Word Relations and Word Sense Disambiguation

Slides adapted from Dan Jurafsky, Jim Martin and Chris Manning

Homework Questions? Schedule

- Next week
  - Finish semantics
  - Begin machine learning for NLP
  - Review for midterm
- Midterm
  - October 27^{th}
  - Will cover everything through semantics
  - A sample midterm will be posted
  - Includes multiple choice, short answer, problem solving
- October 29^{th}
  - Bob Coyne and Words Eye: Not to be missed!
- Class outing to *Where the Wild Things Are*

  Either Friday Oct. 23^{rd} or Sunday Oct. 25^{th}. Sign sheet or send email if interested.

Three Perspectives on Meaning

1. **Lexical Semantics**
   - The meanings of individual words
2. **Formal Semantics** (or Compositional Semantics or Sentential Semantics)
   - How those meanings combine to make meanings for individual sentences or utterances
3. **Discourse or Pragmatics**
   - How those meanings combine with each other and with other facts about various kinds of context to make meanings for a text or discourse
   - **Dialog or Conversation** is often lumped together with Discourse

Outline: Comp Lexical Semantics

- Intro to Lexical Semantics
  - Homonymy, Polysemy, Synonymy
  - Online resources: WordNet
- Computational Lexical Semantics
  - Word Sense Disambiguation
    - Supervised
    - Semi-supervised
  - Word Similarity
    - Thesaurus-based
    - Distributional

Preliminaries

- What's a word?
  - Definitions we've used over the class: Types, tokens, stems, roots, inflected forms, etc...
  - Lexeme: An entry in a lexicon consisting of a pairing of a form with a single meaning representation
  - Lexicon: A collection of lexemes

Relationships between word meanings

- Homonymy
- Polysemy
- Synonymy
- Antonymy
- Hypernomy
- Hyponomy
- Meronomy

Homonymy

- Lexemes that share a form
  - Phonological, orthographic or both
- But have unrelated, distinct meanings
- Clear example:
  - Bat (wooden stick-like thing) vs
  - Bat (flying scary mammal thing)
  - Or bank (financial institution) versus bank (riverside)
- Can be homophones, homographs, or both:
  - Homophones:
    - Write and right
    - Piece and peace

Homonymy causes problems for NLP applications

- Text-to-Speech
  - Same orthographic form but different phonological form
    - bass vs bass
- Information retrieval
  - Different meanings same orthographic form
    - QUERY: bat care
- Machine Translation
- Speech recognition

Polysemy

- The bank is constructed from red brick
  I withdrew the money from the bank
- Are those the same sense?
- Or consider the following WSJ example
  - While some banks furnish sperm only to married women, others are less restrictive
  - Which sense of bank is this?
    - Is it distinct from (homonymous with) the river bank sense?
    - How about the savings bank sense?

Polysemy

- A single lexeme with multiple related meanings (bank the building, bank the financial institution)
- Most non-rare words have multiple meanings
  - The number of meanings is related to its frequency
  - Verbs tend more to polysemy
  - Distinguishing polysemy from homonymy isn't always easy (or necessary)

Metaphor and Metonymy

- Specific types of polysemy
- Metaphor:
  - Germany will pull Slovenia out of its economic slump.
  - I spent 2 hours on that homework.
- Metonymy
  - The White House announced yesterday.
  - This chapter talks about part-of-speech tagging
  - Bank (building) and bank (financial institution)

How do we know when a word has more than one sense?

- ATIS examples
  - Which flights serve breakfast?
  - Does America West serve Philadelphia?
- The “zeugma” test:
  - ?Does United serve breakfast and San Jose?

