



Week 6:

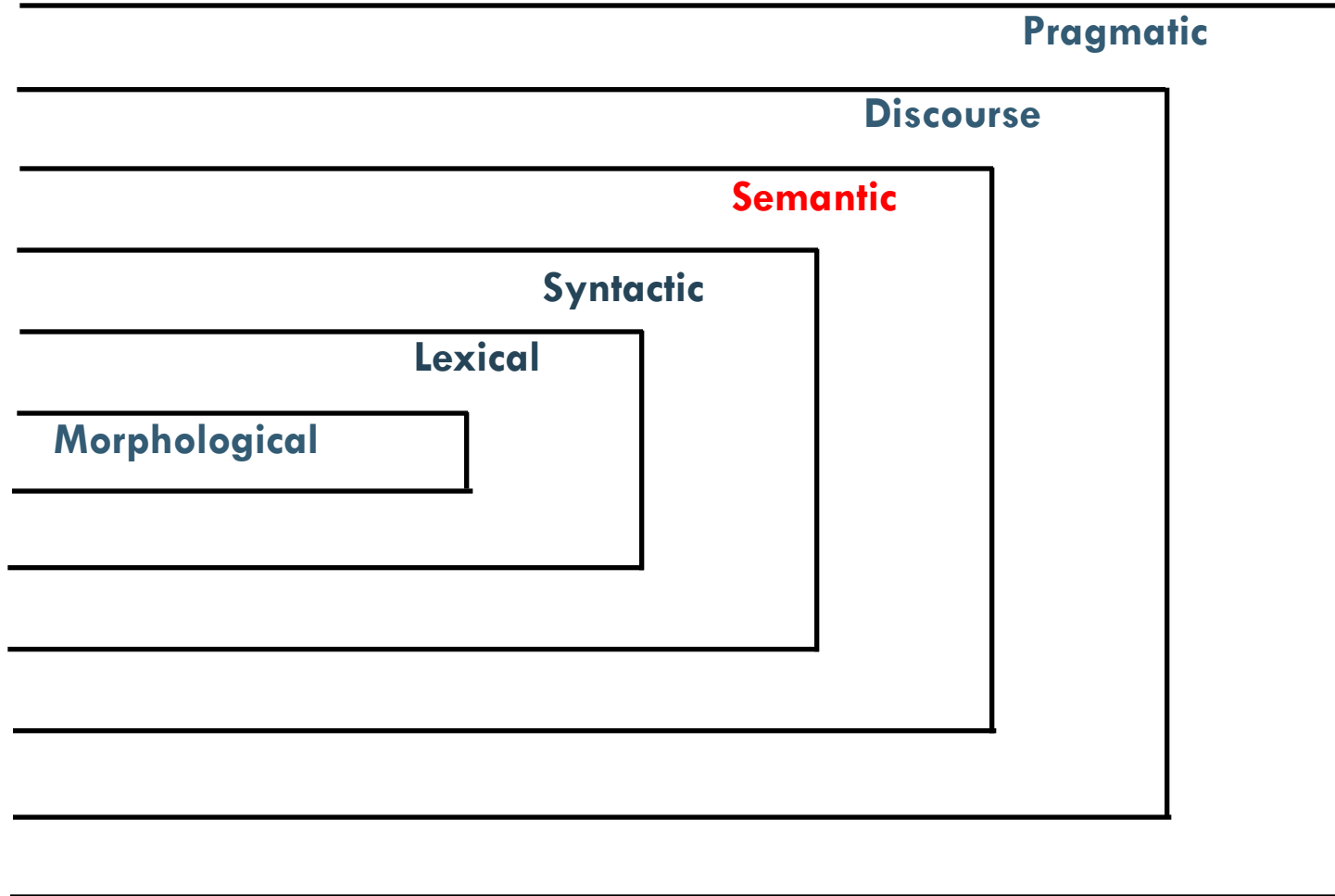
Language Analysis Foundations

School of Information Studies
Syracuse University

Overview

- Mini-talks
- Lecture
 - Lexical semantics
- Lab

Levels of Language





Lexical Semantics

School of Information Studies
Syracuse University

Word Senses

We say that a word has more than one word sense (meaning) if there is more than one definition.

