LING 438/538
Computational Linguistics

Sandiway Fong
Lecture 25: 11/29
Administrivia

• **Homeworks**
  – Britney Spears homework
    • graded (email went out last night)

  – Grammar homework
    • those of you who haven’t submitted it yet
    • tonight
Administrivia

• Reminders
  – 538 Presentations next Tuesday
    • slides due Monday night

  – Next Thursday - (Optional) review class
Chapters (2000 edition)

• II: Syntax
  • 11 Features and Unification – Jeff Berry
  • 12 Lexicalized and Probabilistic Parsing – Mary Dungan (human parsing)
  • 13 Language and Complexity – Roeland Hancock

• III: Semantics
  • 14 Representing Meaning – Mark Siner (FOPC)
  • 15 Semantic Analysis – Sean Humphreys (named entity recognition)
  • 16 Lexical Semantics – Kevin Moffit
  • 17 Word Sense Disambiguation and Information Retrieval – Sunjing Ji – HsinMin Lu
Chapters (2000 edition)

• IV: Pragmatics
  • 18 Discourse
  • 19 Dialog and Conversational Agents
  • 20 Natural Language Generation
    – Brent Ramerth
• V: Multilingual Processing
• 21 Machine Translation
  – Dainon Woudstra

• Other
  – Biomedical Information
    • Tara Paulsen
New Topic

• **Semantic Inference and Language**
  – *A step outside syntax*

• let’s look at computation using
  – **WordNet**
    • project led by a psychologist
      – George Miller (Princeton University)
    • handbuilt network of synonym sets (**synsets**) with semantic relations connecting them
    • very popular in computational linguistics
WordNet

- **what is it?**
  - synonym set (**synset**) network for nouns, verbs, adjectives and adverbs
  - synsets connected by semantic relations (**isa**, **antonymy**,...)
  - 139,000 entries (**word senses**)
  - 10,000 verbs (polysemy 2), 20,000 adjectives (1.5)
  - *originally conceived as a model of human semantic memory*

```
      conveyance, transport
        /          \
       /            \
      motor vehicle, ski tow, ski lift
        /                    \                      \
      car, auto, motorcar, go-cart
        /                      \                        \
      convertible, jeep, limo
                                           \               
                                          hypernymy (isa)
```
WordNet

- **what is it?**
  - synonym set (synset) network for nouns, verbs, adjectives and adverbs
  - synsets connected by semantic relations (isa, antonymy,...)
  - 139,000 entries (word senses), 10,000 verbs (polysemy 2), 20,000 adjectives (1.5)
  - *originally conceived as a model of human semantic memory*

![Diagram of WordNet relationships]
WordNet

- **what is it?**
  - synonym set (**synset**) network for nouns, verbs, adjectives and adverbs
  - synsets connected by semantic relations (**isa**, **antonymy**, ...)
  - 139,000 entries (**word senses**), 10,000 verbs (polysemy 2), 20,000 adjectives (1.5)
  - *originally conceived as a model of human semantic memory*

```
Adjectives

<table>
<thead>
<tr>
<th>Small</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>dinky</td>
<td>ample</td>
</tr>
<tr>
<td>elfin</td>
<td>bulky</td>
</tr>
<tr>
<td>gnomish</td>
<td>colossal</td>
</tr>
<tr>
<td>micro</td>
<td>gigantic</td>
</tr>
<tr>
<td>minuscule</td>
<td>humongous</td>
</tr>
<tr>
<td>puny</td>
<td>spacious</td>
</tr>
<tr>
<td>undersized</td>
<td>titantic</td>
</tr>
<tr>
<td>satellites</td>
<td>antonymy</td>
</tr>
</tbody>
</table>
```

“Dumbell Model”
WordNet

- extremely popular resource for language applications
  - freely available
    - (project cost $3M)
  - many Wordnets based on the Princeton system have been proposed or built for other languages
    - EuroWordNet (EWN)
    - ItalWordNet
    - Tamil WordNet
    - Estonian WordNet
WordNet Applications

