Lexical Semantics:
WordNet and Word Senses, Ontologies, and Semantic Lexical Resources
Lexical Semantics

- Lexicons – words (or lexemes or stems) together with some information
- Dictionaries – a lexicon with definitions for each word sense
  - Most are now available online
- Thesauruses – add synonyms for each word sense
  - Roget Thesaurus
  - WordNet
- Semantic networks – add more semantic relations
  - WordNet
  - EuroWordNet
- Ontologies – add semantic relations and rules about entities, concepts and relations
Word Senses

- We say that a word has more than one word sense if there is more than one definition.

<table>
<thead>
<tr>
<th>Online dictionary definitions for the noun plant</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. a living organism of the kind exemplified by trees, shrubs, herbs, grasses, ferns, and mosses, typically growing in a permanent site, absorbing water and inorganic substances through its roots, and synthesizing nutrients in its leaves by photosynthesis using the green pigment chlorophyll.</td>
</tr>
<tr>
<td>2. a place where an industrial or manufacturing process takes place</td>
</tr>
</tbody>
</table>

- Word senses may be
  - Coarse-grained, if not many distinctions are made
  - Fine-grained, if there are many distinctions of meanings
WordNet

- WordNet is a database of facts about words
  - Meanings and the relations among them
- Words are organized into clusters of synonyms
  - Synsets
- [http://wordnet.princeton.edu/](http://wordnet.princeton.edu/)
- Organized into nouns, verbs, adjectives, and adverbs
  - Currently 170,000 synsets
  - Available for download, arranged in separate files (DBs)
Knowledge Resources - Dictionary

• For each word in the language vocabulary, a dictionary provides:
  – A list of meanings
  – Definitions (for all word meanings)
  – Typical usage examples (for most word meanings)

WordNet definitions(called glosses)/examples for synsets of the noun plant

1. buildings for carrying on industrial labor; "they built a large plant to manufacture automobiles"
2. a living organism lacking the power of locomotion
3. something planted secretly for discovery by another; "the police used a plant to trick the thieves"; "he claimed that the evidence against him was a plant"
4. an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience
Knowledge Resources - synonyms

- A thesaurus adds:
  - An explicit synonymy relation between word meanings

WordNet synsets for the noun “plant”
1. plant, works, industrial plant
2. plant, flora, plant life
A semantic network adds relations for each word sense:
- hypernymy/hyponymy (IS-A),
  - hypernyms are more general, hyponyms are more specific
- meronymy/holonymy (PART-OF),
- antonymy, entailment, etc.

### WordNet related concepts for the meaning “plant life”
{plant, flora, plant life}
- hypernym: {organism, being}
- hyponym: {house plant}, {fungus}, …
- meronym: {plant tissue}, {plant part}
- holonym: {Plantae, kingdom Plantae, plant kingdom}
WordNet Relations

- A more detailed list from Jurafsky and Martin

<table>
<thead>
<tr>
<th>Relation</th>
<th>Also Called</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypernym</td>
<td>Superordinate</td>
<td>From concepts to superordinates</td>
<td>breakfast(^1) → meal(^1)</td>
</tr>
<tr>
<td>Hyponym</td>
<td>Subordinate</td>
<td>From concepts to subtypes</td>
<td>meal(^1) → lunch(^1)</td>
</tr>
<tr>
<td>Instance Hypernym</td>
<td>Instance</td>
<td>From instances to their concepts</td>
<td>Austen(^1) → author(^1)</td>
</tr>
<tr>
<td>Instance Hyponym</td>
<td>Has-Instance</td>
<td>From concepts to concept instances</td>
<td>composer(^1) → Bach(^1)</td>
</tr>
<tr>
<td>Member Meronym</td>
<td>Has-Member</td>
<td>From groups to their members</td>
<td>faculty(^2) → professor(^1)</td>
</tr>
<tr>
<td>Member Holonym</td>
<td>Member-Of</td>
<td>From members to their groups</td>
<td>copilot(^1) → crew(^1)</td>
</tr>
<tr>
<td>Part Meronym</td>
<td>Has-Part</td>
<td>From members to their groups</td>
<td>table(^2) → leg(^3)</td>
</tr>
<tr>
<td>Part Holonym</td>
<td>Part-Of</td>
<td>From parts to wholes</td>
<td>course(^7) → meal(^1)</td>
</tr>
<tr>
<td>Substance Meronym</td>
<td></td>
<td>From substances to their subparts</td>
<td>water(^1) → oxygen(^1)</td>
</tr>
<tr>
<td>Substance Holonym</td>
<td></td>
<td>From parts of substances to wholes</td>
<td>gin(^1) → martini(^1)</td>
</tr>
<tr>
<td>Antonym</td>
<td></td>
<td>Semantic opposition between lemmas</td>
<td>leader(^1) ↔ follower(^1)</td>
</tr>
<tr>
<td>Derivationally</td>
<td></td>
<td>Lemmas w/same morphological root</td>
<td>destruction(^1) ↔ destroy(^1)</td>
</tr>
</tbody>
</table>
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
Origins of Ontology

