Word Senses

Slides adapted from

Dan Jurafsky and James Martin

Recap on words: lemma vs. word form

- A **lemma** or **citation form**
  - Same stem, part of speech, rough semantics
- A **word form**
  - The inflected word as it appears in text

Word form	Lemma
banks	bank
sung	sing
duermes	dormir

Lemmas have senses

- One lemma "bank" can have many meanings:

Sense 1:
  - …a **bank**$_1$ can hold the investments in a custodial account…

Sense 2:
  - "…as agriculture burgeons on the east **bank**$_2$ the river will shrink even more"
- **Sense** (or **word sense**)
  - A discrete representation

    of an aspect of a word's meaning.
- The lemma **bank** here has two senses

Homonymy

Homonyms: words that share a form but have unrelated, distinct meanings:

- $bank_1$: financial institution, $bank_2$: sloping land
- bat_1: club for hitting a ball, bat_2: nocturnal flying mammal

1. Homographs (bank/bank, bat/bat)
2. Homophones:
    1. write and right
    2. piece and peace

Homonymy causes problems for NLP

- Information retrieval
  - "`bat care`"
- Machine Translation
  - `bat`: murciélago (animal) or bate (for baseball)
- Text-to-Speech
  - `bass` (stringed instrument) vs. `bass` (fish)

Polysemy

- 1. The **bank** was constructed in 1875 out of local red brick.
- 2. I withdrew the money from the **bank**
- Are those the same sense?
  - Sense 2: “A financial institution”
  - Sense 1: “The building belonging to a financial institution”
- A **polysemous** word has **related** meanings
  - Most non-rare words have multiple meanings

Metonymy or systematic polysemy

- Lots of types of polysemy are systematic
  - `School, university, hospital`
  - All can mean the institution or the building.
- A systematic relationship:
  - Building ⟺ Organization
- Other such kinds of systematic polysemy:

Author `(Jane Austen wrote Emma)`

⟺ Works of Author `(I love Jane Austen)`

Tree `(Plums have beautiful blossoms)`

⟺ Fruit `(I ate a preserved plum)`

How do we know if more than one sense?

- The "zeugma" test: Two senses of `serve`?
  - `Which flights` **`serve`** `breakfast?`
  - `Does Lufthansa` **`serve`** `Philadelphia?`
  - ?Does Lufthansa serve breakfast and Philadelphia?
- Since this conjunction sounds weird,
  - we say that these are **two different senses of "serve"**

Quiz

- Which of the following pairs exemplify homonymy (as opposed to polysemy)?

1. mouse (animal) vs. mouse (electronic device)
2. bark (of a dog) vs. bark (of a tree)
3. rock (music) vs. rock (hard)
4. chair (for sitting) vs. chair (of a meeting)

Sense Relations

Slides adapted from

Dan Jurafsky and James Martin

Synonyms

- Word that have the same meaning in some or all contexts.
  - filbert / hazelnut
  - couch / sofa
  - big / large
  - automobile / car
  - vomit / throw up
  - Water / H_20
- Two lexemes are synonyms
  - if they can be substituted for each other in all situations
  - If so they have the same **propositional meaning**

Synonyms

- But there are few (or no) examples of perfect synonymy.
  - Even if many aspects of meaning are identical
  - Still may not preserve the acceptability based on notions of politeness, slang, register, genre, etc.
- Example:
  - Water/H_20
  - Big/large
  - Brave/courageous

Synonymy is a relation between senses rather than words

- Consider the words *big* and *large*
- Are they synonyms?
  - How **big** is that plane?
  - Would I be flying on a **large** or small plane?
- How about here:
  - Miss Nelson became a kind of **big** sister to Benjamin.
  - ?Miss Nelson became a kind of **large** sister to Benjamin.
- Why?
  - *big* has a sense that means being older, or grown up
  - *large* lacks this sense

Antonyms

- Senses that are opposites with respect to one feature of meaning
- Otherwise, they are very similar!