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 successfully 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.
  - Why should that be?
  - 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 and H_20

Some more terminology

- Lemmas and wordforms
  - A **lexeme** is an abstract pairing of meaning and form
  - A **lemma** or **citation form** is the grammatical form that is used to represent a **lexeme**.
    - *Carpet* is the lemma for *carpets*
    - *Dormir* is the lemma for *duermes*.
  - Specific surface forms *carpets, sung, duermes* are called **wordforms**
- The lemma *bank* has two **senses**:
  - Instead, a **bank** can hold the investments in a custodial account in the client's name
  - But as agriculture burgeons on the east **bank**, the river will shrink even more.
- A **sense** is a discrete representation of one aspect of the meaning of a word

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, for instance, became a kind of **big** sister to Benjamin.
  - ?Miss Nelson, for instance, 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 their meaning
- Otherwise, they are very similar!
  - dark / light
  - short / long
  - hot / cold
  - up / down
  - in / out
- More formally: antonyms can
  - define a binary opposition or at opposite ends of a scale (*long/short, fast/slow*)
  - Be **reversives**: *rise/fall, up/down*

Hyponymy

- 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*
  - *dog* is a hyponym of *animal*
  - *mango* is a hyponym of *fruit*
- Conversely
  - *vehicle* is a hypernym/superordinate of *car*
  - *animal* is a hypernym of *dog*
  - *fruit* is a hypernym of *mango*

superordinate	vehicle	fruit	furniture	mammal
hyponym	car	mango	chair	dog

Hypernymy more formally

- Extensional:
  - The class denoted by the superordinate
  - 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
- Hyponymy is usually transitive
  - (A hypo B and B hypo C entails A hypo C)

II. WordNet

- A hierarchically organized lexical database
- On-line thesaurus + aspects of a dictionary
  - Versions for other languages are under development

Category	Unique Forms
Noun	117,097
Verb	11,488
Adjective	22,141
Adverb	4,601

WordNet

- Where it is:
  - http://wordnetweb.princeton.edu/perl/webwn

Format of Wordnet Entries

The noun "bass" has 8 senses in WordNet.

1. $bass^1$ - (the lowest part of the musical range)
2. $bass^2$, bass $part^1$ - (the lowest part in polyphonic music)
3. $bass^3$, $basso^1$ - (an adult male singer with the lowest voice)
4. sea $bass^1$, $bass^4$ - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater $bass^1$, $bass^5$ - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. $bass^6$, bass $voice^1$, $basso^2$ - (the lowest adult male singing voice)
7. $bass^7$ - (the member with the lowest range of a family of musical instruments)
8. $bass^8$ - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

The adjective "bass" has 1 sense in WordNet.

1. $bass^1$, $deep^6$ - (having or denoting a low vocal or instrumental range)
   "a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"

WordNet Noun Relations

Relation	Also called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$breakfast^1 \rightarrow meal^1$
Hyponym	Subordinate	From concepts to subtypes	$meal^1 \rightarrow lunch^1$
Member Meronym	Has-Member	From groups to their members	$faculty^2 \rightarrow professor^1$
Has-Instance		From concepts to instances of the concept	$composer^1 \rightarrow Bach^1$
Instance		From instances to their concepts	$Austen^1 \rightarrow author^1$
Member Holonym	Member-Of	From members to their groups	$copilot^1 \rightarrow crew^1$
Part Meronym	Has-Part	From wholes to parts	$table^2 \rightarrow leg^3$
Part Holonym	Part-Of	From parts to wholes	$course^7 \rightarrow meal^1$
Antonym		Opposites	$leader^1 \rightarrow follower^1$

WordNet Verb Relations

Relation	Definition	Example
Hypernym	From events to superordinate events	$fly^9 \rightarrow travel^5$
Troponym	From a verb (event) to a specific manner elaboration of that verb	$walk^1 \rightarrow stroll^1$
Entails	From verbs (events) to the verbs (events) they entail	$snore^1 \rightarrow sleep^1$
Antonym	Opposites	$increase^1 \Longleftrightarrow decrease^1$

