which these terms occurs, and record lexical and syntactic context
  - Find the commonalities among these context and hypothesize patterns

Uses of WordNet

- An on-line lexical database
- A basis to develop *sense disambiguation* tools (many approaches):
  - Considering the distances among synsets or some kind of navigation path in the hierarchy
    - Problem: there are sub-hierarchies that are less dense than other
  - Word Domain Disambiguation
- It is a valuable resource for many possible applications: a large variety of uses
Domain Labels

- We extended Wordnet1.6 adding domain labels (e.g. Medicine, Architecture, Sport) to noun synsets [Magnini and Cavaglià, LREC-2000]

- Objectives:
  - Provide a context for evaluating the quality of this resource
  - Verify the role domain labels may have in a word disambiguation task, in particular as far as sense clustering is concerned
  - Exploit the role of domain labels in a multilingual task

Domain Label Organization

- 250 Domain labels collected from dictionaries
- Four level hierarchy (Dewey Decimal Classification)

- 41 “basic” domains used for the experiment

![Diagram of domain labels and hierarchies]
Domain Labels Annotation in WordNet

- Integrate taxonomic and domain oriented information
  - Cross hierarchy
    - doctor#2 [Medicine] --> person#1
    - hospital#1 [Medicine] --> location#1
  - Cross category relations: operate#3 [Medicine]
  - Cross language information
- Reduce polysemy

Domains in the WN hierarchy
Polysemy Reduction

Semantic similarity measures

- The need to determine the *semantic distance* between two lexical expressed concepts is a problem that pervades the Computational Linguistics
- The problem has a long history in philosophy, psychology, artificial intelligence, ...

⇒ The perspective of semantic relatedness of two lexemes *in a lexical resource*
Terminology in the CL literature

- In the CL literature, many terms are used by different authors: *semantic relatedness, similarity, semantic distance*, ...

- Resnik (1995): “*car* and *gasoline* seem more closely related than *car* and *bicycles*, even if the second pairs are more similar”

- Morris and Hirst (2004) - relatedness covers various kinds of relations: meronymy, antonymy, functional association, and other non-classical relations
  - E.g. antonyms are dissimilar and hence distant, but strongly related

Terminology (cont.)

- Semantic distance in a lexical resource
  - with *concept* we refer to a particular sense of a given *word*
  - we are not talking about similarity of distributional behavior of words (e.g. Dagan, 2000)
  - When we refer to hierarchies and networks of concepts, we use the term *link* (or *edge*) to refer the relationships between nodes
Dictionary-based approaches

- Kozima and Furugori (1993) turned the Longman Dictionary of Contemporary Dictionary (LDOCE) into a network of nodes
- Creating a node for each headword and linking each headword to the nodes for all the words in its definition
- \( \text{Sim}_K^{\text{F}} \) between words is computed by means of spreading activation on this network

\[
\text{Sim}_K^{\text{F}}(\text{linguistics, stylistics}) = \text{Sim}_K^{\text{F}}(\{\text{the, study, of, language, in, general, and, of, particular, languages, and, their, structure, and, grammar, and, history}\}, \\
\{\text{the, study, of, style, in, written, or, spoken, language}\})
\]

Roget Thesauri Approaches

- Roget groups words in a structure based on categories, within which there are several levels of finer clustering and grouped in six primary classes
- While the categories are named, the finer division are not
- The users access through the index, which contains category numbers and labels
Roget thesaurus

- The category reference LOVE is a hypertext link to the entry love in the Category Index.
- **love, dislike**
- **like** noun: affection, sympathy, fellow-feeling, tenderness
- **dislike** noun: disapproval, disgust, dislike, drip, imposition, offense, prejudice, unwelcome, vociferation

=> Roget's 1000 categories
=> http://www.bartleby.com/110/index.html

Similarity on Roget thesaurus

- Methods of semantic similarity rely not only on the category structure but also on pointers within categories that cross-reference other categories.
- **Morris and Hirst (1991)** - two words are semantically related if
  - They have a category in common in their index entries
  - One has a category in its index that contains a pointer to a category of the other
  - One is either a label in the other’s index entry or is in a category of the other
  - They are both contained in the same subcategory
  - They both have categories in their index entries that point to a common category
Approaches using WordNet

- Mainly focused on noun hierarchy
- The backbone of noun hierarchy is the IS-A relation
- At the top of the hierarchy are 11 abstract concepts (unique beginners)
- The maximum depth is 16 nodes
- Beyond IS-A, also other relations: Part-of, Substance-of, ...

