December 2, 2004

Chapter 16

Lexical Semantics

Overview

- What is lexical semantics?; Applications
- Lexeme
- Lexical relations (homonymy, polysemy, synonymy, hyponymy)
- Internal structure of words: thematic roles, selectional restrictions, decomposition
- Generative lexicon: metaphor & metonymy
- Along the way, resources: WordNet, FrameNet

What is lexical semantics? (1/2)

- The study of the meanings of words.
- How do you (as speakers) know what a word means?
- How could you go about systematically describing it?

What is lexical semantics? (2/2)

- Lexical meanings are structured:
  - Predicate-argument structure
  - Relationships between senses
  - Entailments
  - Selectional restrictions
  - Possible potential for decomposition

Applications

- Information extraction
- Information retrieval
- Machine translation
- Natural language understanding

Some terminology: Lexemes

- Lexemes are a kind of signs: pairings of form and meaning.
- Form can be phonological or orthographic.
- Meaning is quantized as ‘senses’.
- A lexeme may have multiple forms (through inflectional morphology).
- . . . and multiple meanings (polysemy).
- The *lexicon* is a finite list of lexemes.

Relationships between lexemes

- Homonymy (homophony, homography)
- Polysemy
- Synonymy
- Hyponymy

Homonymy

- A relation that holds between two lexemes that have the same form but unrelated meanings.
- Complicated by the presence of two kinds of form.

	orth. form	phon. form	meaning
homonyms	same	same	unrelated
homographs	same	different	unrelated
homophones	different	same	unrelated
??	same	different	same

- Typically only consider lexemes with the same part of speech.
- Morphology can cause ‘partial’ homonyms, where paradigms only partially overlap, e.g., *found*.

Homonymy causes problems

- Spelling correction: confusables (*your*/*you're*)
- Speech recognition:
  - Homophones – which form to transcribe?
  - Homonyms – disrupt n-gram statistics
- Text-to-speech: pronunciation selection for homographs
- IR: homonyms, homographs – which documents are really relevant?

Polysemy

- A single lexeme with multiple related senses.
- *bank*: financial institution, biological repository
- How to distinguish polysemy from homonymy?
- How to determine how many senses a word has? (When to split and when to join?)

  serve, uncle, bat

- In what ways can the various senses be related?
- What kinds of problems does polysemy pose?
- Word sense disambiguation: How to tell which sense is intended in a given context.

Synonymy

- Two words are synonyms if they share at least one sense.
- Operationalize as a substitutability test.
- Substitutability might fail because:
  - No senses are shared
  - The sense required in a particular example is not shared
  - Particular shades of meaning are not shared: *price*/*fare*
  - Collocational constraints are violated: *big/large mistake*
  - Register constraints are violated: Example?

Hyponymy

- A relation that holds between two lexemes where one denotes a subclass of the other
- E.g.,: vehicle (hypernym) :: car (hyponym)
- Establish via an entailment test:

  That is a <hyponym>. $\Rightarrow$ That is a <hypernym>.

- Related to ontologies, taxonomies, and object hierarchies.

WordNet: Representing senses and semantic relations

- A large electronic database of lexical relations
- Browser- and (C) library- based access
- Consists of lexical entries corresponding to unique orthographic forms (within a part of speech), accompanied by sets of senses for each form.

WordNet Statistics

Category	Forms	Senses	Word-Sense Pairs
Noun	114648	79689	141690
Verb	11306	13508	24632
Adjective	21436	18563	31015
Adverb	4669	3664	5808
Total	152059	115424	203145

http://www.cogsci.princeton.edu/∼wn/

WordNets around the World

- Global WordNet Association lists WordNets for:

  Avestan, Baluchi, Basque, Bulgarian, Catalan,
  Czech, Danish, Dutch, English, Estonian, French,
  German, Greek, Hebrew, Hindi, Icelandic, Italian,
  Kannada, Kurdish, Latvian, Marathi, Moldavian,
  Norwegian, Old Persian, Oriya, Persian,
  Portuguese, Romanian, Russian, Sanskrit,
  Serbian, Slovenian, Spanish, Swedish, Tamil,
  Thai, Turkish

  http://www.globalwordnet.org/

WordNet Synset

- WordNet representations of meaning and meaning relations are organized around ‘synsets’ (synonym sets).

  { chump, fish, fool, gull, mark, patsy, fall guy, sucker, schlemiel, shlemiel, soft touch, mug }

- A word sense in WordNet *is* a synset.
- Semantic relations (hyponymy, membership, part-of, antonymy) are relations between synsets.

