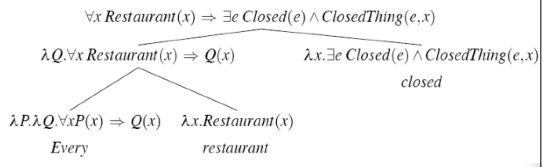
CSCI 5832
Natural Language Processing

Jim Martin
Lecture 20

Today 4/3

- Finish semantics
  - Dealing with quantifiers
  - Dealing with ambiguity
- Lexical Semantics
  - Wordnet
  - WSD

Every Restaurant Closed



Problem

- Every restaurant has a menu.

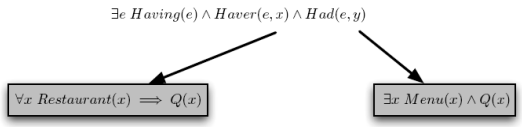
$$\forall x\, Restaurant(x) \Rightarrow \exists y\, (Menu(y) \wedge \exists e\, (Having(e) \wedge Haver(e,x) \wedge Had(e,y)))$$

$$\exists y\, Menu(y) \wedge \forall x\, (Restaurant(x) \Rightarrow \exists e\, (Having(e) \wedge Haver(e,x) \wedge Had(e,y)))$$

Problem

- The current approach just gives us 1 interpretation.
  - Which one we get is based on the order in which the quantifiers are added into the representation.
  - But the syntax doesn't really say much about that so it shouldn't be driving the placement of the quantifiers
    - It should focus on the argument structure mostly

What We Really Want



Store and Retrieve

- Now given a representation like that we can get all the meanings out that we want by
  - Retrieving the quantifiers one at a time and placing them in front.
  - The order determines the scoping (the meaning).

Store

- The Store..

$$\exists e\, Having(e) \wedge Haver(e, s_1) \wedge Had(e, s_2)$$
$$(\lambda Q.\forall x\, Restaurant(x) \Rightarrow Q(x), 1),$$
$$(\lambda Q.\exists x\, Menu(x) \wedge Q(x), 2)$$

Retrieve

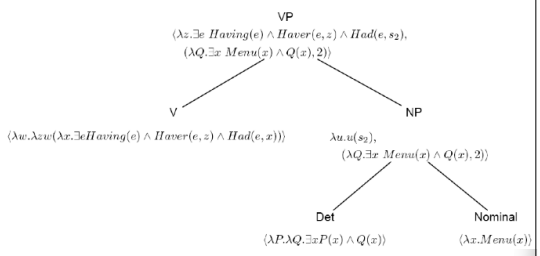
- Use lambda reduction to retrieve from the store incorporate the arguments in the right way.
  - Retrieve element from the store and apply it to the core representation
  - With the variable corresponding to the retrieved element as a lambda variable
  - Huh?

Retrieve

- Example pull out 2 first (that's s2).

$$\lambda Q.\exists x\,(Menu(x) \wedge Q(x))$$
$$(\lambda s_2.\exists e\,Having(e) \wedge Haver(e,s_1) \wedge Had(e,s_2))$$

Retrieve



Break

- CAETE students...
  - Quizzes have been turned in to CAETE for distribution back to you.
  - Next in-class quiz is 4/17.
    - That's 4/24 for you

Break

- Quiz review

WordNet

- WordNet is a database of facts about words
  - Meanings and the relations among them
- www.cogsci.princeton.edu/~wn
  - Currently about 100,000 nouns, 11,000 verbs, 20,000 adjectives, and 4,000 adverbs
  - Arranged in separate files (DBs)

