

CSCI 5832

Natural Language Processing

Lecture 22
Jim Martin

4/24/07

CSCI 5832 Spring 2006

1

Today: 4/12

- **More on meaning**
- **Lexical Semantics**
 - **A seemingly endless set of random facts about words**

4/24/07

CSCI 5832 Spring 2006

2

Meaning

- Traditionally, meaning in language has been studied from three perspectives
 - The meaning of a **text or discourse**
 - The meanings of **individual sentences or utterances**
 - The meanings of **individual words**
- We started in the middle, now we'll move down to words and then back up to discourse.

4/24/07

CSCI 5832 Spring 2006

3

Word Meaning

- We didn't assume much about the meaning of words when we talked about sentence meanings
 - Verbs provided a template-like predicate argument structure
 - **Number of arguments**
 - **Position and syntactic type**
 - **Names for arguments**
 - Nouns were practically meaningless constants
- There has be more to it than that

4/24/07

CSCI 5832 Spring 2006

4

Theory

- From the theory-side we'll proceed by looking at
 - The external relational structure among words
 - The internal structure of words that determines where they can go and what they can do

4/24/07

CSCI 5832 Spring 2006

5

Applications

- We'll take a look at...
 - Enabling resources
 - WordNet, FrameNet
 - Enabling technologies
 - Word sense disambiguation
 - Word-based applications
 - Search engines
- But first the facts and some theorizing

4/24/07

CSCI 5832 Spring 2006

6

Preliminaries

- **What's a word?**
 - Types, tokens, stems, roots, inflected forms, etc... Ugh.
 - **Lexeme**: An entry in a lexicon consisting of a pairing of a base form with a single meaning representation
 - **Lexicon**: A collection of lexemes

4/24/07

CSCI 5832 Spring 2006

7

Complications

- **Homonymy**:
 - Lexemes that share a form
 - Phonological, orthographic or both
 - **Clear example**:
 - Bat (wooden stick-like thing) vs
 - Bat (flying scary mammal thing)

4/24/07

CSCI 5832 Spring 2006

8

Problems for Applications

- **Text-to-Speech**
 - Same orthographic form but different phonological form
 - Content vs content
- **Information retrieval**
 - Different meanings same orthographic form
 - QUERY: *router repair*
- **Translation**
- **Speech recognition**

4/24/07

CSCI 5832 Spring 2006

9

Homonymy

- The problematic part of understanding homonymy isn't with the forms, it's the meanings.
 - An intuition with true homonymy is coincidence
 - It's a coincidence in English that bat and bat mean what they do.
 - Nothing particularly important would happen to anything else in English if we used a different word for flying rodents

4/24/07

CSCI 5832 Spring 2006

10

Polysemy

- The case where a single lexeme has multiple meanings associated with it.
 - Most words with moderate frequency have multiple meanings
 - The actual number of meanings is related to a word's frequency
 - Verbs tend more to polysemy
 - Distinguishing polysemy from homonymy isn't always easy (or necessary)

4/24/07

CSCI 5832 Spring 2006

11

Polysemy

- 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?

4/24/07

CSCI 5832 Spring 2006

12

Polysemy Tests

- **ATIS examples**
 - Which flights serve breakfast?
 - Does America West serve Philadelphia?
 - Does United serve breakfast and San Jose?

4/24/07

CSCI 5832 Spring 2006

13

Relations

- **Inter-word relations...**
 - Synonymy
 - Antonymy
 - Hyponymy
 - Metonymy
 - ...

4/24/07

CSCI 5832 Spring 2006

14

Synonyms

- There really aren't any...
- Maybe not, but people think and act like there are so maybe there are...
- One test...
 - Two lexemes are synonyms if they can be successfully substituted for each other in all situations

4/24/07

CSCI 5832 Spring 2006

15

Synonyms

- What the heck does successfully mean?
 - Preserves the meaning
 - But may not preserve the acceptability based on notions of politeness, slang, register, genre, etc.
- Example:
 - Big and large?
 - That's my big brother
 - That's my large brother

4/24/07

CSCI 5832 Spring 2006

16

Hyponymy

- A hyponymy relation can be asserted between two lexemes when the meanings of the lexemes entail a subset relation
 - Since dogs are canids
 - Dog is a hyponym of canid and
 - Canid is a hypernym of dog