- In philosophy, ontology studies existence/being of the world.
  - We can think of ontology as categorizing everything in the world.

- In his work "categories", Aristotle listed ten categories that all things of the world should belong to.

http://www.jfsowa.com/talks/ontology.htm
Ontology in Information Science

• Ontology is an approach of knowledge organization.
• In general, ontologies are about the representations of semantics:
  – Concepts, e.g. *person*, *animal*, *food*, *table*, *movie*, etc.
  – Instances (or entities), e.g. Barack Obama is an *instance* of the concept “person”.
  – Properties, e.g. a person has properties of *gender*, *height*, *weight*, *father*, *mother*, etc.
  – Relations, e.g. Syracuse University is *located in* Syracuse.
  – Rules between concepts, properties, and relations, e.g. if someone is *married*, then he/she should have a *spouse*. 
Ontology Example: UMLS

- The Unified Medical Language System (UMLS) aggregates various controlled vocabularies and mapped them to a comprehensive biomedical ontology. It has three knowledge sources:
  - Metathesaurus. Mapping concepts and terms in different thesaurus and organizing them in the UMLS structure
  - Semantic network. Connecting semantic types of concepts in metathesaurus by semantic relations.
  - Specialist Lexicon. Containing lexical information of biomedical terms.

- This is an example of a word and phrase level resource
- Online, but not publically available
Semantic Lexicons

• Lexicon where each word is assigned to a semantic class
• Lexical resources have been developed to assign words to semantic classes in support of applications that need to detect opinion, sentiment, or other more subjective meanings
• Three examples given here; additional examples will be given when we cover sentiment analysis
Semantic classes: Subjectivity Lexicon

• Subjectivity Lexicon from the MPQA project with Jan Wiebe
  – Gives a list of 8,000+ words that have been judged to be weakly or strongly positive, negative or neutral in subjectivity
  – Examples:

```
type=weaksubj len=1 word1=abandoned pos1=adj stemmed1=n priorpolarity=negative

type=weaksubj len=1 word1=abandonment pos1=noun stemmed1=n priorpolarity=negative


type=weaksubj len=1 word1=abandon pos1=verb stemmed1=y priorpolarity=negative


type=strongsubj len=1 word1=abase pos1=verb stemmed1=y priorpolarity=negative


type=strongsubj len=1 word1=abasement pos1=anypos stemmed1=y priorpolarity=negative


type=strongsubj len=1 word1=abash pos1=verb stemmed1=y priorpolarity=negative


type=weaksubj len=1 word1=abate pos1=verb stemmed1=y priorpolarity=negative


type=strongsubj len=1 word1=absolve pos1=verb stemmed1=y priorpolarity=positive


type=strongsubj len=1 word1=absolute pos1=adj stemmed1=n priorpolarity=neutral
```
Semantic classes: LIWC

• Linguistic Inquiry and Word Count
  – Text analysis software based on dictionaries of word dimensions
  – Dimensions can be syntactic
    • Pronouns, past-tense verbs
  – Dimensions can be semantic
    • Social words, affect, cognitive mechanisms
  – Other categories
    • See http://www.liwc.net/comparedicts.php

• James Pennebaker, Univ. of Texas at Austin
  – http://www.liwc.net/

• Often used for positive and negative emotion words in opinion mining
Semantic classes for words: ANEW

- Affective Norms for English Words
  - Provides a set of emotional ratings for a large number of words in the English language

- Participants gave graded reactions from 1-9 on three dimensions
  - Good/bad, psychological valence
  - Active/passive, arousal valence
  - Strong/weak, dominance valence

- From the NIMH Center for the Study of Emotion and Attention at the University of Florida
  - http://csea.phhp.ufl.edu/Media.html
  - See also the paper by Dodds and Danforth on Happiness of Large-Scale Written Expressions