A person’s knowledge store of words and their properties is known as a **lexicon**
- In this lecture we will use word in place of *lexeme* or *lexical unit* with the understanding that some words are made up of multiple words
- Everyone has a unique lexicon, although you would expect significant overlap within the same community

**Syntactic properties and collocation preferences are stored here**
- Is the word a verb, noun, etc.
- Are there any words that “sound better” to use with this word?

**Semantics is also stored here**
- What the word means
- How this word combines with other words to form larger units of meaning
The organization of the lexicon

- The existence of the lexicon is a theory
  - We can’t directly observe a lexicon
- People can pick the right word very quickly
  - How can they sort through such a large amount of words so quickly?
- Numerous models of how words are organized and retrieved in the lexicon have been posited
- Today we are going to focus on the relational model of the lexicon
  - Words are connected to other words through a series of relations
  - There are actually many different relational models of the lexicon, but they all share in common the fact that they are organized around relationships
    - Differ in what the relationships are
    - Differ if the relationships are between words or concepts
Semantic Relationships

- We will go over specific examples of relationships soon
  - An example of a semantic relationship is **synonymy**
- How do we know if something is a useful semantic relationship?
- Cruse offers these guidelines
  - Recurrence
    - The relationship should be something that holds between many pairs of words
  - Discrimination
    - The relationship should be something that discriminates between pairs of words
      - We don’t want a relationship that exist between any random pair of words
  - Accessibility
    - The relationship should be something people have an intuition about and can name or talk about

Cruse, 2011. Meaning in Language pp 129-130
Common Semantic Relationships

- **Hyponymy / Hypernymy**
  - *Apple - Fruit*
  - The “is-a” relationship
  - The hyponym is the more specific of the pair (*Apple* is a hyponym of *fruit*)
  - The hypernym is the more generic of the pair (*Fruit* is a hypernym of *apple*)

- **Meronymy**
  - *Finger - hand*
  - The part-whole relationship
  - The meronym is the part (*Finger* is a meronym of *hand*)
  - The holonym is the whole (*Hand* is a holonym of *finger*)

- **Synonymy**
  - Words with very similar meanings
  - Exact meanings are very rare to find
    - Language is usually efficient, no reason to carry two words in the lexicon when one will do
Common Semantic Continued

- **Opposites**
  - Much less agreement on what it is
    - Numerous classifications exist, including antonyms, complementaires, reversives, etc.

- **Antonymy**
  - Some relational models use antonym to mean opposite
  - One definition requires both words to be *gradable*
    - They can be combined with degree modifiers like *very, slightly*, etc.
  - The indicate degrees on a unidimensional property
  - Is symmetric

- **Troponymy**
  - Applies only to verbs
  - Specifies a more specific manner of doing the action described in the verb
    - *Stroll* is a troponym of *walk*
  - Might be represented as hypernym-hyponym pair in some models
How do we know this

- It is hard to say for certain these are the models that actually exist in humans
- BUT, we do have some decent evidence from psycholinguistics using carefully controlled studies, that suggest this is the case
- Semantic Priming establishes that there is a general relationship between words
  - Participants are shown sequences in rapid succession, and asked if the letters make up a word
  - By varying the words that come before, the time it takes to recognize a sequence of letters can be manipulated
    - If the participant sees cat a few examples prior, then their recognition of dog will be much faster
Knowledge of semantic relationships is very useful in NLP and a number of related fields.

In Information Retrieval (IR), knowledge of synonyms can be used to find more results that just using a query string alone.

For Natural Language Understanding, semantic relations can be used to perform inference.

- I saw an elephant last week on safari. The animal was so majestic.

When automatically evaluating machine translated text, semantically related words can be used to more fairly grade a translation than just exact matching.

- METEOR does this.
Word Senses

- Clearly semantic relationships directly between words isn’t optimal
- Instead we’ll create a **sense** for each different meaning a word can have
- A word with multiple senses is said to be polysemous

It is a 30 minute car ride to the bank.

- Most of the time we would think this refers to the financial institution
  - But depending on context we might select the edge of the water interpretation
- **Word Senses are really hard to differentiate**
  - Lumpers vs Splitters
How many senses of the word **bass** do you think there is here?

1. The largest **bass** ever caught at this lake was 20 pounds
2. I was fishing for **bass** in the Atlantic but we had to head back because of the hurricane warning.
3. The band has a really good **bass** player.
4. The tuba is a **bass** instrument.
5. My friend sings **bass** in the chorus
The NLP task of assigning a word its correct sense in context is Word Sense Disambiguation (WSD).

