

**IF CORN
FLOWER IS
MADE OUT OF CORN**



**DOES THAT MEAN
BABY POWDER IS
MADE OUT OF BABIES**



**BUT IF FACE
MASKS ARE
MADE OUT OF FACES**



**WHY DO WE
PUT THEM ON OUR
FACES IF WE
ALREADY HAVE A FACE**



Lexical Semantics

CSC485/2501

Lecture 7

Announcement

- Essay 1 scores are out on Quercus!
- A1 Tutorial 3 on Wednesday, Sep 25.

Lexical Semantics

- Word meanings and their internal structure.
- The structure of the relations among words and meanings.

Current CL Research

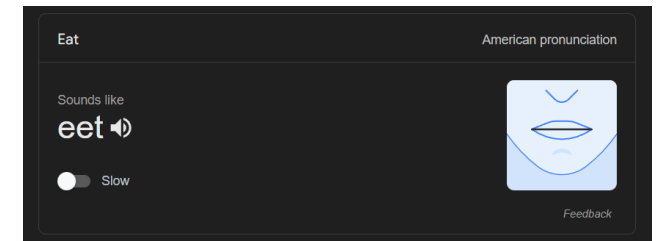
- Current focus in CL on lexical semantics:
 - word senses;
 - detailed lexical representations/vectors;
 - organization of senses, or lexical entries more generally (like a dictionary entry? Probably not).

Knowledge about Words

Lexicon with entry for each word (or fixed phrase).

- Senses (meanings). For each:
 - Surface form:
 - Orthography (spelling), phonology (pronunciation), ...
 - Syntax:
 - Part-of-speech, morphology, subcategorization, ...
 - Behaviour, usage, ...:
 - Collocations, register, genre, ...

eat



present

verb

eat
eat obj

- Collocations

- heavy rain
- high temperature
- thick rain
- tall temperature

- Register

- Variety of language used for a particular situations.



Sorry

I sincerely
apologize for my
past acts, which
seem to have upset you.

President Joseph R. Biden, Jr. spoke today with President Xi Jinping of the People's Republic of China (PRC). The call follows the two leaders' meeting in Woodside, California in November 2023. The two leaders held a candid and constructive discussion on a range of bilateral, regional, and global issues, including areas of cooperation and areas of difference. They reviewed and encouraged progress on key issues discussed at the Woodside Summit, including counternarcotics cooperation, ongoing military-to-military communication, talks to address AI-related risks, and continuing efforts on climate change and people-to-people exchanges.

- Genre

ANNUAL REVIEW OF EARTH AND PLANETARY SCIENCES Volume 34, 2006

Review Article

WATER, MELTING, AND THE DEEP EARTH H₂O CYCLE

Marc M. Hirschmann¹

View Affiliations

Vol. 34:629-653 (Volume publication date May 2006) | <https://doi.org/10.1146/annurev.earth.34.031405.125211>

First published as a Review in Advance on January 31, 2006

© Annual Reviews



r/hiking • 1 yr. ago
Peanutbutter_Warrior

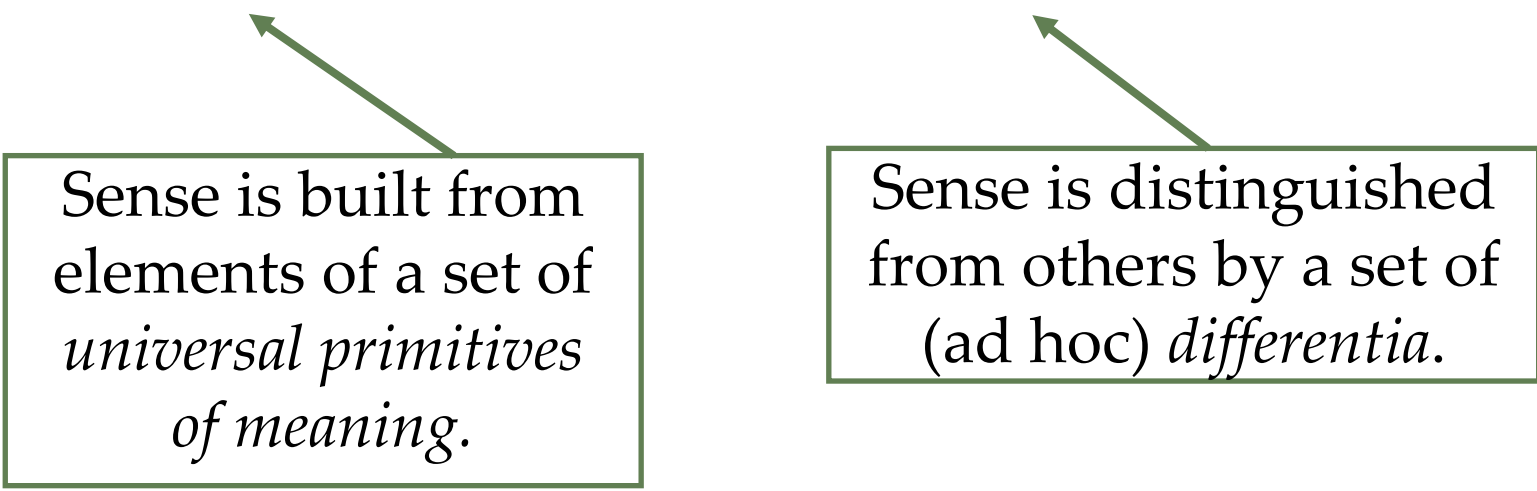
How much water do you carry?