4/24/07

CSCI 5832 Spring 2006

17

Resources

- There are lots of lexical resources available these days...
 - Word lists
 - On-line dictionaries
 - Corpora
- The most ambitious one is WordNet
 - A database of lexical relations for English
 - Versions for other languages are under development

4/24/07

CSCI 5832 Spring 2006

18

WordNet

- Some out of date numbers

Category	Unique Forms	Number of Senses
Noun	94474	116317
Verb	10319	22066
Adjective	20170	29881
Adverb	4546	5677

4/24/07

CSCI 5832 Spring 2006

19

WordNet

- The critical thing to grasp about WordNet is the notion of a **synset**; it's their version of a sense or a concept
- Example: **table** as a verb to mean defer
 - > {postpone, hold over, table, shelve, set back, defer, remit, put off}
- For WordNet, the meaning of this sense of **table** is this list.

4/24/07

CSCI 5832 Spring 2006

20

WordNet Relations

Relation	De nition	Example
Hypernym	From concepts to superordinates	<i>breakfast</i> → <i>meal</i>
Hyponym	From concepts to subtypes	<i>meal</i> → <i>lunch</i>
Has-Member	From groups to their members	<i>faculty</i> → <i>professor</i>
Member-Of	From members to their groups	<i>copilot</i> → <i>crew</i>
Has-Part	From wholes to parts	<i>table</i> → <i>leg</i>
Part-Of	From parts to wholes	<i>course</i> → <i>meal</i>
Antonym	Opposites	<i>leader</i> → <i>follower</i>

