Outline of today’s lecture

Semantic relations

Polysemy

Word sense disambiguation

Grounding
Lexical semantics

- Limited domain: mapping to some knowledge base term(s). Knowledge base constrains possible meanings.
- Issues for broad coverage systems:
  - Boundary between lexical meaning and world knowledge.
  - Representing lexical meaning.
  - Acquiring representations.
  - Polysemy and multiword expressions.
Gary Larson’s approach to lexical meaning

“Now! That should clear up a few things around here!”
Approaches to lexical meaning

- Formal semantics: **extension** — what words denote (e.g., cat': the set of all cats).

- Semantic primitives: e.g., \textit{kill} means CAUSE (NOT (ALIVE)).

- Meaning postulates:

\[
\forall e, x, y [\text{kill}'(e, x, y) \rightarrow \exists e'[\text{cause}'(e, x, e') \land \text{die}'(e', y)]]
\]

- Ontological relationships: informal or formal (description logics): this lecture (informal approaches).

- Distributional approaches (lecture 8 and 9).
Approaches to lexical meaning

- **Formal semantics:** extension — what words denote (e.g., cat': the set of all cats).
- **Semantic primitives:** e.g., *kill* means CAUSE (NOT (ALIVE)).
- **Meaning postulates:**
  \[
  \forall e, x, y[\text{kill}'(e, x, y) \rightarrow \exists e'[\text{cause}'(e, x, e') \land \text{die}'(e', y)]]
  \]
- **Ontological relationships:** informal or formal (description logics): this lecture (informal approaches).
- **Distributional approaches** (lecture 8 and 9).
Approaches to lexical meaning

- Formal semantics: **extension** — what words denote (e.g., cat': the set of all cats).
- Semantic primitives: e.g., *kill* means CAUSE (NOT (ALIVE)).
- Meaning postulates:

  \[ \forall e, x, y [\text{kill}'(e, x, y) \rightarrow \exists e'[\text{cause}'(e, x, e') \land \text{die}'(e', y)]] \]

- Ontological relationships: informal or formal (description logics): this lecture (informal approaches).
- Distributional approaches (lecture 8 and 9).
Approaches to lexical meaning

- Formal semantics: **extension** — what words denote (e.g., cat′: the set of all cats).
- Semantic primitives: e.g., *kill* means CAUSE (NOT (ALIVE)).
- Meaning postulates:

  \[ \forall e, x, y [\text{kill}'(e, x, y) \rightarrow \exists e' [\text{cause}'(e, x, e') \land \text{die}'(e', y)]] \]

- Ontological relationships: informal or formal (description logics): this lecture (informal approaches).
- Distributional approaches (lecture 8 and 9).
Is this object a table?
Other examples to think about

- tomato
- thought
- democracy
- push
- sticky
Hyponymy: IS-A

- (a sense of) *dog* is a hyponym of (a sense of) *animal*
- *animal* is a hypernym of *dog*
- hyponymy relationships form a taxonomy
- works best for concrete nouns
Some issues concerning hyponymy

- not useful for all words: *thought*, *democracy*, *push*, *sticky*?
- individuation differences: is *table* a hyponym of *furniture*?
- multiple inheritance: e.g., is *coin* a hyponym of both *metal* and *money*?
- what does the top of the hierarchy look like?
Other semantic relations

Classical relations:

Meronomy: PART-OF  e.g., arm is a meronym of body, steering wheel is a meronym of car (piece vs part)

Synonymy  e.g., aubergine/eggplant.

Antonymy  e.g., big/little

Also:

Near-synonymy/similarity  e.g., exciting/thrilling
  e.g., slim/slender/thin/skinny
WordNet

- [http://wordnetweb.princeton.edu/perl/webwn](http://wordnetweb.princeton.edu/perl/webwn)
- large scale, open source resource for English
- hand-constructed
- wordnets being built for other languages
- organized into synsets: synonym sets (near-synonyms)

Overview of adj red:

- **S**: (adj) **red**, reddish, ruddy, blood-red, carmine, cerise, cherry, cherry-red, crimson, ruby, ruby-red, scarlet (of a color at the end of the color spectrum (next to orange); resembling the color of blood or cherries or tomatoes or rubies)
- **S**: (adj) **crimson**, red, violent (characterized by violence or bloodshed) "writes of crimson deeds and barbaric days"– Andrea Parke; "fann'd by Conquest's crimson wing"– Thomas Gray; "convulsed with red rage"– Hudson Strode
- **S**: (adj) crimson, red, reddened, red-faced, flushed ((especially of the face) reddened or suffused with or as if with blood from emotion or exertion) "crimson with fury"; "turned red from exertion"; "with puffy reddened eyes"; "red-faced and violent"; “flushed (or crimson) with embarrassment"
Hyponymy in WordNet

Sense 6
big cat, cat
  => leopard, Panthera pardus
    => leopardedess
    => panther
  => snow leopard, ounce, Panthera uncia
  => jaguar, panther, Panthera onca, Felis onca
  => lion, king of beasts, Panthera leo
    => lioness
    => lionet
  => tiger, Panthera tigris
    => Bengal tiger
    => tigress
Using hyponymy

- Semantic classification: e.g., for named entity recognition. e.g., *JJ Thomson Avenue* is a place.
- RTE style inference: *find/discover*
- Query expansion in search
Collocation

- two or more words that occur together more often than expected by chance (informal description — there are others)
- some collocations are multiword expressions (MWE): striped bass
- non-MWEs: heavy snow
Polysemy

- homonymy: unrelated word senses. *bank* (raised land) vs *bank* (financial institution)
- polysemy: related but distinct senses. *bank* (financial institution) vs *bank* (in a casino)
- *bank* (N) (raised land) vs *bank* (V) (to create some raised land): regular polysemy. Compare *pile*, *heap* etc
- In WN, homonyms and polysemous word forms are therefore associated with multiple (different) synsets.

No clearcut distinctions.
Dictionaries are not consistent.
WN example – “interest”

Noun

- **S (n) interest, involvement** (a sense of concern with and curiosity about someone or something) “an interest in music”
- **S (n) sake, interest** (a reason for wanting something done) “for your sake”; “died for the sake of his country”; “in the interest of safety”; “in the common interest”
- **S (n) interest, interestingness** (the power of attracting or holding one’s attention (because it is unusual or exciting etc.)) “they said nothing of great interest”; “primary colors can add interest to a room”
- **S (n) interest** (a fixed charge for borrowing money; usually a percentage of the amount borrowed) “how much interest do you pay on your mortgage?”
- **S (n) interest, stake** (law) a right or legal share of something; a financial involvement with something) “they have interests all over the world”; “a stake in the company’s future”
- **S (n) interest, interest group** (usually plural) a social group whose members control some field of activity and who have common aims) “the iron interests stepped up production”
- **S (n) pastime, interest, pursuit** (a diversion that occupies one’s time and thoughts (usually pleasantly)) “sailing is her favorite pastime”; “his main pastime is gambling”; “he counts reading among his interests”; “they criticized the boy for his limited pursuits”

Verb:

- **S (v) interest** (excite the curiosity of; engage the interest of)
- **S (v) concern, interest, occupy, worry** (be on the mind of) “I worry about the second Germanic consonant shift”
- **S (v) matter to, interest** (be of importance or consequence) “This matters to me!”
“interest/4” – a closer look

**S: (n) interest** (a fixed charge for borrowing money; usually a percentage of the amount borrowed) “how much interest do you pay on your mortgage?”

**direct hyponym / full hyponym**

- **S: (n) compound interest** (interest calculated on both the principal and the accrued interest)
- **S: (n) simple interest** (interest paid on the principal alone)

**direct hyponym/ inherited hypernym / sister term:**

- **S: (n) fixed charge, fixed cost, fixed costs** (a periodic charge that does not vary with business volume (as insurance or rent or mortgage payments etc.))
- **S: (n) charge** (the price charged for some article or service) "the admission charge"
- **S: (n) cost** (the total spent for goods or services including money and time and labor)
  - **S: (n) outgo, spending, expenditure, outlay** (money paid out; an amount spent)
  - **S: (n) transferred property, transferred possession** (a possession whose ownership changes or lapses)
- **S: (n) possession** (anything owned or possessed)
- **S: (n) relation** (an abstraction belonging to or characteristic of two entities or parts together)
  - **S: (n) abstraction, abstract entity** (a general concept formed by extracting common features from specific examples)
- **S: (n) entity** (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
“interest/5” – a closer look