Online dictionary definitions for the noun *plant*

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.
2. a place where an industrial or manufacturing process takes place
3. a person placed in a group as a spy or informer.

Word senses may be

- Coarse-grained, if not many distinctions are made
- Fine-grained, if there are many distinctions of meanings

Lexical Semantics

- **Lexicons** – list of words (or lexemes or stems) with basic info
- **Dictionaries** – a lexicon with definitions for each word sense
 - Most are now available online
- **Thesauruses** – add synonyms/ antonym for each word sense
 - WordNet
- **Semantic networks** – add more semantic relations, including semantic categories
 - WordNet, EuroWordNet
- **Ontologies** – add rules about entities, concepts and relations, semantic categories
 - UMLS
- **Semantic Lexicon** – Lexicon where each word is assigned to a semantic class
 - LIWC, ANEW, Subjectivity Lexicon

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

Thesauruses - 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

WordNet

- WordNet is a database of facts about words
 - Meanings and the relations among them
 - <http://wordnet.princeton.edu/>
- Words are organized into clusters of synonyms
 - Synsets
- Organized into nouns, verbs, adjectives, and adverbs
 - Currently 117,000 synsets
 - More developed for nouns and verbs

Sense of “bat” in Wordnet

Noun

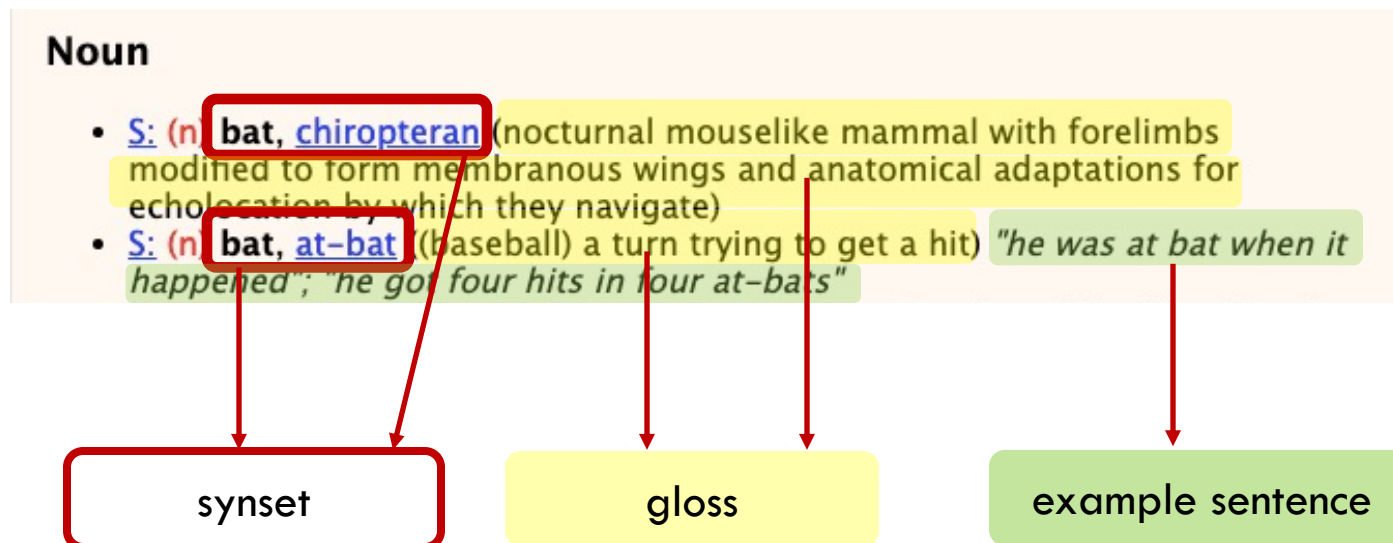
- **S: (n) bat, [chiropteran](#)** (nocturnal mouselike mammal with forelimbs modified to form membranous wings and anatomical adaptations for echolocation by which they navigate)
- **S: (n) bat, [at-bat](#)** ((baseball) a turn trying to get a hit) *"he was at bat when it happened"; "he got four hits in four at-bats"*
- **S: (n) [squash racket](#), [squash racquet](#), bat** (a small racket with a long handle used for playing squash)
- **S: (n) [cricket bat](#), bat** (the club used in playing cricket) *"a cricket bat has a narrow handle and a broad flat end for hitting"*
- **S: (n) bat** (a club used for hitting a ball in various games)