Internal Structure of Words

- Thematic roles
- Selectional restrictions
- Primitive decomposition
- Semantic fields

Thematic roles

- Deep roles:
  - Kim broke a bat.
  - $\exists e, x, y$ *Isa*(*e, Breaking*) $\wedge$ *Breaker*(*e,Kim*), $\wedge$ *BrokenThing*(*e,y*) $\wedge$ *Isa*(*y, bat*)
- Shallower 'theta' roles:
  - $\exists e, x, y$ *Isa*(*e, Breaking*) $\wedge$ **Agent**(*e,Kim*), $\wedge$ **Theme**(*e,y*) $\wedge$ *Isa*(*y, bat*)
- Theta roles intend to capture the similarities between participants across events.
- Possibly useful in syntax-semantics interface for statements of 'linking theory'.

Problems with Theta Roles

- Linking theory is notoriously prone to exceptions
- Originally confined only to NP and PP arguments of verbs
- Some verbs describe the same event from different perspectives.

  $\rightarrow$ Can't predict theta roles from the event itself

  - Amie bought the sandwich from Benson for $3
  - Benson sold Amie the sandwich for $3
  - Amie paid Benson $3 for the sandwich
- A modern, corpus-based approach: FrameNet

Selectional Restrictions

- Senses of lexemes enforce selectional restrictions on their arguments:
  - Which airlines serve Denver?
    - *ServedThing* is a geographical location
  - Which airlines serve breakfast?
    - *ServedThing* is a meal
- $\rightarrow$ Can be leveraged for word-sense disambiguation.
- Note that selectional restriction violations usually lead to incongruity and/or coerced readings, rather than strict ungrammaticality.

Specificity of selectional restrictions

- Varied:
  - In rehearsal, I often ask the musicians to imagine a tennis game.
  - They tell of jumping over beds they can't imagine clearing while awake.
  - I cannot even imagine what this lady does all day.
  - Atlantis lifted Galileo from the launch pad.
  - Mr. Kruger lifted the fish from the water.
  - To diagonalize a matrix is to find its eigenvalues.
- Not drawn from a limited set of primitives.

Representing Selectional Restrictions

- Semantic contribution of a verb like *eat*:

  $\exists e, x, y$ *Isa*(*e*, *Eating*) $\wedge$ *Agent*(*e*,*x*), $\wedge$ *Theme*(*e*,*y*)
  $\wedge$ *Isa*(*y*, *EdibleThing*)

- Representation of the phrase *ate a hamburger*:

  $\exists e, x, y$ *Isa*(*e*, *Eating*) $\wedge$ *Agent*(*e*,*x*), $\wedge$ *Theme*(*e*,*y*)
  $\wedge$ *Isa*(*y*, *EdibleThing*) $\wedge$ *Isa*(*y*, *hamburger*)

- Instead of defining concepts corresponding to the classes required for selectional restrictions, use WordNet synsets: { food, nutrient }

- { hamburger, beefburger } is a hyponym of { food, nutrient }

Creativity and the Lexicon

- We can use more word meanings than can be explicitly listed in the lexicon.
- The extensions are systematic, not haphazard.
- Productive processes for creating new senses include:
  - Metaphor
  - Metonymy

Metaphor

- Using metaphor, we refer to and reason about concepts using terminology appropriate to completely different kinds of concepts.
- CORPORATION AS PERSON
  - That doesn't **scare** Digital, which has grown to be the world's second-largest. . .
  - Triton Group Ltd., as company it helped **resuscitate**, has begun acquiring Fuqua shares.
  - But if it **changed its mind**, however, it would do so for investment reasons, the filing said.

Metonymy

- The use of one concept to refer to another concept closely related to it.
  - PRODUCT FOR PROCESS:

    GM killed the Fiero because it had dedicated a full-scale factory to building the plastic bodied car. . .
  - AUTHOR FOR WORKS

    He likes Shakespeare.
  - PLACE FOR INSTITUTION

    The White House had no comment.

Approaches to Metaphor and Metonymy

- Convention-based approaches hard-code metaphors such as CORPORATION AS PERSON and metonymies like PRODUCT FOR PROCESS.
- Reasoning-based approaches treat this as a problem for general, not necessarily linguistic, reasoning, such as analogical reasoning.

Summary

- What is lexical semantics?; Applications
- Lexeme
- Lexical relations (homonymy, polysemy, synonymy, hyponymy)
- Internal structure of words: thematic roles, selectional restrictions, decomposition
- Generative lexicon: metaphor & metonymy
- Along the way, resources: WordNet, FrameNet
- Next time: Dialogue & conversation agents