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


Sense 7
bass --
(the member with the lowest range of a family of
musical instruments)
=> musical instrument, instrument
   => device
      => instrumentality, instrumentation
         => artifact, artefact
            => whole, unit
               => object, physical object
                  => physical entity
                     => entity
```

How is “sense” defined in WordNet?

- The set of near-synonyms for a WordNet sense is called a synset (synonym set); it’s their version of a sense or a concept
- Example: chump as a noun to mean
  - ‘a person who is gullible and easy to take advantage of’

{chump^{1}, fool^{2}, gull^{1}, mark^{9}, patsy^{1}, fall guy^{1}, sucker^{1}, soft touch^{1}, mug^{2}}

- Each of these senses share this same gloss
- Thus for WordNet, the meaning of this sense of chump *is* this list.

Word Sense Disambiguation (WSD)

- Given
  - a word in context,
  - A fixed inventory of potential word senses
- decide which sense of the word this is.
  - English-to-Spanish MT
    - Inventory is set of Spanish translations
  - Speech Synthesis
    - Inventory is homographs with different pronunciations like *bass* and *bow*
  - Automatic indexing of medical articles
    - MeSH (Medical Subject Headings) thesaurus entries

Two variants of WSD task

- Lexical Sample task
  - Small pre-selected set of target words
  - And inventory of senses for each word
- All-words task
  - Every word in an entire text
  - A lexicon with senses for each word
  - Sort of like part-of-speech tagging
    - Except each lemma has its own tagset

Approaches

- Supervised

- Semi-supervised
  - Unsupervised
    - Dictionary-based techniques
    - Selectional Association
  - Lightly supervised
    - Bootstrapping
    - Preferred Selectional Association

Supervised Machine Learning Approaches

- Supervised machine learning approach:
  - a training corpus of ?
  - used to train a classifier that can tag words in new text
  - Just as we saw for part-of-speech tagging, statistical MT.
- 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?

WordNet Bass

The noun ``bass" has 8 senses in WordNet

1. bass - (the lowest part of the musical range)
2. bass, bass part - (the lowest part in polyphonic music)
3. bass, basso - (an adult male singer with the lowest voice)
4. sea bass, bass - (flesh of lean-fleshed saltwater fish of the family Serranidae)
5. freshwater bass, bass - (any of various North American lean-fleshed freshwater fishes especially of the genus Micropterus)
6. bass, bass voice, basso - (the lowest adult male singing voice)
7. bass - (the member with the lowest range of a family of musical instruments)
8. bass -(nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

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

Supervised WSD 3: Extract feature vectors

- 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?''

- dishes

- bass

- washing *dishes*.
- simple *dishes* including
- convenient *dishes* to
- of *dishes* and

- free *bass* with
- pound *bass* of
- and *bass* player
- his *bass* while

- “In our house, everybody has a career and none of them includes washing *dishes*,” he says.
- In her tiny kitchen at home, Ms. Chen works efficiently, stir-frying several simple *dishes,* including braised pig’s ears and chcken livers with green peppers.
- Post quick and convenient *dishes* to fix when your in a hurry.
- Japanese cuisine offers a great variety of *dishes* and regional specialties

- We need more good teachers – right now, there are only a half a dozen who can play the free *bass* with ease.
- Though still a far cry from the lake's record 52-pound *bass* of a decade ago, "you could fillet these fish again, and that made people very, very happy." Mr. Paulson says.