Verb

- **S: (v) bat** (strike with, or as if with a baseball bat) *"bat the ball"*
- **S: (v) bat, [flutter](#)** (wink briefly) *"bat one's eyelids"*
- **S: (v) bat** (have a turn at bat) *"Jones bats first, followed by Martinez"*
- **S: (v) bat** (use a bat) *"Who's batting?"*
- **S: (v) [cream](#), bat, [clobber](#), [drub](#), [thrash](#), [lick](#)** (beat thoroughly and conclusively in a competition or fight) *"We licked the other team on Sunday!"*

How is “sense” defined in WordNet

- The synset (synonym set), the set of near-synonyms, a brief definition (‘gloss’), and in most cases, example sentence.



Semantic Network - Relations

A semantic network adds relations for each word sense:

- hypernymy/hyponymy (“IS-A” Relationship),
 - “hyponyms” are more general in meaning. In essence a **parent**
 - (*hyper is super*)
 - “hyponyms” are more specific in meaning
- Meronymy (It is a “PART-OF” it - Relationship),
- Holonymy (It is a “PART-OF” - Relationship),

WordNet related concepts for the meaning “plant life”

{plant, flora, plant life}

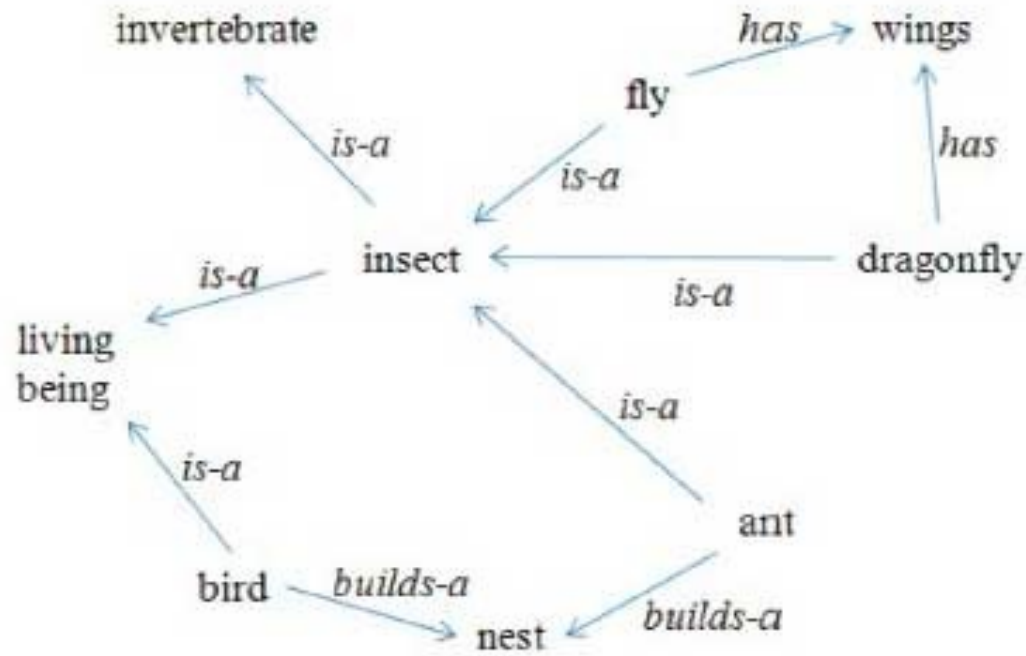
hypernym: {organism, being}

hypomym: {house plant}, {fungus}, ...

meronym: {plant tissue}, {plant part}

holonym: {Plantae, kingdom Plantae, plant kingdom}

Semantic Network



Semantic Network

is a knowledge base that represents semantic relations between concepts in a network. This is often used as a form of knowledge representation. It is a directed or undirected graph consisting of **vertices**, which represent concepts, and **edges**, which represent semantic relations between concepts, mapping or connecting semantic fields.