4/24/07

CSCI 5832 Spring 2006

21

WordNet Hierarchies

```

Sense 3
bass, basso --
(an adult male singer with the lowest voice)
=> singer, vocalist
    => musician, instrumentalist, player
        => performer, performing artist
            => entertainer
                => person, individual, someone...
                    => life form, organism, being...
                        => entity, something
                            => causal agent, cause, causal agency
                                => entity, something

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

4/24/07

CSCI 5832 Spring 2006

22

Break

Quiz...

Average was 44 (out of 55)

SD was 7

Most popular month is May

4/24/07

CSCI 5832 Spring 2006

23

Break

1. **May**
2. **True**
3. **Treebank rules**
 - Nom -> Noun
 - Nom -> Noun Noun
 - Nom -> Noun Noun Noun...
4. **False**
5. **Next slide**
6. **[A flight][from][Boston][to][Miami]**
7. **Count and divide**

4/24/07

CSCI 5832 Spring 2006

24

Break

<i>An</i>	<i>evening</i>	<i>flight</i>
Det	NP	NP
	Nom Noun	Nom
		Nom Noun

4/24/07

CSCI 5832 Spring 2006

25

Break

<i>An</i>	<i>evening</i>	<i>flight</i>
Det ←	NP	NP
	Nom ←	Nom
	Noun	Noun

4/24/07

CSCI 5832 Spring 2006

26

Inside Words

- **Thematic roles: more on the stuff that goes on inside verbs.**
- **Qualia theory: what must be going inside nouns (they're not really just constants)**

4/24/07

CSCI 5832 Spring 2006

27

Inside Verbs

- **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 (or try to) across other roles as well**

4/24/07

CSCI 5832 Spring 2006

28

Thematic Roles

Thematic Role	De nition
AGENT	The volitional causer of an event
EXPERIENCER	The experiencer of an event
FORCE	The non-volitional causer of the event
THEME	The participant most directly affected by an event
RESULT	The end product of an event
CONTENT	The proposition or content of a propositional event
INSTRUMENT	An instrument used in an event
BENEFICIARY	The bene ciary of an event
SOURCE	The origin of the object of a transfer event
GOAL	The destination of an object of a transfer event

4/24/07

CSCI 5832 Spring 2006

29

Thematic Role Examples

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

4/24/07

CSCI 5832 Spring 2006

30

Why Thematic Roles?

- It's not the case that every verb is unique and has to introduce unique labels for all of its roles; thematic roles let us specify a fixed set of roles.
- More importantly it permits us to distinguish surface level shallow semantics from deeper semantics

4/24/07

CSCI 5832 Spring 2006

31

Example

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

4/24/07

CSCI 5832 Spring 2006

32

Tasks

- **Shallow semantic analysis is defined as**
 - **Assigning the right labels to the arguments of verb in a sentence**
 - Case role assignment
 - Thematic role assignment

Thematic Role	De nition
AGENT	The volitional causer of an event
EXPERIENCER	The experiencer of an event
FORCE	The non-volitional causer of the event
THEME	The participant most directly affected by an event
RESULT	The end product of an event
CONTENT	The proposition or content of a propositional event
INSTRUMENT	An instrument used in an event
BENEFICIARY	The bene ciary of an event
SOURCE	The origin of the object of a transfer event
GOAL	The destination of an object of a transfer event

4/24/07

CSCI 5832 Spring 2006

33

Example

- **Newswire text**
 - [agent *British forces*] [target *believe*] that [theme *Ali was killed in a recent air raid*]
 - British forces believe that [theme *Ali*] was [target *killed*] [temporal *in a recent air raid*]

4/24/07

CSCI 5832 Spring 2006

34

Resources

- **PropBank**
 - Annotate every verb in the Penn Treebank with its semantic arguments.
 - Use a fixed (25 or so) set of role labels (Arg0, Arg1...)
 - Every verb has a set of frames associated with it that indicate what its roles are.
 - So for *Give* we're told that Arg0 -> *Giver*

4/24/07

CSCI 5832 Spring 2006

35

Resources

- **Propbank**
 - Since it's built on the treebank we have the trees and the parts of speech for all the words in each sentence.
 - Since it's a corpus we have the statistical coverage information we need for training machine learning systems.

4/24/07

CSCI 5832 Spring 2006

36

Resources

- **Propbank**
 - Since it's the WSJ it contains some fairly odd (domain specific) word uses that don't match our intuitions of the normal use of the words
 - Similarly, the word distribution is skewed by the genre from "normal" English (whatever that means).
 - There's no unifying semantic theory behind the various frame files (*buy* and *sell* are essentially unrelated).

4/24/07

CSCI 5832 Spring 2006

37

Resources

- **FrameNet**
 - Instead of annotating a corpus, annotate domains of human knowledge a domain at a time (called frames)
 - Then within a domain annotate lexical items from within that domain.
 - Develop a set of semantic roles (called frame elements) that are based on the domain and **shared across** the lexical items in the frame.

4/24/07

CSCI 5832 Spring 2006

38

Cause_Harm Frame

Cause_harm

Definition:

The words in this frame describe situations in which an **Agent** or a **Cause** injures a **Victim**. The **Body_part** of the **Victim** which is most directly affected may also be mentioned in the place of the **Victim**. In such cases, the **Victim** is often indicated as a genitive modifier of the **Body_part**, in which case the **Victim** FE is indicated on a second FE layer.

4/24/07

CSCI 5832 Spring 2006

39

Lexical Units

Lexical Units

bash.v, batter.v, bayonet.v, beat up.v, beat.v, belt.v, biff.v, bludgeon.v, boil.v, break.v, bruise.v, buffet.v, burn.v, butt.v, cane.v, chop.v, claw.v, clout.v, club.v, crack.v, crush.v, cudgel.v, cuff.v, cut.v, elbow.v, electrocute.v, electrocution.n, flagellate.v, flog.v, fracture.v, gash.v, hammer.v, hit.v, horsewhip.v, hurt.v, impale.v, injure.v, jab.v, kick.v, knee.v, knife.v, knock.v, lash.v, maim.v, maul.v, mutilate.v, pelt.v, poison.v, poisoning.n, pummel.v, punch.v, slap.v, slice.v, smack.v, smash.v, spear.v, squash.v, stab.v, stone.v, strike.v, thwack.v, torture.v, transfix.v, welt.v, whip.v, wound.v

Created by hcb on Wed May 23 15:26:58 PDT 2001

4/24/07

CSCI 5832 Spring 2006

40

FrameNet

- Frames and frame elements are entities in a hierarchy.
 - **Cause_Harm** inherits from **Transitive_Action**
 - **Corporal_Punishment** inherits from **Cause_Harm**
 - The **victim** FE in **Cause_Harm** inherits from the **patient** FE of **Transitive_Action**
 - And the **evaluatee** of the **Corporal_Punishment** frame inherits from the **victim** of the **Cause_Harm** frame.

4/24/07

CSCI 5832 Spring 2006

41

FrameNet

- Framenet.icsi.berkeley.edu

4/24/07

CSCI 5832 Spring 2006

42

Next Time

I'll post readings for Ch. 19.

Tuesday we'll return to and finish information extraction

Thursday we'll turn to discourse (Chapter 20).

Final quiz will be on May 1.