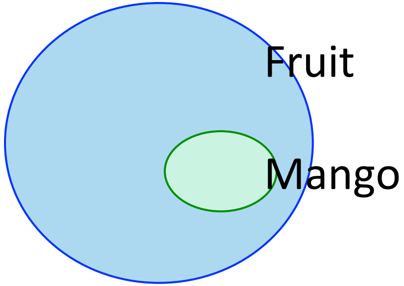
`dark/light`	`short/long`	`fast/slow`	`rise/fall`
`hot/cold`	`up/down`	`in/out`	

- Can define a binary opposition or be at opposite ends of a scale
  - `alive/dead`
  - `fast/slow`
- Scale can be context-sensitive:
  - a short basketball player can be a tall person

Hyponymy and Hypernymy

- One sense is a **hyponym** of another if the first sense is more specific, denoting a subclass of the other
  - *car* is a hyponym of *vehicle*
  - *mango* is a hyponym of *fruit*
- Conversely **hypernym/superordinate** ("hyper is super")
  - *vehicle* is a hypernym of *car*
  - *fruit* is a hypernym of *mango*

Hyponymy more formally



- Extensional:
  - The class denoted by the hypernym extensionally includes the class denoted by the hyponym
- Entailment:
  - A sense A is a hyponym of sense B if *being an A* entails *being a B*
- Another name: the **IS-A hierarchy**
  - A IS-A B (or A ISA B)
  - B **subsumes** A

Hyponyms and Instances

- Hyponymy holds between **classes**
- Classes have specific **instances**.
- An **instance** is an individual, a proper noun that is a unique entity
    - San Francisco is an **instance** of city
  - But city is a class
    - city is a **hyponym** of municipality, ..., location...

Meronymy

- The part-whole relation
  - A *leg* is part of a *chair*; a *wheel* is part of a *car*.
- *Wheel* is a **meronym** of *car*, and *car* is a **holonym** of *wheel*.

Quiz

- Which of the following pairs exemplify hyponymy/hypernymy?

1. dog – animal
2. dog – tail
3. dog – beagle
4. dog – Snoopy

WordNet

Slides adapted from

Dan Jurafsky and James Martin

WordNet 3.0

- A hierarchically organized lexical database
- On-line thesaurus + aspects of a dictionary
    - Some other languages available or under development
        - (Arabic, Finnish, German, Portuguese…)

Category	Unique Strings
Noun	117,798
Verb	11,529
Adjective	22,479
Adverb	4,481

Senses of “bass” in Wordnet

Noun

- S: (n) **bass** (the lowest part of the musical range)
- S: (n) **bass**, bass part (the lowest part in polyphonic music)
- **S: (n) bass, basso (an adult male singer with the lowest voice)**
- S: (n) sea bass, **bass** (the lean flesh of a saltwater fish of the family Serranidae)
- S: (n) freshwater bass, **bass** (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
- S: (n) **bass**, bass voice, basso (the lowest adult male singing voice)
- S: (n) **bass** (the member with the lowest range of a family of musical instruments)
- S: (n) **bass** (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Adjective

- S: (adj) **bass**, deep (having or denoting a low vocal or instrumental range) *"a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"*

How is "sense" defined in WordNet?

- **The synset (synonym set),** the set of near-synonyms, instantiates a sense or concept, with a gloss
- Example: chump as a noun with the gloss:

  "a person who is gullible and easy to take advantage of"
- This sense of "chump" is shared by 9 words:

  chump^1, fool^2, gull^1, mark^9, patsy^1, fall guy^1, sucker^1, soft touch^1, mug^2
- Each of **these** senses have this same gloss
  - (Not **every** sense; sense 2 of gull is the aquatic bird)

WordNet Hypernym Hierarchy for “bass”