WordNet Hypernym Hierarchy for “bat”

- [S: \(n\) placental](#), [placental mammal](#), [eutherian](#), [eutherian mammal](#) (mammals having a placenta; all mammals except monotremes and marsupials)
 - [S: \(n\) mammal](#), [mammalian](#) (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
 - [S: \(n\) vertebrate](#), [craniate](#) (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
 - [S: \(n\) chordate](#) (any animal of the phylum Chordata having a notochord or spinal column)
 - [S: \(n\) animal](#), [animate being](#), [beast](#), [brute](#), [creature](#), [fauna](#) (a living organism characterized by voluntary movement)
 - [S: \(n\) organism](#), [being](#) (a living thing that has (or can develop) the ability to act or function independently)
 - [S: \(n\) living thing](#), [animate thing](#) (a living (or once living) entity)
 - [S: \(n\) whole](#), [unit](#) (an assemblage of parts that is regarded as a single entity)
"how big is that part compared to the whole?"; *"the team is a unit"*
 - [S: \(n\) object](#), [physical object](#) (a tangible and visible entity; an entity that can cast a shadow)
"it was full of rackets, balls and other objects"
 - [S: \(n\) physical entity](#) (an entity that has physical existence)
 - [S: \(n\) entity](#) (that which is perceived or known or inferred to have its own)

WordNet Noun Relations

A more detailed list for nouns from Jurafsky and Martin

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

WordNet Verb Relations

A more detailed list for verbs from Jurafsky and Martin

| Relation | Definition | Example |
|--------------------------------|---|---|
| Hypernym | From events to superordinate events | <i>fly</i> ⁹ → <i>travel</i> ⁵ |
| Troponym | From events to subordinate event
(often via specific manner) | <i>walk</i> ¹ → <i>stroll</i> ¹ |
| Entails * | From verbs (events) to the verbs (events) they entail | <i>snore</i> ¹ → <i>sleep</i> ¹ |
| Antonym | Semantic opposition between lemmas | <i>increase</i> ¹ ⇔ <i>decrease</i> ¹ |
| Derivationally
Related Form | Lemmas with same morphological root | <i>destroy</i> ¹ ⇔ <i>destruction</i> ¹ |