Question

I was on a 3 day hike recently, going for about 6 hours each day. I took 3 litres of water each day assuming it to be enough but the temperature was much higher than I expected and ended up running out of water. It got me wondering how other people decide how much water to carry.

Word Senses

- How are word senses defined?
 - Grounded in world knowledge?
- Are they defined and fixed at all?
 - Or wholly context-dependent? (See also slide 9)
- **Constructional** versus **differential** approaches.



Sense is built from
elements of a set of
universal primitives
of meaning.

Sense is distinguished
from others by a set of
(ad hoc) *differentia*.

Relations between **words** and **senses**

- **Synonymy:** Two (or more) words (synonyms) having the same meaning.



What does this mean?

- Gottlob Frege: sense vs. reference
 - Venus: the morning star is the evening star.
- **Homonymy, polysemy:** Two (or more) meanings having the same word (*homonym, polyseme*).
 - Lexical ambiguity

Lexical Ambiguity: Homonymy

- **Homonymy**: meanings are unrelated.
[Etymology or history of word is not a deciding factor.]
- Due to same spelling (*homography*):
 - *bank* for money, *bank* of river, *bank* of switches,
... *bank* → *banque* or *bord* or *rangée* or ...?
 - *bass*: “bàss” fish, “bāss” guitar;
bow: “bau” to the audience, tie a “bō”.
- Due to same sound (homophony):
 - *wood*, *would*; *weather*, *whether*; *you*, *ewe*, *yew*; *bough*, *bow*.

Lexical ambiguity: Polysemy

- **Polysemy:** meanings are related.
 - *run*: of humans, rivers, buses, bus routes, ...
 - *line*: of people, of type, drawn on paper, transit, route, ...
- Often, no clear line between polysemy and homonymy.

Lexical ambiguity: Polysemy

- Sense modulation by context:
 - fast train, fast typist, fast road.
- Systematic polysemy or sense extension:
 - Arrive
 - to come to locations *arrive at the gate*
 - to come to an event *arrive at a concert*
 - to achieve a goal or cognitive state *arrive at a conclusion*
 - Applies to most or all senses of certain semantic classes.

Relations between senses

- ***Hyponymy, hyperonymy (hypernym)***: subtype, supertype:
 - *sedan* is a hyponym of *car*;
car is a hyperonym of *sedan*.
[*hypo-* = under; *hyper-* = over]
 - The fundamental relation for creating a taxonomy: a tree-like structure that expresses classes and inheritance of properties.
- [Terminology:
 - **is-a** relation in ontologies of (language-independent) concepts;
 - hyponymy relation in taxonomies of (language-dependent) senses.]

Relations between senses

- ***Meronymy, holonymy***: part/whole, or membership:
 - *leg* is a meronym of *chair*;
chair is a holonym of *leg* and a meronym of *dining-set*.
 - Many subtypes of meronym relations.
Component-of: kitchen–apartment
Member-of: soldier–army
Portion-of: slice–pie

Relations between senses

- ***Entailment, implicature:*** various kinds:
 - *snore* entails *sleep*;
manage implies *try*.