- S: (n) **bass**, basso (an adult male singer with the lowest voice)
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) singer, vocalist, vocalizer, vocaliser (a person who sings)
      - S: (n) musician, instrumentalist, player (someone who plays a musical instrument (as a profession))
        - S: (n) performer, performing artist (an entertainer who performs a dramatic or musical work for an audience)
          - S: (n) entertainer (a person who tries to please or amuse)
            - S: (n) person, individual, someone, somebody, mortal, soul (a human being) *"there was too much for one person to do"*
              - 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 distinct existence (living or nonliving))

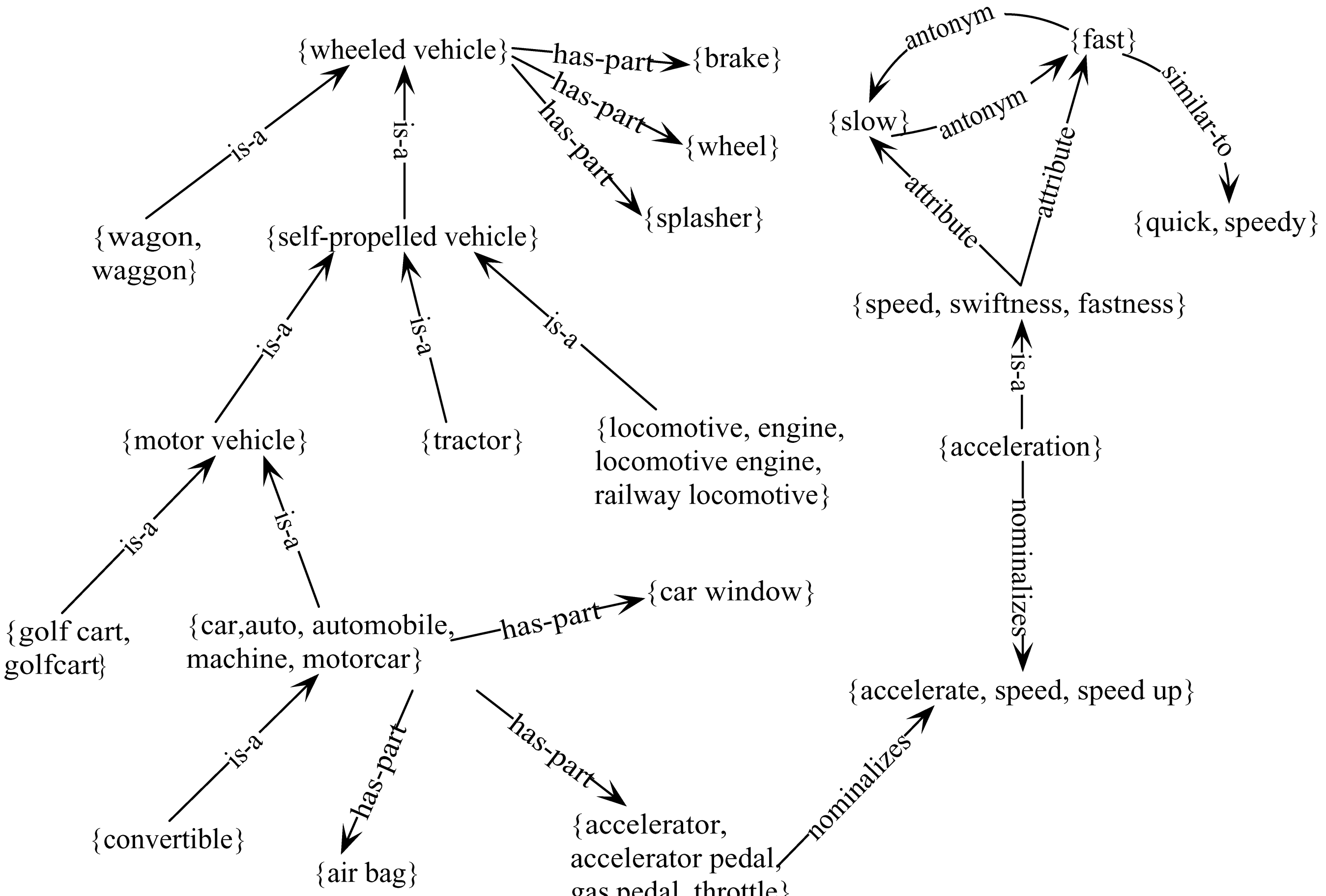
WordNet Noun Relations

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

WordNet VerbRelations

Relation	Definition	Example
Hypernym	From events to superordinate events	$\textit{fly}^{9} \rightarrow \textit{travel}^{5}$
Troponym	From events to subordinate event (often via specific manner)	$\textit{walk}^{1} \rightarrow \textit{stroll}^{1}$
Entails	From verbs (events) to the verbs (events) they entail	$\textit{snore}^{1} \rightarrow \textit{sleep}^{1}$
Antonym	Semantic opposition between lemmas	$\textit{increase}^{1} \iff \textit{decrease}^{1}$
Derivationally Related Form	Lemmas with same morphological root	$\textit{destroy}^{1} \iff \textit{destruction}^{1}$

WordNet: Viewed as a graph



"Supersenses"
The top level hypernyms in the hierarchy

(counts from Schneider and Smith 2013's Streusel corpus)

Noun

GROUP	1469	*place*
PERSON	1202	*people*
ARTIFACT	971	*car*
COGNITION	771	*way*
FOOD	766	*food*
ACT	700	*service*
LOCATION	638	*area*
TIME	530	*day*
EVENT	431	*experience*
COMMUNIC.*	417	*review*
POSSESSION	339	*price*
ATTRIBUTE	205	*quality*
QUANTITY	102	*amount*
ANIMAL	88	*dog*
BODY	87	*hair*
STATE	56	*pain*
NATURAL OBJ.	54	*flower*
RELATION	35	*portion*
SUBSTANCE	34	*oil*
FEELING	34	*discomfort*
PROCESS	28	*process*
MOTIVE	25	*reason*
PHENOMENON	23	*result*