WordNet Approaches: terminology

- The length of the shortest path from synset \( c_i \) to \( c_j \) (measured in edges) is denoted by \( \text{len}(c_i,c_j) \)
- The depth of a node is the length from the node to the global root \( \text{depth}(c_i) = \text{len}(\text{root},c_i) \)
- \( \text{mscs}(c_i,c_j) \) is the most specific common subsumer of \( c_i \) and \( c_j \)
- Given any formula \( \text{rel}(c_i,c_j) \) between two concepts, \( \text{rel}(w_i,w_j) \) between two words is

\[
\text{rel}(w_i,w_j) = \max_{c_i \in s(w_i), c_j \in s(w_j)} \left[ \text{rel}(c_i,c_j) \right]
\]

where \( s(w_j) \) is the set of concepts that are senses of \( w_j \)
Most specific common subsumer

- $\text{mscs}(\text{nickel}, \text{dime}) = \text{coin}$
- $\text{mscs}(\text{nickel}, \text{credit card}) = \text{medium of exchange}$

Computing taxonomic path

- Compute the path between the concepts
- “The shorter the path, the more similar they are” (Resnik, 1995)
  - Two concepts are semantically related if the path connecting the synsets is not too long and does not change direction too often
Hirst and St-Onge

- Two words are strongly related if one of the following holds:
  1. They have a synset in common (e.g. human and person)
  2. They are associated with two different synsets that are connected by IS-A relation
  3. One of the words is compound that include the other (e.g. school and private school)

Hirst and St-Onge (cont.)

- Two words are said to be in regular relation if there exists an allowable path connecting a synset associated with each word (e.g. carrot and apple)
- A path is allowable if
  - it contains no more than five links
  - it conforms to one of eight patterns that take into consideration the direction variation - upward (hypernymy and meronymy), downward (hyponymy and holonymy) and horizontal (antonymy)

\[ rel_{HS}(c_1,c_2) = C - len(c_1,c_2) - k \times turns(c_1,c_2) \]

where \( C \) and \( k \) are constants (they used \( C = 8 \) and \( k = 1 \))
and \( turns(c_1,c_2) \) are the number of direction changes in the connecting path
Scaling the network

- The distances we have seen rely on the fact that links in the taxonomy represent uniform distances
- But this is not true: certain sub-taxonomies are much denser than others
  - rabbit_ears IS-A television_antenna => narrow distance
  - white_elephant IS-A possession => wide distance
  - but in both cases, only one links

Sussna’s depth-relative scaling

- Sussna (1997) idea: sibling concepts deep in a taxonomy are more closely related than those higher up
- Each relation has a weight or a range $[\min_r , \max_r]$ of weights associated (e.g. hyper/hyponymy, holonymy and meronymy have $\min_r = 1, \max_r = 2$)
- The weight of an edge of type $r$ from a node $c_i$ is reduced by a factor $\text{edges} (\cdot)$: the number of edges of the same type leaving $c_i$

\[
\text{weight}(c_i \rightarrow_r) = \max_r - \frac{\max_r - \min_r}{\text{edges}(c_i)}
\]
Sussna’s depth-relative scaling (2)

- The distance between two adjacent nodes is the average of the weights in each direction (i.e. $r'$ being the inverse of $r$)

$$\text{dist}_s(c_1, c_2) = \frac{\text{weight}(c_1 \rightarrow r) + \text{weight}(c_2 \rightarrow r')}{2 \times \max\{\text{depth}(c_1), \text{depth}(c_2)\}}$$

- The semantic distance between two arbitrary nodes is the sum of the distances of the adjacent nodes along the shortest path

Wu and Palmer’s conceptual similarity

- Wu and Palmer (1994) introduce a notion of conceptual similarity between a pair of concepts $c_1$ and $c_2$: 

$$\text{sim}_{wp} = \frac{2 \times \text{depth}(\text{mscs}(c_1, c_2))}{\text{len}(c_1, \text{mscs}(c_1, c_2)) + \text{len}(c_2, \text{mscs}(c_1, c_2)) + 2 \times \text{depth}(\text{mscs}(c_1, c_2))}$$

where $\text{depth}(\text{mscs}(c_1, c_2))$ is the `global’ depth in the hierarchy
Leacock and Chodorow