* If you are snoring it implies that you are sleeping

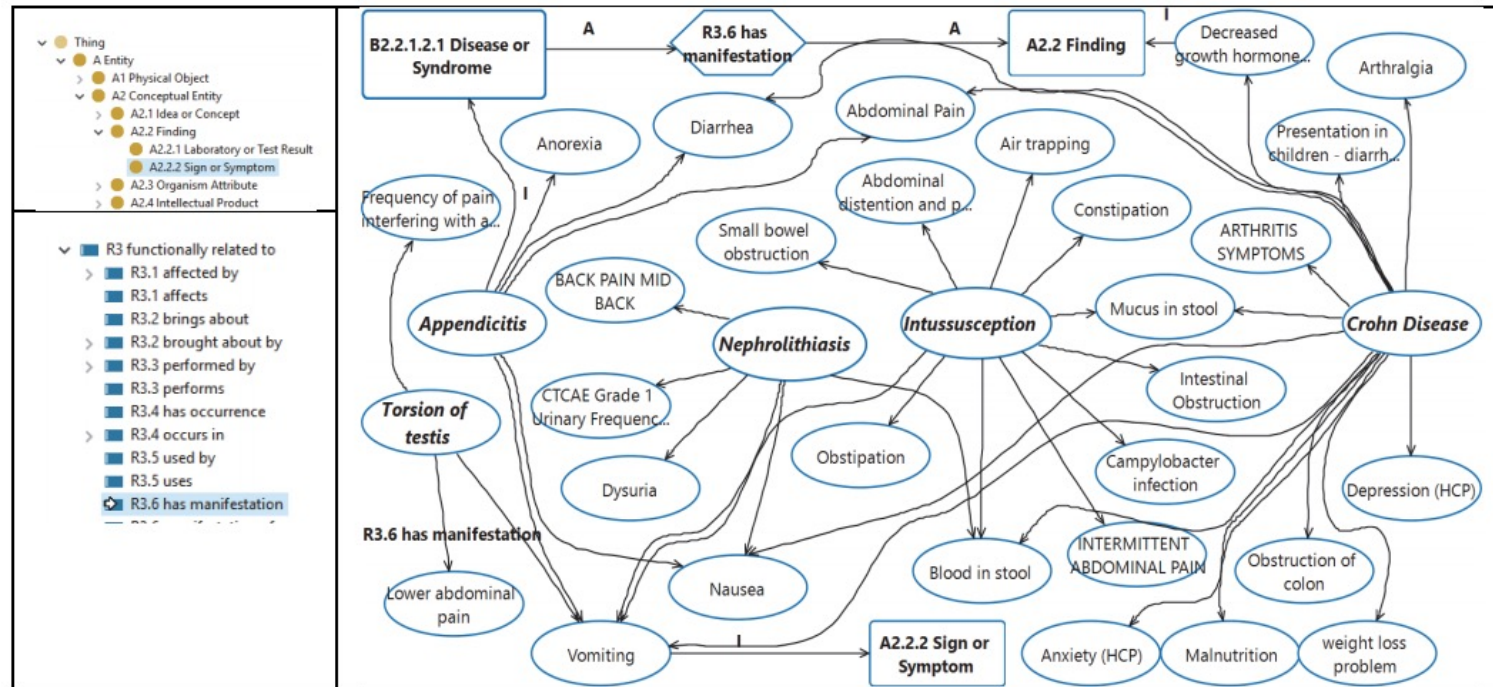
Ontology in Information Science

- 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. Joe Biden is an instance of the concept “president”.
 - **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.

Purpose of ontology

- To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To analyze domain knowledge

Example of UMLS



Semantic level of UMLS-OWL and factual level of a specific scenario

Source: <https://iswc2017.semanticweb.org/wp-content/uploads/papers/PostersDemos/paper546.pdf>

Semantic Lexicons

- Lexicon where each word is assigned to a semantic class or category
- 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 for **sentiment lexicons**

Sentiment: Subjectivity Lexicon

Subjectivity Lexicon from the MPQA project with Jan Wiebe

- Gives a list of 8,000+ words that have been judged to be positive, negative or neutral in subjectivity
- Examples:

Polarity =
Sentiment

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

Another well known Lexicon

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://liwc.wpengine.com/compare-dictionaries/>
 - James Pennebaker, Univ. of Texas at Austin
- <http://www.liwc.net/>

Semantic classes: ANEW

A third well known Lexicon

- 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>

Ranked on Scale



Lexical Semantics: Word Sense Disambiguation

School of Information Studies
Syracuse University

Defining a Word Sense

- How can we define different word senses?
 - Give a list of synonyms
 - Give a definition
- Coarse-grained senses distinguish core aspects of meaning
- Task is more difficult for Fine-grained senses
 - Example: In WordNet, which has fairly fine-grained senses, the word “bass” has five senses related to music and 3 senses related to fish (and the sense determines how you pronounce the word).

Word Sense Disambiguation (WSD)

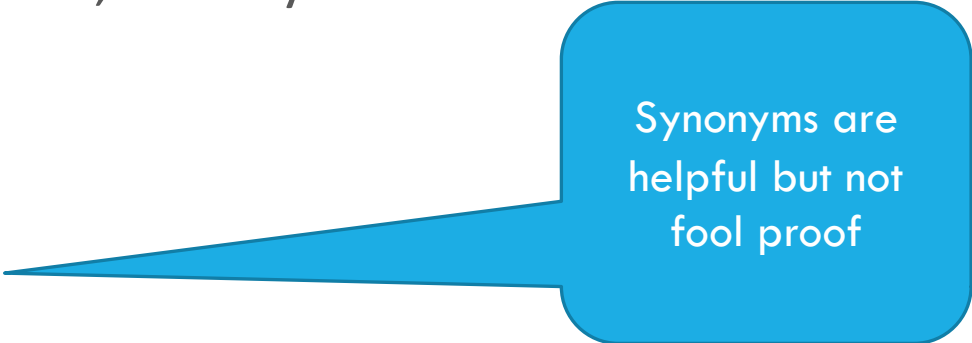
- We need to look up words in lexicons to find semantics, but how do we know which meaning (sense) of the word should we use?
- Definition of Word Sense Disambiguation
 - look up a word and try to figure out the which “sense” is the correct sense for a given context

Why WSD in NLP?