WordNet Relations

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

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
```


Inside Words

- Paradigmatic relations connect lexemes together in particular ways but don't say anything about what the meaning representation of a particular lexeme should consist of.
- That's what I mean by inside word meanings.

Inside Words

- Various approaches have been followed to describe the semantics of lexemes. We'll look at only a few…
  - Thematic roles in predicate-bearing lexemes
  - Selection restrictions on thematic roles
  - Decompositional semantics of predicates
  - Feature-structures for nouns

Inside Words

- Thematic roles: more on the stuff that goes on inside verbs.
  - Thematic roles are semantic generalizations over the specific roles that occur with specific verbs.
  - I.e. Takers, givers, eaters, makers, doers, killers, all have something in common
    - -er
    - They're all the agents of the actions
  - We can generalize across other roles as well to come up with a small finite set of such roles

Thematic Roles

Thematic Role	Example
AGENT	*The waiter* spilled the soup.
EXPERIENCER	*John* has a headache.
FORCE	*The wind* blows debris from the mall into our yards.
THEME	Only after Benjamin Franklin broke *the ice*...
RESULT	The French government has built a *regulation-size baseball diamond*...
CONTENT	Mona asked *"You met Mary Ann at a supermarket?"*
INSTRUMENT	He turned to poaching catfish, stunning them *with a shocking device*...
BENEFICIARY	Whenever Ann Callahan makes hotel reservations *for her boss*...
SOURCE	I flew in *from Boston*.
GOAL	I drove *to Portland*.

Thematic Roles

- Takes some of the work away from the verbs.
  - It's not the case that every verb is unique and has to completely specify how all of its arguments uniquely behave.
  - Provides a locus for organizing semantic processing
  - It permits us to distinguish near surface-level semantics from deeper semantics

Linking

- Thematic roles, syntactic categories and their positions in larger syntactic structures are all intertwined in complicated ways. For example…
  - AGENTS are often subjects
  - In a `VP->V NP NP` rule, the first NP is often a GOAL and the second a THEME

Resources

- There are 2 major English resources out there with thematic-role-like data
  - PropBank
    - Layered on the Penn TreeBank
      - Small number (25ish) labels
  - FrameNet
    - Based on a theory of semantics known as frame semantics.
      - Large number of frame-specific labels

Deeper Semantics

- From the WSJ…
  - *He melted her reserve with a husky-voiced paean to her eyes.*
  - If we label the constituents He and her reserve as the Melter and Melted, then those labels lose any meaning they might have had.
  - If we make them Agent and Theme then we don't have the same problems

Problems

- What exactly is a role?
- What's the right set of roles?
- Are such roles universals?
- Are these roles atomic?
  - I.e. Agents
    - Animate, Volitional, Direct causers, etc
- Can we automatically label syntactic constituents with thematic roles?

Selection Restrictions

- Last time
  - *I want to eat someplace near campus*
  - Using thematic roles we can now say that eat is a predicate that has an AGENT and a THEME
    - What else?
  - And that the AGENT must be capable of eating and the THEME must be something typically capable of being eaten

As Logical Statements

- For eat…
  - Eating(e) ^Agent(e,x)^ Theme(e,y)^Food(y)

  (adding in all the right quantifiers and lambdas)

Back to WordNet

- Use WordNet hyponyms (type) to encode the selection restrictions

```
Sense 1
hamburger, beefburger --
(a fried cake of minced beef served on a bun)
=> sandwich
   => snack food
      => dish
         => nutriment, nourishment, nutrition...
            => food, nutrient
               => substance
                  => matter
                     => physical entity
                        => entity
```


Specificity of Restrictions

- Consider the verbs imagine, lift and diagonalize in the following examples
  - To *diagonalize* a matrix is to find its eigenvalues
  - Atlantis *lifted* Galileo from the pad
  - *Imagine* a tennis game
- What can you say about THEME in each with respect to the verb?
- Some will be high up in the WordNet hierarchy, others not so high…

Problems

- Unfortunately, verbs are polysemous and language is creative… WSJ examples…
  - *… ate glass on an empty stomach accompanied only by water and tea*
  - *you can't eat gold for lunch if you're hungry*
  - *… get it to try to eat Afghanistan*

Solutions

- Eat glass
  - Not really a problem. It is actually about an eating event
- Eat gold
  - Also about eating, and the can't creates a scope that permits the THEME to not be edible
- Eat Afghanistan
  - This is harder, its not really about eating at all

Discovering the Restrictions

- Instead of hand-coding the restrictions for each verb, can we discover a verb's restrictions by using a corpus and WordNet?
  1. Parse sentences and find heads
  2. Label the thematic roles
  3. Collect statistics on the co-occurrence of particular headwords with particular thematic roles
  4. Use the WordNet hypernym structure to find the most meaningful level to use as a restriction

Motivation

- Find the lowest (most specific) common ancestor that covers a significant number of the examples