• examples
  – information retrieval and extraction
    • query term expansion (synonyms etc.)
    • cross-linguistic information retrieval (multilingual WordNet)
  – concept identification in natural language
    • word sense disambiguation
    • WordNet senses and ontology (isa-hierarchy)
  – semantic distance computation
    • relatedness of words
  – document structuring and categorization
    • determine genre of a paper (WordNet verb categories)
Using WordNet Online

- URL: http://wordnet.princeton.edu/

WordNet - a lexical database for the English language

Cognitive Science Laboratory  Princeton University  221 Nassau St.  Princeton, NJ 08542


Word to search for: [Input field]  Search WordNet

Display Options: [Select option to change]  Change

WordNet home page

current version: 3.0
## WordNet Relations

<table>
<thead>
<tr>
<th>Relation</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>x HYP y</td>
<td>x is a hypernym of y</td>
<td>x: repair, y: improve</td>
</tr>
<tr>
<td>x ENT y</td>
<td>x entails y</td>
<td>x: breathe, y: inhale</td>
</tr>
<tr>
<td>x SIM y</td>
<td>y is similar to x (A)</td>
<td>x: achromatic, y: white</td>
</tr>
<tr>
<td>x CS y</td>
<td>y is a cause of x</td>
<td>x: anesthetize, y: sleep</td>
</tr>
<tr>
<td>x VGP y</td>
<td>y is similar to x (V)</td>
<td>x: behave, y: pretend</td>
</tr>
<tr>
<td>x ANT y</td>
<td>x and y are antonyms</td>
<td>x: present, y: absent</td>
</tr>
<tr>
<td>x SA y</td>
<td>x, see also y</td>
<td>x: breathe, y: breathe out</td>
</tr>
<tr>
<td>x PPL y</td>
<td>x participle of y</td>
<td>x: applied, y: apply</td>
</tr>
<tr>
<td>x PER y</td>
<td>x pertains to y</td>
<td>x: abaxial, y: axial</td>
</tr>
</tbody>
</table>
Semantic Opposition

• *Event-based Models of Change and Persistence in Language* (Pustejovsky, 2000):
  – John mended the *torn* dress
  – John mended the *red* dress

  – what kind of knowledge is invoked here?

• from Artificial Intelligence (AI):
  – an instance of the *frame problem*
  – *aka* the *update problem*
  – computation:
    • *what changes in the world and what doesn’t?*
Persistence and Change of State Verbs

- Event-based Models of Change and Persistence in Language (Pustejovsky, 2000):
  - John mended the *torn* dress
  - John mended the *red* dress

- **Verb Classes**: *Aspectual Classes* (Vendler 1967)
  - Mary cleaned the *dirty* table
  - The waiter filled every *empty* glass
  - Mary fixed the *flat* tire
  - Bill swept the *dirty* floor
  - Bill swept the *dirty* floor clean
  - Nero built the *gleaming* temple
  - Nero ruined the *splendid* temple
Event Representation

- **Change of State Verbs:**
  - John mended the **torn/red** dress
  - **mend**: x CAUS y BECOME <STATE (mended)>
  - John CAUS the **torn/red** dress BECOME <STATE (mended)>

- **antonym** relation between adjective and the end state
Using WordNet Online

• Let’s explore *mend*:

Noun

- **S: (n) mend, patch, darn** (sewing that repairs a worn or torn hole (especially in a garment)) *"her stockings had several mends"
- **S: (n) repair, fix, fixing, fixture, mend, mending, reparation** (the act of putting something in working order again)

Verb

- **S: (v) repair, mend, fix, bushel, doctor, furbish up, restore, touch on** (restore by replacing a part or putting together what is torn or broken) *"She repaired her TV set": "Repair my shoes please"
- **S: (v) mend, heal** (heal or recover) *"My broken leg is mending"*
Using WordNet Online

• Let’s explore mend:

  • S: (v) repair, mend, fix, bushel, doctor, furnish up, restore, touch on (restore by replacing a part broken) "She repaired her TV set"; "Repair my shoes please"
    - direct troponym / full troponym
    - direct hypernym / inherited hypernym / sister term
    - antonym
  
  • W: (v) break [Opposed to: repair] (ruin completely) "He busted my radio!"
    - phrasal verb
    - antonym
    - derivationally related form
    - Overview
    - synset
      - S: (v) break, bust (ruin completely) "He busted my radio!"
        - derivationally related form
        - sentence frame