- Leacock and Chodorow normalized path length (1998)

\[
sim_{LC}(c_1, c_2) = -\log \frac{\text{len}(c_1, c_2)}{2 \times \max_{c \in \text{WordNet}} \text{depth}(c)}
\]

the denominator includes the maximum depth of the hierarchy

Information-based approaches

- Approaches incorporating an additional and qualitatively different knowledge source (e.g. information from a corpus)
  - Resnik’s information based approach
  - Jiang and Conrath’s combined approach
  - Lin’s universal similarity measure
Resnik’s information-based approach

- In IS-A taxonomy => inspecting the relative position of the most specific concept that subsumes the concepts
- But if the minimal path of IS-A links is too long, it is often necessary to go very high in the taxonomy

```
mscs(nickel,dime) = coin
mscs(nickel,credit_card) = medium_of_exchange
```

**Resnik’s information-based approach**

- For any concept $c$ in the taxonomy, let $p(c)$ be the probability of encountering an instance of $c$
- The information content of $c$, $IC(c)$, is $-\log p(c)$
- The semantic similarity of a pair of concept $c_1$ and $c_2$ is

\[
sim_R = -\log p(mscs(c_1,c_2)) = IC(mscs(c_1,c_2))
\]

- If $c_1 IS-A c_2$ then $p(c_j) \leq p(c_i)$
  - e.g. whenever we have encountered a nickel we have encountered a coin, so $p(nickel) \leq p(coin)$
- The higher the position of mscs of two concepts, the lower their similarity
- If the mscs is the top, the similarity will be $-\log(1) = 0$
Resnik’s information-based approach

- Find the probability of a concept:
  - Through an estimation of noun frequencies from Brown Corpus of American English
  - How to do the counting: any noun in the corpus was counted as an occurrence of each taxonomic class containing it (e.g. an occurrence of noun *nickel* was counted towards the frequency of *nickel*, *coin*, etc.)
  - As a consequence of using non-disambiguated data, a polysemous word contributes to the counts of all its senses.

\[ p(c) = \frac{\sum_{w \in \text{words}(c)} \text{count}(w)}{N} \]

- where \( \text{words}(c) \) is the set of nouns in the corpus whose senses are subsumed by \( c \), and \( N \) is the total number of nouns in the corpus also present in WordNet

Resnik’s information-based approach

- This rather selective use of the taxonomy has its drawbacks:
  - In terms of semantic distance, two concepts with the same mscs are indistinguishable
  - \( \text{sim}_R(\text{money}, \text{credit}) = \text{sim}_R(\text{dime}, \text{credit\_card}) \) being in each case mscs is \text{medium\_of\_exchange}
  - For a method edge-based (such as Leacock and Chodorow) this is not

<table>
<thead>
<tr>
<th>Word1</th>
<th>Word2</th>
<th>Similarity</th>
<th>Most Informative Subsumer</th>
</tr>
</thead>
<tbody>
<tr>
<td>doctor</td>
<td>nurse</td>
<td>9.4823</td>
<td>health professional</td>
</tr>
<tr>
<td>doctor</td>
<td>lawyer</td>
<td>7.2240</td>
<td>professional, person</td>
</tr>
<tr>
<td>doctor</td>
<td>man</td>
<td>2.9683</td>
<td>person individual</td>
</tr>
<tr>
<td>doctor</td>
<td>medicine</td>
<td>1.0105</td>
<td>entity</td>
</tr>
<tr>
<td>doctor</td>
<td>hospital</td>
<td>1.0105</td>
<td>entity</td>
</tr>
<tr>
<td>doctor</td>
<td>health</td>
<td>0.0</td>
<td>Virtual root</td>
</tr>
<tr>
<td>doctor</td>
<td>sickness</td>
<td>0.0</td>
<td>Virtual root</td>
</tr>
</tbody>
</table>
Disambiguation of Word Sets

- Horse => has six senses
- Queen => has seven senses
- King => has six senses

⇒ There are 252 possible combinations!