Lexical acquisition

- *Problem*: Need a complete lexicon for each natural language.
- Dictionary as starting point?
- Text (corpus) as starting point?
- Build by hand (*lexicographers*) or automatically?

Lexical acquisition

- Corpus-based pattern recognition methods.
 - Accurate, representative information.
 - Includes statistical information.
- Extraction from online dictionary.
 - More knowledge-based.
 - Can treat dictionary as highly specialized corpus.

WordNet

- **WordNet:** A hierarchical (taxonomic) lexicon and thesaurus of English.
 - Developed by lexicographers at Princeton, 1990s to present.
- Graph structure
 - Nodes are **synsets** (“synonym sets”) (\approx word senses).
 - `wn.synsets('dog')`
- Assignment 2 question 0: play with wordnet

<http://wordnetweb.princeton.edu/perl/webwn>

Noun

Synonyms of this sense

Gloss

- S: (n) faux pas, gaffe, solecism, **slip**, gaucherie (a socially awkward or tactless act)
- S: (n) **slip**, slip-up, miscue, parapraxis (a minor inadvertent mistake usually observed in speech or writing or in small accidents or memory lapses etc.)
- S: (n) **slip** (potter's clay that is thinned and used for coating or decorating ceramics)
- S: (n) cutting, **slip** (a part (sometimes a root or leaf or bud) removed from a plant to propagate a new plant through rooting or grafting)
- S: (n) **slip** (a young and slender person) *"he's a mere slip of a lad"*
- S: (n) mooring, moorage, berth, **slip** (a place where a craft can be made fast)
- S: (n) **slip**, trip (an accidental misstep threatening (or causing) a fall) *"he blamed his slip on the ice"; "the jolt caused many slips and a few spills"*
- S: (n) slickness, slick, slipperiness, **slip** (a slippery smoothness) *"he could feel the slickness of the tiller"*
- S: (n) strip, **slip** (artifact consisting of a narrow flat piece of material)
- S: (n) **slip**, slip of paper (a small sheet of paper) *"a receipt slip"; "a withdrawal slip"*
- S: (n) chemise, shimmy, shift, **slip**, teddy (a woman's sleeveless undergarment)
- S: (n) case, pillowcase, **slip**, pillow slip (bed linen consisting of a cover for a pillow) *"the burglar carried his loot in a pillowcase"*
- S: (n) skid, **slip**, sideslip (an unexpected slide)
- S: (n) **slip**, sideslip (a flight maneuver; aircraft slides sideways in the air)
- S: (n) **slip**, elusion, eluding (the act of avoiding capture (especially by cunning))