- An electric guitar and *bass* player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations again.
- Lowe caught his *bass* while fishing with pro Bill Lee of Killeen, Texas, who is currently in 144th place with two bass weighing 2-09.

Feature vectors

- A simple representation for each observation (each instance of a target word)
  - Vectors of sets of feature/value pairs
    - I.e. files of comma-separated values
  - These vectors should represent the window of words around the target

    How big should that window be?

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 (regardless of position)
      - Typically limited to frequency counts

Examples

- Example text (WSJ)
  - An electric guitar and **bass** player stand off to one side not really part of the scene, just as a sort of nod to gringo expectations perhaps
  - Assume a window of +/- 2 from the target

Examples

- Example text
  - An electric guitar and bass player stand off to one side not really part of the scene, just as a sort of nod to gringo expectations perhaps
  - Assume a window of +/- 2 from the target

Collocational

- Position-specific information about the words in the window
- guitar and bass player stand
  - [guitar, NN, and, CC, player, NN, stand, VB]
  - $Word_{n-2,}$ $POS_{n-2,}$ $word_{n-1,}$ $POS_{n-1,}$ $Word_{n+1}$ POS_{n+1}…
  - In other words, a vector consisting of
  - [position n word, position n part-of-speech…]

Bag-of-words

- Information about the words that occur within the window.
- First derive a set of terms to place in the vector.
- Then note how often each of those terms occurs in a given window.

Co-Occurrence Example

- Assume we've settled on a possible vocabulary of 12 words that includes guitar and player but not and and stand
- guitar and bass player stand
  - [0,0,0,1,0,0,0,0,0,1,0,0]
  - Which are the counts of words predefined as e.g.,
  - [fish,fishing,viol, guitar, double,cello…

Classifiers

- Once we cast the WSD problem as a classification problem, then all sorts of techniques are possible
  - Naïve Bayes (the easiest thing to try first)
  - Decision lists
  - Decision trees
  - Neural nets
  - Support vector machines
  - Nearest neighbor methods…

Classifiers

- The choice of technique, in part, depends on the set of features that have been used
  - Some techniques work better/worse with features with numerical values
  - Some techniques work better/worse with features that have large numbers of possible values
    - For example, the feature **the word to the left** has a fairly large number of possible values

Naïve Bayes

- ŝ = $\arg\max_{s \in S}$ p(s|V), or $\arg\max_{s \in S} \frac{p(V|s)p(s)}{p(V)}$
- Where s is one of the senses S possible for a word w and V the input vector of feature values for w
- Assume features *independent*, so probability of V is the product of probabilities of each feature, given s, so
- $p(V|s) = \prod_{j=1}^{n} p(v_j|s)$ p(V) same for any ŝ
- Then $\hat{s} = \arg\max_{s \in S} p(s) \prod_{j=1}^{n} p(v_j|s)$

- How do we estimate p(s) and $p(v_j|s)$?
  - $p(s_i)$ is max. likelihood estimate from a sense-tagged corpus ($count(s_i,w_j)/count(w_j)$) – how likely is bank to mean 'financial institution' over all instances of bank?
  - $P(v_j|s)$ is max. likelihood of each feature given a candidate sense ($count(v_j,s)/count(s)$) – how likely is the previous word to be 'river' when the sense of bank is 'financial institution'
- Calculate $\hat{s} = \arg\max_{s \in S} p(s) \prod_{j=1}^{n} p(v_j|s)$ for each possible sense and take the highest scoring sense as the most likely choice

Naïve Bayes Test

- On a corpus of examples of uses of the word line, naïve Bayes achieved about 73% correct
- Good?

Decision Lists: another popular method