It is done in two settings:
- All-words - Disambiguate every word in the sentence
- Lexical Sample - Disambiguate specified words in a sentence

Why it is useful:
- In Machine Translation it should theoretically help to translate into the correct word
  - Isn’t often used in practice
- Speech synthesis
  - Different senses have different pronunciations
- Natural Language Understanding
  - To perform computations over language (QA, inference) you usually need to know the correct sense
  - Representing the meaning of a sentence abstractly (we’ll talk about this next week)
How to Do WSD

- Just take the first sense

- Machine Learning
  - Train a tagger or classifier over annotated data
  - Use parallel corpora to do this in a semi-supervised way
    - Each different translation in the target language is a different sense

- Dictionary Based Methods
  - Use the definition (gloss) of each sense and calculate overlap with the context you found the word in (Lesk Algorithm)
  - Use the definition of each sense and its ancestors in WordNet and calculate overlap as before (Extended Lesk)
Lesk Algorithm Example

- Use the Lesk Algorithm to assign senses to the word *cone* in the following sentences

1. That pine *cone* has a lovely color.
2. I can’t be in a bright room because I have a problem with my *cones*
3. The pine *cone* is missing half of it’s scales

The definitions are:

**cone#1** cone-shaped mass of ovule- or spore-bearing scales or bracts

**cone#2** a visual receptor cell in the retina that is sensitive to bright light and to color
WordNet

- WordNet is a computational semantic lexicon
- Each sense is a group of synonyms called a *synset*
- For the senses of bass we were discussing before, the synsets are
  - Bass, basso
  - Bass, bass voice, basso
  - Sea bass, bass
  - Freshwater bass, bass
- Each synset has a gloss (definition) and usually an example sentence
- Each synset is related to other synsets through semantic relations
- Antonymy is represented between specific words, rather than synsets
- Almost every programming language has an interface to WordNet
- WordNet is now a general term for this type of structure that has been produced for many natural languages
Let's look at the WordNet entries for:
  - Bass
  - Walk
  - Hot
Word Similarity

- Going back to semantic priming, it is clear to see that there is a notion of similarity between words.
  - Comparing against human similarity scores is one way to evaluate a lexicon
- Many famous datasets have been created that list a pair of words and their similarity score (e.g. quick-rapid 9.7)
  - These are created by showing many annotators pairs of words and asking them to judge how similar they are
    - This is a very hard task
  - Popular datasets include
    - WordSim353
    - SimLex999
- Recent work by Ettinger and Linzen suggests using the priming timings themselves as a gold standard
Word Similarity in WordNet

- Numerous methods to judge similarity of words from WordNet exist
- Most use the path distance between two words
- Resnik gives similarity between two words as the maximum similarity between any two of their senses
- Numerous other methods are based on the idea of the Lowest Common Subsumer (LCS), which is the lowest synset in the tree that both senses we are interested in have in common
- It is also useful to define the probability of a sense in WordNet
  - Rather than just the count the number of times a word appears, we count the number of times each word under a sense appears in a corpus and divide that by the total number of words in the corpus
● Let our corpus be

The cat sat on the mat under the dog who was staring at the fox.

Find P(animal)
Word Similarity in WordNet

- **Types of Similarities**
- **Resnik**
  - SIM = -log P(LCS (c1, c2))
- **Lin**
  - SIM = \(-2 \times \log P(LCS(c1, c2)) / \log P(c1) + \log P(c2)\)
- **Extended Lesk**
  - Just as before, but now rather than between a context and a sense, it is between two senses
  - Square each overlaps length so that longer overlaps are weighted higher
- **Practice**
  - Hill and coast
Finding New Semantic Relations using Patterns

- WordNet is a very commonly used resource, but naturally it cannot cover everything.
- Finding new word pairs that participate in a relationship is another active area of research in the community.
- Originally this was done using lexico-syntactic patterns, sometimes called Hearst Patterns after Marti Hearst.
  - Search a corpus for “X is a [(kind|type) of] Y” and if X and Y occur in this pattern over some threshold then X is a hyponym of Y.
  - Further research as been done on discovering new patterns.
- **Current state of the art** research uses patterns in combination with neural network based classifiers to perform this task.
Semantic Lexicons Outside of WordNet

● Besides WordNet, other popular lexical resources are FrameNet and VerbNet

● **FrameNet** is based on the semantic theory known as frame semantics, popularized by Charles J. Fillmore
  ○ Semantic frames specify the meaning of words by listing out a number of attributes that are either filled or must be filled by other words in a sentence.

● **VerbNet** primary contribution is the listing of the arguments of verbs and the roles they take (We will cover semantic roles later)