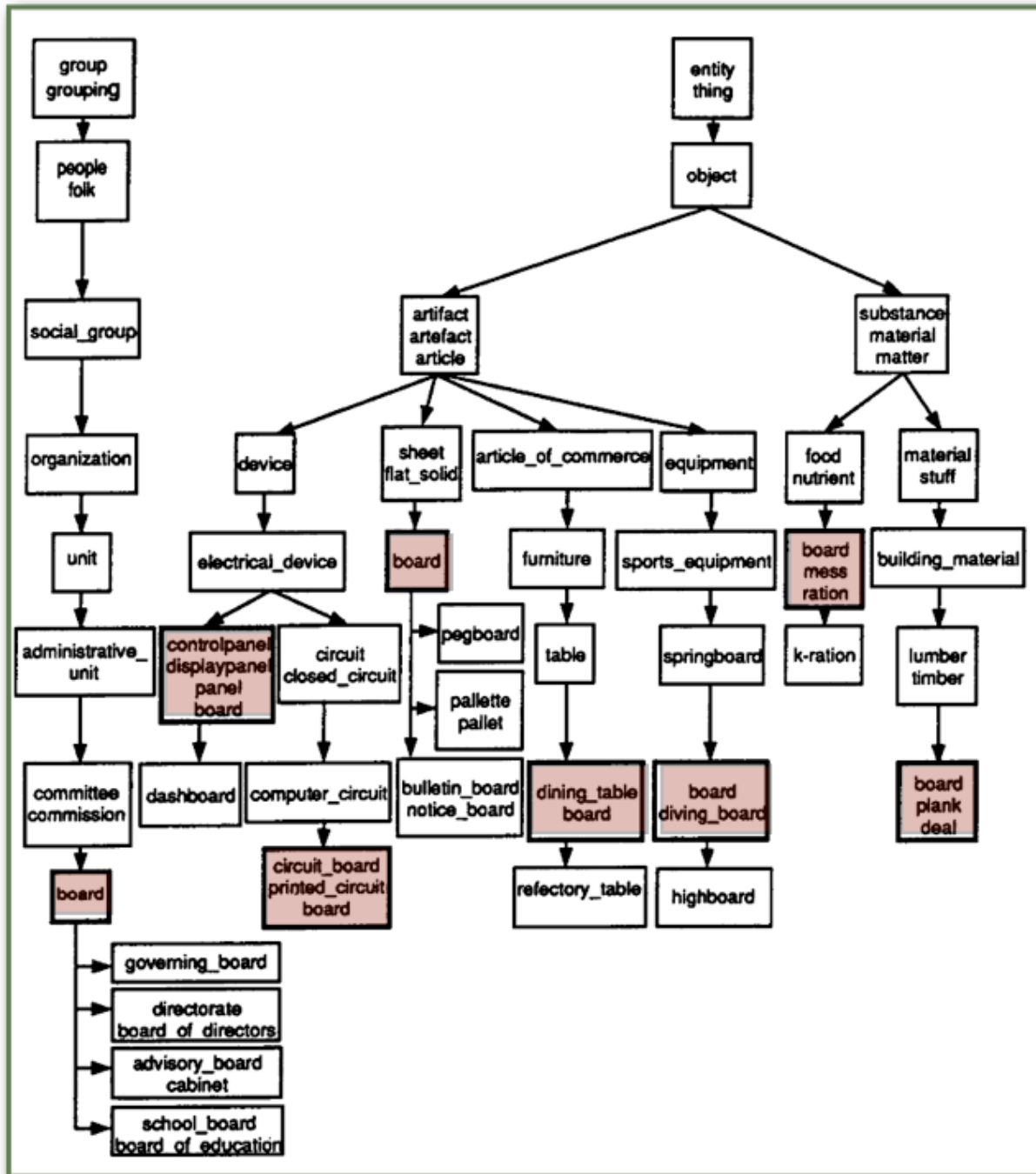
Example

Wordnet hyperonyms

- [S: \(n\) slip](#), [slip of paper](#) (a small sheet of paper) *"a receipt slip"; "a withdrawal slip"*
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - [S: \(n\) sheet](#), [piece of paper](#), [sheet of paper](#) (paper used for writing or printing)
 - [S: \(n\) paper](#) (a material made of cellulose pulp derived mainly from wood or rags or certain grasses)
 - [S: \(n\) material](#), [stuff](#) (the tangible substance that goes into the makeup of a physical object) *"coal is a hard black material"; "wheat is the stuff they use to make bread"*
 - [S: \(n\) substance](#) (the real physical matter of which a person or thing consists) *"DNA is the substance of our genes"*
 - [S: \(n\) matter](#) (that which has mass and occupies space) *"physicists study both the nature of matter and the forces which govern it"*
 - [S: \(n\) physical entity](#) (an entity that has physical existence)
 - [S: \(n\) entity](#) (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))
 - [S: \(n\) part](#), [portion](#), [component part](#), [component](#), [constituent](#) (something determined in relation to something that includes it) *"he wanted to feel a part of something bigger than himself"; "I read a portion of the manuscript"; "the smaller component is hard to reach"; "the animal constituent of plankton"*
 - [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))

Sister terms belong to synsets

- S: (n) **slip**, slip of paper (a small sheet of paper) *"a receipt slip"; "a withdrawal slip"*
 - direct hypernym / inherited hypernym / **sister term**
 - S: (n) sheet, piece of paper, sheet of paper (paper used for writing or printing)
 - S: (n) **slip**, slip of paper (a small sheet of paper) *"a receipt slip"; "a withdrawal slip"*
 - S: (n) signature (a sheet with several pages printed on it; it folds to page size and is bound with other signatures to form a book)
 - S: (n) leaf, folio (a sheet of any written or printed material (especially in a manuscript or book))
 - S: (n) tear sheet (a sheet that can be easily torn out of a publication)
 - S: (n) foolscap (a size of paper used especially in Britain)
 - S: (n) style sheet (a sheet summarizing the editorial conventions to be followed in preparing text for publication)
 - S: (n) worksheet (a sheet of paper with multiple columns; used by an accountant to assemble figures for financial statements)
 - S: (n) revenue stamp, stamp (a small piece of adhesive paper that is put on an object to show that a government tax has been paid)



Eight senses of *board* in WordNet, and their hyperonyms and hyponyms.