- A case statement….

Rule		Sense
fish within window	$\Rightarrow$	$\mathbf{bass}^1$
striped bass	$\Rightarrow$	$\mathbf{bass}^1$
guitar within window	$\Rightarrow$	$\mathbf{bass}^2$
bass player	$\Rightarrow$	$\mathbf{bass}^2$
piano within window	$\Rightarrow$	$\mathbf{bass}^2$
tenor within window	$\Rightarrow$	$\mathbf{bass}^2$
sea bass	$\Rightarrow$	$\mathbf{bass}^1$
play/V bass	$\Rightarrow$	$\mathbf{bass}^2$
river within window	$\Rightarrow$	$\mathbf{bass}^1$
violin within window	$\Rightarrow$	$\mathbf{bass}^2$
salmon within window	$\Rightarrow$	$\mathbf{bass}^1$
on bass	$\Rightarrow$	$\mathbf{bass}^2$
bass are	$\Rightarrow$	$\mathbf{bass}^1$

Learning Decision Lists

- Restrict the lists to rules that test a single feature (1-decisionlist rules)
- Evaluate each possible test and rank them based on how well they work.
- Glue the top-N tests together and call that your decision list.

Yarowsky

- On a binary (homonymy) distinction used the following metric to rank the tests

$$\frac{P(\mathrm{Sense}_1 \mid Feature)}{P(\mathrm{Sense}_2 \mid Feature)}$$

- This gives about 95% on this test…

WSD Evaluations and baselines

- *In vivo* versus *in vitro* evaluation
- In vitro evaluation is most common now
  - Exact match **accuracy**
    - % of words tagged identically with manual sense tags
  - Usually evaluate using held-out data from same labeled corpus
    - Problems?
    - Why do we do it anyhow?
- Baselines
  - Most frequent sense
  - The Lesk algorithm

Most Frequent Sense

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

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%

Unsupervised Methods WSD: Dictionary/Thesaurus methods

- The Lesk Algorithm
- Selectional Restrictions

Simplified Lesk

The **bank** can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.

given the following two WordNet senses:

$bank^1$	Gloss:	a financial institution that accepts deposits and channels the money into lending activities
	Examples:	"he cashed a check at the bank", "that bank holds the mortgage on my home"
$bank^2$	Gloss:	sloping land (especially the slope beside a body of water)
	Examples:	"they pulled the canoe up on the bank", "he sat on the bank of the river and watched the currents"

Original Lesk: pine cone

pine	1	kinds of evergreen tree with needle-shaped leaves
	2	waste away through sorrow or illness
cone	1	solid body which narrows to a point
	2	something of this shape whether solid or hollow
	3	fruit of certain evergreen trees

Corpus Lesk

- Add corpus examples to glosses and examples
- The best performing variant

Disambiguation via Selectional Restrictions

- "Verbs are known by the company they keep"
  - Different verbs select for different thematic roles

    wash the *dishes* (takes washable-thing as patient)

    serve delicious *dishes* (takes food-type as patient)
- Method: another semantic attachment in grammar
  - Semantic attachment rules are applied as sentences are syntactically parsed, e.g.

    VP --> V NP

    V→ serve <theme> {theme:food-type}
  - Selectional restriction violation: no parse

- But this means we must:
  - Write selectional restrictions for each sense of each predicate – or use FrameNet
    - Serve alone has 15 verb senses
  - Obtain hierarchical type information about each argument (using WordNet)
    - How many hypernyms does dish have?
    - How many words are hyponyms of dish?
- But also:
  - Sometimes selectional restrictions don't restrict enough (Which dishes do you like?)
  - Sometimes they restrict too much (Eat dirt, worm! I'll eat my hat!)
- Can we take a statistical approach?

Semi-supervised Bootstrapping

- What if you don't have enough data to train a system…
- Bootstrap
  - Pick a word that you as an analyst think will co-occur with your target word in particular sense