  • S: (v) mend, heal (heal or recover) "My broken leg is mending"
Using WordNet Online

Let's explore *mend*:

**Word to search for:** bust

**Display Options:**

Key: "S." = Show Synset (semantic) relations, "W." = Show Word (lexical) relations

**Noun**

- S: (n) **flop**, bust, fizzle (a complete failure) "the play was a dismal flop"
- S: (n) **female chest**, bust (the chest of a woman)
- S: (n) **bust** (a sculpture of the head and shoulders of a person)
- S: (n) **bust**, tear, binge, bout (an occasion for excessive eating or drinking) "they went on a bust that lasted three days"

**Verb**

- S: (v) **break**, bust (ruin completely) "He busied my radio!"
- S: (v) **raid**, bust (search without warning, make a sudden surprise attack on) "The police raided the crack house"
- S: (v) **tear**, rupture, snap, bust (separate or cause to separate abruptly) "The rope snapped"; "tear the paper"
- S: (v) **break**, wear, wear out, bust, fall apart (go to pieces) "The lawn mower finally broke"; "The gears wore out"; "The old chair finally fell apart completely"
- S: (v) **burst**, bust (break open or apart suddenly and forcefully) "The dam burst"
Using wnconnect

• A program on my homepage
  – [http://dingo.sbs.arizona.edu/~sandiway/wnconnect/index.html](http://dingo.sbs.arizona.edu/~sandiway/wnconnect/index.html)
  – written in 2003 (for a research paper)
  – uses an older version of WordNet (1.6)
  – graphs connections
Using wnconnect

- Find shortest link between mend and tear in WordNet:
Using wnconnect

- Find shortest link between mend and tear in WordNet:
  - mend/v is in [repair, mend, fix, bushel, doctor, furbish_up, restore, touch_on]
  - repair and break/v are antonyms
  - bust in [break, bust] and bust/v related by verb.contact
  - tear/v is in the synset [tear, rupture, snap, bust]
Using wnconnect

two senses of **bust**: (1) to ruin completely, (2) to separate or cause to separate abruptly
Using wnconnect

- John CAUS the red dress BECOME <STATE (mended)>

mend/n is in [mend,patch]
mend,patch is an instance of [sewing,stickery]
[sewing,stickery] is an instance of [needlework,needlecraft]
[needlework,needlecraft] is an instance of [creation]
[creation] is an instance of [artifact,artefact]
[artifact,artefact] is an instance of [object,physical_object]
[object,physical_object] is an instance of [entity,physical_thing]
[causal_agent,cause,causal_agency] is an instance of [entity,physical_thing]
[person,individual,someone,somebody,mortal,human,soul] is an instance of [causal_agent,cause,
disputant,controversialist] is an instance of [person,individual,someone,somebody,mortal,human,
[radical] is an instance of [disputant,controversialist]
[Bolshevik,Marxist,pinko,red,bolshie] is an instance of [radical]
red/n is in the synset [Bolshevik,Marxist,pinko,red,bolshie]
Using wnconnect

- John CAUS the red dress BECOME <STATE (mended)>
Shortest Path

mend–tear

all paths
Point of Discussion

• some issues
  – knowledge representation for semantic inferencing
  – do we have a network like WordNet?
  – how is it built?
  – is it wholly outside the “language faculty”?
Another Application

- Solving GRE word quiz drills
Demo

• 65 quizzes
  – 92% correct
  – remaining errors reveal gaps in WordNet’s lexical knowledge

• Code still under development...
Where next?

• **LING 438/538: Computational Linguistics**
  – is an introductory course
  – offered every Fall
  – no formal pre-requisites
  – a compromise between depth and breadth
    • *we cover basic topics but we can’t cover anything in real depth*

• **LING 581: Advanced Computational Linguistics**
  – designed as a follow-on course
  – offered every Spring
  – lectures are once a week (double length)
  – project-oriented (we do research-style work)
  – use software tools in depth (last time we did verb classes and treebanks plus statistical parsers)
  – homeworks are tasks using software tools