S: (n) interest, stake ((law) a right or legal share of something; a financial involvement with something) "they have interests all over the world"; "a stake in the company’s future"

direct hyponym/ inherited hypernym / sister term:
  • S: (n) share, portion, part, percentage (assets belonging to or due to or contributed by an individual person or group) “he wanted his share in cash"
  • S: (n) assets (anything of material value or usefulness that is owned by a person or company)
  • S: (n) possession (anything owned or possessed)
  • S: (n) relation (an abstraction belonging to or characteristic of two entities or parts together)
  • S: (n) abstraction, abstract entity (a general concept formed by extracting common features from specific examples)
  • S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
interest/4 and interest/5

entity

abstraction, abstract entity

relation

possession

transferred property, transferred possession

assets

outgo, spending, expenditure, outlay

share, portion, part, percentage

cost

charge

security interest

grubstake

controlling interest

fixed charge, fixed cost, fixed costs

fee
due

interest/4

cover charge, cover

compound interest

simple interest
Interest – all senses
Word sense disambiguation

Needed for many applications, problematic for large domains. Assumes that we have a standard set of word senses (e.g., WordNet)

- frequency: e.g., *diet*: the food sense (or senses) is much more frequent than the parliament sense (Diet of Wurms)

- collocations: e.g. *striped bass* (the fish) vs *bass guitar*: syntactically related or in a window of words (latter sometimes called ‘cooccurrence’). Generally ‘one sense per collocation’.

- selectional restrictions/preferences (e.g., *Kim eats bass*, must refer to fish)
WSD techniques

- supervised learning: cf. POS tagging from lecture 3. But sense-tagged corpora are difficult to construct, algorithms need far more data than POS tagging
- unsupervised learning (see below)
- Machine readable dictionaries (MRDs): e.g., look at overlap with words in definitions and example sentences
- selectional preferences: don’t work very well by themselves, useful in combination with other techniques
Standalone WSD

Once a very common research topic, now less studied:

- Evaluation issues
- Lack of a good standard
- Not application-independent:
  - Speech synthesis: e.g., *bass* Homonyms are not always homophones, but mostly are.
  - SMT and similar applications: WSD part of the model. Translation differences don’t necessarily correspond to source language ambiguity.
Grounding

- meaning isn’t (just) about symbols: humans need to recognize and manipulate things in the world.
- ‘grounding’: relate symbols to the real world (often associated with Harnad, but other authors too).
- is grounding an essential part of meaning?
- preliminary/abstract discussion here — more concrete in later lectures.
Turing: ‘Computing machinery and Intelligence’

- introduces the ‘Turing Test’ to replace the question ‘Can machines think?’
- ‘The Imitation Game’: a man (A), a woman (B) and an interrogator (C).
- Questions put to both A and B: both pretend to be a woman. C must decide.
- Replace A with machine, B remains human, how often will C get the identification wrong (after 5 minutes)?

(Picture adapted from Saygin, 2000)
Intelligence as ungrounded imitation?

- Turing described an abstract test (avoiding the complications of robotics, vision etc).
- But communication is central.
- Deception is key to the test: computer ‘pretends’ to be human.
- Many have argued that the point is not deception per se, but application of intelligence in tricking a human. The woman acts as a neutral control.
- Searle ‘Chinese Room’: discussion of consciousness, criticism of Strong AI.
Lexical meaning: what doesn’t work

- meaning of *tomato* is tomato’ or TOMATO
- meaning postulates
- dictionary definition
  *tomato*: mildly acid red or yellow pulpy fruit eaten as a vegetable
  good dictionary definition allows reader with some familiarity with a concept to identify it
Lexical meaning: unanswered questions

- how far does distributional semantics (next lecture) get us?
- grounding often claimed for systems combining vision and language: is this enough?
- are virtual worlds a possible basis for grounding?
- or do we really need robots?