The algorithm assigns the best score to the following senses:

{king} - weakest but most important chess piece
{queen} - the most powerful chess piece
{knight, horse} - a chessman in the shape of horse’s head

Jiang and Conrath’s combined approach

- Jiang and Conrath (1997) also use the notion of information content
- They take into account the information content of the two nodes as well that of their mscs

\[
\text{dist}_{IC} = IC(c_1) + IC(c_2) - 2 \times IC(\text{mscs}(c_1, c_2)) \\
= 2\log p(\text{mcs}(c_1, c_2)) - (\log p(c_1) + \log p(c_2))
\]
Lin’s universal similarity measure

- Lin (1998) tried to define a similarity measure applicable to arbitrary objects and “not presuming a particular form of knowledge representation” and “derived form a set of assumptions”
- Given two objects A and B:
  - The more commonality they share, the more similar they are
  - The more differences they have, the less similar they are
  - The maximum similarity is reached when A and B are identical

Lin defines:
- the commonality between A and B as the information content of “the preposition that states the commonalities”
  \[ IC(\text{comm}(A,B)) \]
- the difference between A and B as
  \[ IC(\text{descr}(A,B)) - IC(\text{comm}(A,B)) \]
  where descr(A,B) is the preposition describing A and B
- Similarity theorem: the similarity is given by the ratio between the amount of information needed to state the commonality and the information needed to fully describe them

\[
\text{sim}_{L} = \frac{\log p(\text{comm}(A,B))}{\log p(\text{descr}(A,B))} \Rightarrow \text{sim}_{L} = \frac{2 \times \log p(\text{mscs}(c_{1},c_{2}))}{\log p(c_{1}) + \log p(c_{2})}
\]
Evaluation

- Rubenstein and Goodenough (1965) obtained *synonymy judgments* from 51 human subjects on 65 pairs of words.

- The pairs ranged from high synonymous (*gem-jewel*) to semantically unrelated (*noon-string*).

- The subjects were asked to rate them on a scale of 0.0 to 4.0 according to their similarity of meaning.

- Miller and Charles (1991) extracted 30 pairs from the original 65, and then obtained similarity judgments from 38 subjects.

### Miller-Charles set of word pairs

<table>
<thead>
<tr>
<th>#</th>
<th>Pair</th>
<th>Humans</th>
<th>Msim</th>
<th>dbrk</th>
<th>simb</th>
<th>simc</th>
<th>simd</th>
<th>sime</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cat</td>
<td>automobile</td>
<td>3.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>gem</td>
<td>jewel</td>
<td>3.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>journey</td>
<td>voyage</td>
<td>3.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>boy</td>
<td>lad</td>
<td>3.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>coast</td>
<td>shore</td>
<td>3.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>asylum</td>
<td>madhouse</td>
<td>3.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>magician</td>
<td>wizard</td>
<td>3.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>midnight</td>
<td>noon</td>
<td>3.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>furnace</td>
<td>stove</td>
<td>3.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>food</td>
<td>fruit</td>
<td>3.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>bird</td>
<td>cock</td>
<td>3.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>bird</td>
<td>canoe</td>
<td>2.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>boot</td>
<td>implement</td>
<td>2.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>brother</td>
<td>monk</td>
<td>2.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>lad</td>
<td>brother</td>
<td>1.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>crime</td>
<td>implement</td>
<td>1.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>journey</td>
<td>car</td>
<td>1.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>monk</td>
<td>oracle</td>
<td>1.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>cemetery</td>
<td>woodland</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>food</td>
<td>rooster</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>coast</td>
<td>hall</td>
<td>0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>forest</td>
<td>graveyard</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>shore</td>
<td>woodland</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>monk</td>
<td>slave</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>coast</td>
<td>forest</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>lad</td>
<td>wizard</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>closed</td>
<td>smile</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>glass</td>
<td>magician</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>rooster</td>
<td>voyage</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>noon</td>
<td>string</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The calculation is left as an exercise*
Evaluation (cont.)

- Miller and Charles calculates an upper bound (referring to human performance) of .884.