- Information retrieval
 - “bank”
- Machine translation
 - “bank”: 銀行 (financial institution) , or 河岸 (sloping land)
- Text-to-speech
 - How to pronounce “bass”

Difficulties with synonym lists

- True synonyms non-existent, or very rare
- Near-synonyms
 - Examples:
 - Error, blunder, mistake
 - Order, tell, enjoin, say



Synonyms are helpful but not fool proof

- Dimensions of synonym differentiation
 - Stylistic variation
 - Pissed, drunk, inebriated
 - Expressive variation
 - Attitude: skinny, thin, slim
 - Emotion: father, dad, daddy
 - . . .

In-class activity

Let's first check out the meanings of “**right**” from WordNet.

Then determine the meaning of “right” in the following sentences: “Make a **right** turn at the light. Don't turn left!”

Discuss how/why you choose that meaning for “right” in the context .

Human Sense Disambiguation

How humans do this ?

- Sources of influence known from psycholinguistics research:
 - local context (most important)
 - the sentence or other surrounding text containing the ambiguous word restricts the interpretation of the ambiguous word
 - (e.g., book in a sentence that has flight, travel, etc.)
 - domain knowledge
 - the fact that a text is concerned with a particular domain activates only the sense appropriate to that domain
 - (e.g., plant in a biology article)
 - frequency data
 - the frequency of each sense in general usage affects its accessibility to the mind

Lesk Algorithm

Many algorithms look at words in context

- Original Lesk definition: measure overlap between sense definitions for all words in context. (Michael Lesk 1986)
 - Identify simultaneously the correct senses for all words in context
- Simplified Lesk (Kilgarrriff & Rosensweig 2000): measure overlap between sense definitions of a word and current context
 - Identify the correct sense for one word at a time
 - Current context is the set of words in the surrounding sentence/paragraph/document.

Lesk Algorithm: A Simplified Version

- **Algorithm** for simplified Lesk:
 1. Retrieve from lexicon all sense definitions of the target word
 2. Determine the overlap between each sense definition and the current context
 3. Choose the sense that leads to highest overlap

Example: disambiguate PINE in

“Pine cones hanging in a tree”

- PINE

1. kinds of evergreen tree with needle-shaped leaves
2. waste away through sorrow or illness

Pine#1 \cap Sentence = 1

Pine#2 \cap Sentence = 0

Sense #1 has 1 word in
sentence “Tree”

Classification approach to WSD

- Train a classification algorithm that can label each word with the correct sense, given the context of the word
 - Training set is the hand-labeled corpus of senses
- The context is represented as a set of “features” of the word and includes information about the surrounding words
 - Typical features shown on next slides
- Result of training is a model that is used by the classification algorithm to label words in the test set, and ultimately, in new text examples

Machine learning 101:

<https://www.youtube.com/watch?v=nKW8Ndu7Mjw>

Steps in machine learning

1. Data collection
2. Data preparation
3. Choose model
4. Training
5. Evaluation
6. Parameter tuning
7. Prediction

WSD classification features

- Collocational features from the target word
 - Information about words in specific positions (i.e. previous word)
 - Typical features include the word itself, its stem and its POS tag
- Similar word features
 - Whether words surrounding the target word are “similar” to those in the word definition
- Associated word features
 - for each word to be disambiguated, from a large corpus collect a small number of frequently-used context words.
 - Example: for each word, collect the 12 most frequent context words from a collection of sentences drawn from the corpus.

For **bass**, the 12 most frequent context words from the WSJ are:

- [fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band]

Word Similarity Features

- Important in these classifications is that instead of just using words that overlap in context, look for “similar” words
- Similarity measures
 - Can be defined from a semantic relation lexicon, such as WordNet
 - One example is path similarity
 - For any two words, gives a number between 0 and 1 based on the shortest path between the two words in the WordNet hypernym/hyponym hierarchy
 - For example, the words “plant” and “tree” should have a shorter path through a common ancestor than words like “plant” and “piano”.



Lab

School of Information Studies
Syracuse University

Tasks

1. Understand the structure of WordNet
2. Calculate the similarity of two words



3. Figure out the sentiment implied in a word