SHAPE	6	*square*
PLANT	5	*tree*
OTHER	2	*stuff*

Verb

STATIVE	2922	*is*
COGNITION	1093	*know*
COMMUNIC.*	974	*recommend*
SOCIAL	944	*use*
MOTION	602	*go*
POSSESSION	309	*pay*
CHANGE	274	*fix*
EMOTION	249	*love*
PERCEPTION	143	*see*
CONSUMPTION	93	*have*
BODY	82	*get...done*
CREATION	64	*cook*
CONTACT	46	*put*
COMPETITION	11	*win*
WEATHER	0	—

Supersenses

- A word's supersense can be a useful coarse-grained representation of word meaning for NLP tasks

I googled$_{\texttt{communication}}$ restaurants$_{\texttt{GROUP}}$ in the area$_{\texttt{LOCATION}}$ and Fuji_Sushi$_{\texttt{GROUP}}$ came_up$_{\texttt{communication}}$ and reviews$_{\texttt{COMMUNICATION}}$ were$_{\texttt{stative}}$ great so I made_ a carry_out$_{\texttt{possession}}$ _order$_{\texttt{communication}}$

Word Sense Disambiguation

Slides adapted from

Dan Jurafsky and James Martin

Word Sense Disambiguation (WSD)

- Task
  - A word in context + a fixed inventory of potential word senses
  - Decide which sense of the word this is
- Why?
  - Machine translation, QA, speech synthesis, …
- What set of senses?
  - English-to-Spanish MT: set of Spanish translations
  - Speech Synthesis: homographs like *bass* and *bow*
  - In general: the senses in a thesaurus like WordNet

Two variants of WSD task

- Lexical Sample task
  - Small pre-selected set of target words (*line, plant*)
  - And inventory of senses for each word
  - **Supervised machine learning: train a classifier for each word**
- All-words task
  - Every word in an entire text
  - A lexicon with senses for each word
  - Data sparseness: can't train word-specific classifiers

Supervised Machine Learning Approaches

- Supervised machine learning approach:
  - a training corpus of words tagged in context with their sense
  - used to train a classifier that can tag words in new text
- Summary of what we need:
  - the **tag set** (“sense inventory”)
  - the **training corpus**
  - A set of **features** extracted from the training corpus
  - A **classifier**

Supervised WSD 1: WSD Tags

- What's a tag?