  - Grep through your corpus for your target word and the hypothesized word
  - Assume that the target tag is the right one

Bootstrapping

- For bass
  - Assume play occurs with the music sense and fish occurs with the fish sense

Sentences extracting using "fish" and "play"

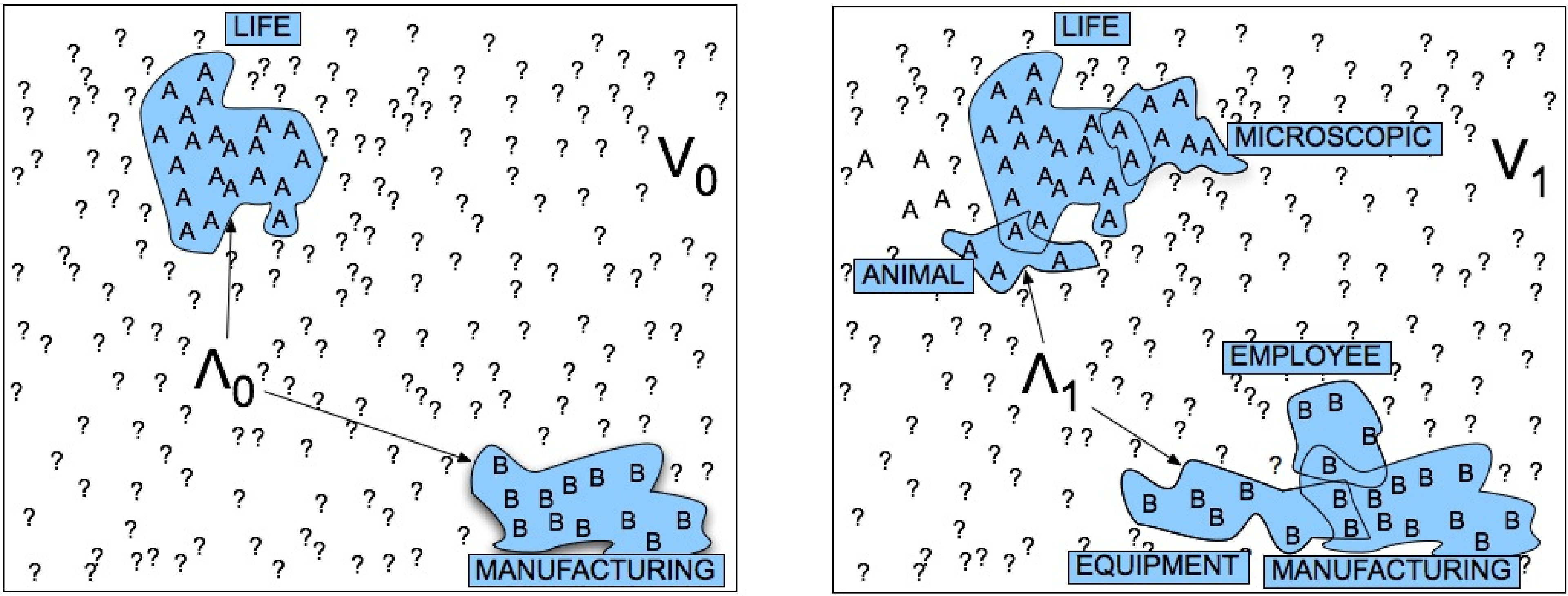
We need more good teachers – right now, there are only a half a dozen who can **play** the free **bass** with ease. An electric guitar and **bass play**er stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps. When the New Jersey Jazz Society, in a fund-raiser for the American Jazz Hall of Fame, honors this historic night next Saturday, Harry Goodman, Mr. Goodman's brother and **bass play**er at the original concert, will be in the audience with other family members.	
The researchers said the worms spend part of their life cycle in such **fish** as Pacific salmon and striped **bass** and Pacific rockfish or snapper. And it all started when **fish**ermen decided the striped **bass** in Lake Mead were too skinny. Though still a far cry from the lake's record 52-pound **bass** of a decade ago, "you could fillet these **fish** again, and that made people very, very happy," Mr. Paulson says.	

Where do the seeds come from?

1) Hand labeling
2) "One sense per discourse":
   - The sense of a word is highly consistent within a document – Yarowsky (1995)
   - True for topic dependent words
   - Not so true for other POS like adjectives and verbs, e.g. make, take
   - Krovetz (1998) "More than one sense per discourse" argues it isn't true at all once you move to fine-grained senses
3) One sense per collocation:
   - A word reoccurring in collocation with the same word will almost surely have the same sense.

Slide adapted from Chris Manning

Stages in the Yarowsky bootstrapping algorithm



(a) (b)

Problems

- Given these general ML approaches, how many classifiers do I need to perform WSD robustly
  - One for each ambiguous word in the language
- How do you decide what set of tags/labels/senses to use for a given word?
  - Depends on the application

WordNet Bass

- Tagging with this set of senses is an impossibly hard task that's probably overkill for any realistic application

1. bass - (the lowest part of the musical range)
2. bass, bass part - (the lowest part in polyphonic music)
3. bass, basso - (an adult male singer with the lowest voice)
4. sea bass, bass - (flesh of lean-fleshed saltwater fish of the family Serranidae)
5. freshwater bass, bass - (any of various North American lean-fleshed freshwater fishes especially of the genus Micropterus)
6. bass, bass voice, basso - (the lowest adult male singing voice)
7. bass - (the member with the lowest range of a family of musical instruments)
8. bass -(nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Senseval History

- ACL-SIGLEX workshop (1997)
  - Yarowsky and Resnik paper
- SENSEVAL-I (1998)
  - Lexical Sample for English, French, and Italian
- SENSEVAL-II (Toulouse, 2001)
  - Lexical Sample and All Words
  - Organization: Kilkgarriff (Brighton)
- SENSEVAL-III (2004)
- SENSEVAL-IV -> SEMEVAL (2007)

WSD Performance

- Varies widely depending on how difficult the disambiguation task is
- Accuracies of over 90% are commonly reported on some of the classic, often fairly easy, WSD tasks (pike, star, interest)
- Senseval brought careful evaluation of difficult WSD (many senses, different POS)
- Senseval 1: more fine grained senses, wider range of types:
  - Overall: about 75% accuracy
  - Nouns: about 80% accuracy
  - Verbs: about 70% accuracy

Summary

- Lexical Semantics
  - Homonymy, Polysemy, Synonymy
  - Thematic roles
- Computational resource for lexical semantics
  - WordNet
- Task
  - Word sense disambiguation