WordNet

- Graph structure (cont.):
 - Edges from hyponymy relations: near-tree.
 - Edges from meronymy relations: network.
- Index maps each word to all of its synsets.
- Separate trees for nouns, verbs, adjectives, adverbs (with derivational cross-connections).
- Differential approach to meaning:
 - The hyponyms of a node are differentiations of its meaning.

WordNet

- WordNets now available or under construction for many languages.

Afrikaans, Albanian, Arabic, Bantu, Basque, Bengali, Bulgarian, Catalan, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, Farsi (Persian), Finnish, French, German, Greek, Hebrew, Hindi, Hungarian, Icelandic, Indonesian, Italian, Irish, Japanese, Kannada, Korean, Latin, Latvian, Macedonian, Maltese, Marathi, Moldavian, Mongolian, Myanmar, Nepali, Norwegian, Oriya, Polish, Portuguese, Romanian, Russian, Sanskrit, Serbian, Slovenian, Spanish, Swedish, Tamil, Thai, Turkish, Vietnamese...

- <http://globalwordnet.org/resources/wordnets-in-the-world/>

Building and updating WordNets

Problem: Need a complete lexicon and lexical relations for each natural language.

Hearst's Algorithm: Discovering lexical relations

- Corpus-based method.
- Makes “suggestions” for lexicographers.
- Scan partially-parsed text looking for instances of patterns:
“*such* NP1 *as* {NPi}* {*or* | *and*} NPi”
→ NP1 is a **hyperonym** of the NPi

Hearst's Results (good)

- Some relations already in WordNet:
 - *fabric–silk, grain–barley, disorders–epilepsy, ...*
- Some relations not already in WordNet (but the words were):
 - *crops–milo, perishables–fruit, conditions–epilepsy, ...*
- Some relations with words not yet in WordNet:
 - *companies–Shell, institutions–Tufts, ...*

Hearst's Results (less good)

- Some too-general relations:
 - *things–exercise, topics–nutrition, areas–Sacramento*
- Some too-context-specific relations:
 - *others–Meadowbrook, classics–Gaslight, categories–drama, ...*
- Some really bad relations (usually due to parsing errors, not detecting full NP):
 - *children–Headstart, jobs–computer, companies–sports*

Hearst: Limitations

- Problems:

- Which word is the hyperonym?

A bearing is a structure that supports a rotating part of a machine, such as a shaft, axle, spindle, or wheel.

```
graph LR; A["A bearing"] -- is --> B["a structure"]; B -- supports --> C["a rotating part"]; C -- part of --> D["a machine"]; E["a shaft"] -- part of --> D; F["axle"] -- part of --> D; G["spindle"] -- part of --> D; H["wheel"] -- part of --> D;
```

- Can't find good patterns for meronyms.
 - How to evaluate method quantitatively?

Since Hearst's Paper

- Methods that use syntactic (not just lexical) patterns, and which derive the patterns from corpora.
- Methods that use senses, not words.
- Methods for finding coordinate (sister) terms by distributional similarity in text.
- Methods that combine the evidence from all of these to identify additional hyponym relations.
 - $\text{sister}(X,Y) \wedge \text{hyponym}(Y,Z) \Rightarrow \text{hyponym}(X,Z)$

Since Hearst's Paper

- Methods for meronymous relations.
 - Each subtype tends to have its own indicators.
 - These tend to have much more ambiguous patterns than hyponymy.
 - Complex methods for learning additional semantic constraints on the patterns.
- Methods for causal relations.
 - Look esp. for verbs such as give rise to, induce, generate, cause, ...

Quiz

Match the terms with the definitions!

- Hyperonymy
 - Homonymy
 - Polysemy
 - Holonymy
 - Hyponymy
 - Meronymy
 - Synonymy
- A relation to a more generic word (subtype).
[sedan is a TERM of car]
 - A relation to a specific instance (supertype).
[car is a TERM of sedan]
 - Two (or more) unrelated meanings having the same word.
 - Two (or more) related meanings having the same word.
 - Two (or more) words having the same meaning.
 - A relation of membership from part to whole.
[leg is a TERM of chair]
 - A relation of membership from whole to part.
[chair is a TERM of leg]