  A dictionary sense?
- For example, for WordNet an instance of "bass" in a text has 8 possible tags or labels (bass1 through bass8).

Inventory of sense tags for *bass*

WordNet Sense	Spanish Translation	Roget Category	Target Word in Context
$bass^4$	lubina	FISH/INSECT	…fish as Pacific salmon and striped **bass** and…
$bass^4$	lubina	FISH/INSECT	…produce filets of smoked **bass** or sturgeon…
$bass^7$	bajo	MUSIC	…exciting jazz **bass** player since Ray Brown…
$bass^7$	bajo	MUSIC	…play **bass** because he doesn’t have to solo…

Supervised WSD 2: Get a corpus

- Lexical sample task:
  - *Line-hard-serve* corpus - 4000 examples of each
  - *Interest* corpus - 2369 sense-tagged examples
- All words:
  - **Semantic concordance**: a corpus in which each open-class word is labeled with a sense from a specific dictionary/thesaurus.
    - SemCor: 234,000 words from Brown Corpus, manually tagged with WordNet senses
    - SENSEVAL-3 competition corpora - 2081 tagged word tokens

SemCor

<wf pos=PRP>**He**</wf>

<wf pos=VB lemma=recognize wnsn=4 lexsn=2:31:00::>**recognized**</wf>

<wf pos=DT>**the**</wf>

<wf pos=NN lemma=gesture wnsn=1 lexsn=1:04:00::>**gesture**</wf>

<punc>.</punc>

Supervised WSD 3: Extract feature vectors

- Intuition from Warren Weaver (1955):

  "If one examines the words in a book, one at a time as through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of the words…

  But if one lengthens the slit in the opaque mask, until one can see not only the central word in question but also say N words on either side, then if N is large enough one can unambiguously decide the meaning of the central word…

  The practical question is : ``What minimum value of N will, at least in a tolerable fraction of cases, lead to the correct choice of meaning for the central word?"

Feature vectors

- A simple representation of each target word instance
  - **Vectors** of sets of feature/value pairs
  - Represented as an ordered list of values
  - Representing, e.g., the window of words around the target

Two kinds of features in the vectors

- **Collocational** features and **bag-of-words** features
  - **Collocational**
    - Features about words at **specific** positions near target word
      - Often limited to just word identity and POS
  - **Bag-of-words**
    - Features about words that occur anywhere in the window
      - Typically limited to frequency counts

Feature Example

- Example text (WSJ):

  An electric guitar and **bass** player stand off to one side not really part of the scene

- Assume a window of +/- 2 from the target

Feature Example

- Example text (WSJ)

  An electric $\boxed{\text{guitar}}\boxed{\text{and}}$ **bass** $\boxed{\text{player}}\boxed{\text{stand}}$ off to one side not really part of the scene,

- Assume a window of +/- 2 from the target

Collocational features

- Position-specific information about the words and collocations in window
- $\boxed{\text{guitar}}\boxed{\text{and}}$ bass $\boxed{\text{player}}\boxed{\text{stand}}$

$$[w_{i-2}, \mathrm{POS}_{i-2}, w_{i-1}, \mathrm{POS}_{i-1}, w_{i+1}, \mathrm{POS}_{i+1}, w_{i+2}, \mathrm{POS}_{i+2}, w_{i-2}^{i-1}, w_{i}^{i+1}]$$

```
[guitar, NN, and, CC, player, NN, stand, VB, and guitar, player stand]
```