- Coefficients of correlation between human rating of similarity and five computational measures:

<table>
<thead>
<tr>
<th>Measure</th>
<th>M&amp;C</th>
<th>R&amp;G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hirst and St-Onge, $rel_{JS}$</td>
<td>.744</td>
<td>.786</td>
</tr>
<tr>
<td>Jiang and Conrath, $dist_{JC}$</td>
<td>.850</td>
<td>.781</td>
</tr>
<tr>
<td>Leacock and Chodorow, $sim_{LC}$</td>
<td>.816</td>
<td>.838</td>
</tr>
<tr>
<td>Lin, $sim_{L}$</td>
<td>.829</td>
<td>.819</td>
</tr>
<tr>
<td>Resnik, $sim_{R}$</td>
<td>.774</td>
<td>.779</td>
</tr>
</tbody>
</table>

An application-based evaluation

- **Malapropism**: the unintentional misuse of a word by confusion with one that sounds similar (e.g. *the sacred and the propane* -- instead of *profane*).

- Hirst and St-Onge (1998): malapropism detection as a testbed for the evaluation of similarity measures.

- Data: 500 articles from *Wall Street Journal*, and they replaced one word every 200 with a spelling variation.

- A final corpus with 107,233 words, 1408 of which were malapropism.

- Malapropism detection evaluated in terms of precision, recall and F-measure.

- Jiang and Conrath slightly outperform the other measures (then Lin, Leacock and Chodoow, Resnik, and finally Hist and St-Onge).
References

- Dagan - Contextual word similarity - In Handbook of Natural Language Processing, Marcel Dekker Inc. - 2000
- Jiang, Conrath - Semantic similarity based on corpus statistics and lexical taxonomy - ROCLING X - 1997
- Kozima, Furugori - Similarity between words computed by spreading activation on an English dictionary - EACL 93, 1993
- Hirst, St-Onge - Lexical chains as representation of context for detection and correction of malapropism - In Fellbaum (ed) WordNet - MIT press 1998
- Miller, Charles - Contextual correlates of semantic similarity - Language and Cognitive Processes 6(1) 1991
- Morris, Hirst - Lexical cohesion computed by thesaural relations as an indicator of the structure of text - Computational Linguistics 17(1), March 1991
- Morris, Hirst - Non-classical lexical semantic relations - In Workshop of Computational Lexical Semantics NAACL May 2004
- Resnik - Using information content to evaluate semantic similarity - In Proceedings of IJCAI95 - August 1995
- Rubenstein, Goodenough - Contextual correlates of synonymy - Communication of ACM 8(10) 1965
- Sussna - Text retrieval using inference in semantic metanetworks - PhD thesis University of California, San Diego 1997
- Wu, Palmer - Verb semantics and lexical selection - In proceedings of ACL 94 - June 1994

Semantic Relatedness using Lexical Resources

Laboratory/Project

Carlo Strapparava
FBK-Irst Istituto per la ricerca scientifica e tecnologica
I-38050 Povo, Trento, ITALY
strappa@fbk.eu
Task

- Testing the semantic relatedness of set of word pairs
- Data
  - Rubenstein and Goodenough (1965) obtained synonymy judgments from 51 human subjects on 65 pairs of words
  - The pairs range from high synonymous to semantically unrelated
  - The humans rated on a scale of 0 to 4
  - Miller and Charles (1991) chose 30 pairs from the original 65, 10 from high level (between 3-4), 10 from intermediate level (between 1-3), and 10 from low level (between 0-1) of semantic similarity

Task

- Considering five implemented similarity measures, calculate similarity or related scores
- When either or both of the words has more than one synset in WordNet, took the most-related pair of synsets
- The proposed measures are:
  - Hirst StOnge
  - Jiang Conrath
  - Leacock Chodorow
  - Lin
  - Resnik
Tool

- There is a package (*WordNet::Similarity*) developed by Ted Pedersen
- This package consists of Perl modules that implement the semantic relatedness measures described at lesson
- There is a web interface at [http://maraca.d.umn.edu/cgi-bin/similarity/similarity.cgi](http://maraca.d.umn.edu/cgi-bin/similarity/similarity.cgi)

- Try also with *Latent Semantic Analysis* tool
- There is a web interface at [http://lsa.colorado.edu/](http://lsa.colorado.edu/)

List of CL conferences and journals

- Association of Computational Linguistics (ACL)

- *Journals & Conferences*
  ex. a list of journals & conferences