```
Sense 1
hamburger, beefburger --
(a fried cake of minced beef served on a bun)
=> sandwich
   => snack food
      => dish
         => nutriment, nourishment, nutrition...
            => food, nutrient
               => substance
                  => matter
                     => physical entity
                        => entity
```

WSD and Selection Restrictions

- Word sense disambiguation refers to the process of selecting the right sense for a word from among the senses that the word is known to have
- Semantic selection restrictions can be used to disambiguate
  - Ambiguous arguments to unambiguous predicates
  - Ambiguous predicates with unambiguous arguments
  - Ambiguity all around

WSD and Selection Restrictions

- Ambiguous arguments
  - Prepare a dish
  - Wash a dish
- Ambiguous predicates
  - Serve Denver
  - Serve breakfast
- Both
  - Serves vegetarian dishes

WSD and Selection Restrictions

- This approach is complementary to the compositional analysis approach.
  - You need a parse tree and some form of predicate-argument analysis derived from
    - The tree and its attachments
    - All the word senses coming up from the lexemes at the leaves of the tree
    - Ill-formed analyses are eliminated by noting any selection restriction violations

Problems

- As we saw last time, selection restrictions are violated all the time.
- This doesn't mean that the sentences are ill-formed or preferred less than others.
- This approach needs some way of categorizing and dealing with the various ways that restrictions can be violated

Supervised ML Approaches

- That's too hard… try something empirical
- In supervised machine learning approaches, a training corpus of words tagged in context with their sense is used to train a classifier that can tag words in new text (that reflects the training text)

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).

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)

Representations

- Most supervised ML approaches require a very simple representation for the input training data.
  - Vectors of sets of feature/value pairs
    - I.e. files of comma-separated values
- So our first task is to extract training data from a corpus with respect to a particular instance of a target word
  - This typically consists of a characterization of the window of text surrounding the target

Representations

- This is where ML and NLP intersect
  - If you stick to trivial surface features that are easy to extract from a text, then most of the work is in the ML system
  - If you decide to use features that require more analysis (say parse trees) then the ML part may be doing less work (relatively) if these features are truly informative

Surface Representations

- Collocational and co-occurrence information
  - Collocational
    - Encode features about the words that appear in specific positions to the right and left of the target word
      - Often limited to the words themselves as well as they're part of speech
  - Co-occurrence
    - Features characterizing the 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, CJC, player, NN, stand, VVB]
  - In other words, a vector consisting of
  - [position n word, position n part-of-speech…]

Co-occurrence

- 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]

Classifiers

- Once we cast the WSD problem as a classification problem, then all sorts of techniques are possible
  - Naïve Bayes (the right thing to try first)
  - Decision lists
  - Decision trees
  - MaxEnt
  - 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

- Argmax P(sense|feature vector)
- Rewriting with Bayes and assuming independence of the features

Naïve Bayes

- P(s) … just the prior of that sense.
  - Just as with part of speech tagging, not all senses will occur with equal frequency
- $P(v_i|s)$… conditional probability of some particular feature/value combination given a particular sense
- You can get both of these from a tagged corpus with the features encoded

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…

Rule		Sense
fish within window	⇒	bass^1
striped bass	⇒	bass^1
guitar within window	⇒	bass^2
bass player	⇒	bass^2
piano within window	⇒	bass^2
tenor within window	⇒	bass^2
sea bass	⇒	bass^1
play/V *bass*	⇒	bass^2
river within window	⇒	bass^1
violin within window	⇒	bass^2
salmon within window	⇒	bass^1
on bass	⇒	bass^2
bass are	⇒	bass^1

Learning DLs

- Restrict the lists to rules that test a single feature (1-dl 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

- This gives about 95% on this test…
- Is this better than the 73% on line we noted earlier?

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

Bass Results

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 player** 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 player** 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.

Bootstrapping

- Perhaps better
  - Use the little training data you have to train an inadequate system
  - Use that system to tag new data.
  - Use that larger set of training data to train a new system

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)

Next Time

- On to Chapter 22 (Information Extraction)