- word 1,2,3 grams in window of ±3 is common

Bag-of-words features

- An unordered set of words – position ignored
- Counts of words that occur within the window
  - Choose a vocabulary
  - Count how often each word occurs in a given window
  - Sometimes just a binary “indicator”: 1 or 0

Co-Occurrence Example

- Assume we've settled on a possible vocabulary of 12 words in "bass" sentences:

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

- The vector for:

  guitar and bass player stand

  [0,0,0,1,0,0,0,0,0,0,1,0]

Supervised WSD 4: Classifier

- Input:
  - a word w in a text window d (which we'll call a "document")
  - a fixed set of classes (senses) $C = \{c_1, c_2, \ldots, c_J\}$
  - A training set of m hand-labeled text windows again called "documents" $D = \{(d_1, c_1), \ldots, (d_m, c_m)\}$
- Output:
  - a learned classifier $f(d) = c$

Naïve Bayes classifier

- Probability of class/sense given document/context:

$$P(c \mid d) = P(c)\, P(d \mid c) \,/\, P(d)$$

- Assume independence between context words:

$$P(d \mid c) = \prod_i P(w_i \mid c)$$

- Find most probable class/sense:

$$f(d) = argmax_j\, P(c_j) \prod_i P(w_i \mid c_j)$$

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w \mid c) = \frac{count(w,c)+1}{count(c)+|V|}$$

	Doc	Words	Class
Training	1	fish smoked fish	f
	2	fish line	f
	3	fish haul smoked	f
	4	guitar jazz line	g
Test	5	line guitar jazz jazz	?

V = {fish, smoked, line, haul, guitar, jazz}

Priors:

$P(f) = \frac{3}{4}$

$P(g) = \frac{1}{4}$

Conditional Probabilities:

P(line|*f*) = (1+1) / (8+6) = 2/14

P(guitar|*f*) = (0+1) / (8+6) = 1/14

P(jazz|*f*) = (0+1) / (8+6) = 1/14

P(line|*g*) = (1+1) / (3+6) = 2/9

P(guitar|*g*) = (1+1) / (3+6) = 2/9

P(jazz|*g*) = (1+1) / (3+6) = 2/9

Choosing a class:

P(f|d5) $\propto$ 3/4 * 2/14 * $(1/14)^2$ * 1/14 $\approx$ 0.00003

P(g|d5) $\propto$ 1/4 * 2/9 * $(2/9)^2$ * 2/9 $\approx$ 0.0006

WSD Evaluations and baselines

- Best evaluation: **extrinsic (end-to-end, task-based) evaluation**
  - Embed WSD algorithm in a task and see if you can do the task better!
- What we often do for convenience: **intrinsic evaluation**
  - Exact match **sense accuracy**
    - % of words tagged identically with the human-manual sense tags
  - Usually evaluate using **held-out data** from same labeled corpus
- Baselines
  - Random guessing
  - Most frequent sense

Most Frequent Sense

- WordNet senses are ordered in frequency order
- So "most frequent sense" in WordNet = "take the first sense"
- Sense frequencies come from the *SemCor* corpus

Freq	Synset	Gloss
338	plant^1, works, industrial plant	buildings for carrying on industrial labor
207	plant^2, flora, plant life	a living organism lacking the power of locomotion
2	plant^3	something planted secretly for discovery by another
0	plant^4	an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience

Ceiling

- Human inter-annotator agreement
  - Compare annotations of two humans
  - On same data
  - Given same tagging guidelines
- Human agreements on all-words corpora with WordNet style senses
  - 75%–80%

WordNet 3.0

- Where it is:
  - http://wordnetweb.princeton.edu/perl/webwn
- Libraries
  - Python: WordNet from NLTK
    - http://www.nltk.org/Home
  - Java:
    - JWNL, extJWNL on sourceforge