

L113 Word Meaning and Discourse Understanding

Session 1: Background to Lexical Semantics and Word Senses

Simone Teufel

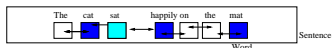
MPhil in Advanced Computer Science
Computer Laboratory Natural Language and Information Processing (NLP)
Group



Simone.Teufel@cl.cam.ac.uk

2012/2013

Units of Linguistic Meaning



[L100 Introduction to NLP](#) – how to assemble meaning from individual words inside a sentence (compositional semantics).

[L113 Word Meaning and Discourse Understanding](#) covers Lexical Semantics, which looks at meanings of *individual* words. In the standard NLP pipeline, the semantics of words is left “atomic”/not treated.

What does a word *mean*?

burn

Different Kinds of Ambiguity

- Purely syntactic:
 - *young women and men*
- Quasi-syntactic:
 - *a red pencil*
 - *the astronaut entered the atmosphere again*
- Lexico/syntactic:
 - *we saw her duck*
- Purely lexical:
 - *He reached the bank*

What can we do with Word Meaning?

- Describe and examine word senses

Example:

- We unhinged the door.*
- We walked through the door.*
- ?We unhinged the door and walked through it.*

door, window = $\begin{cases} \text{opening in wall} \\ \text{cover for opening} \end{cases}$

- Automatically recognise word senses in text
- Recognise and interpret figurative use of words
- Define similarities between words
- Determine how strongly a verb “goes with” its subject
- Describe relations between words (or rather, between word senses)

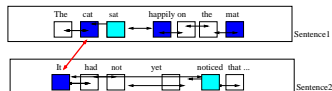


X-proofing

acid-proof, affair-proof, air-proof, ant-proof, baby-proof, bat-proof, bear-proof, bite-proof, bomb-proof, bullet-proof, burglar-proof, cat-proof, cannon-proof, claw-proof, coyote-proof, crash-proof, crush-proof, deer-proof, disaster-proof, dust-proof, dog-proof, elephant-proof, escape-proof, explosion-proof, fade-proof, fire-proof, flame-proof, flood-proof, fool-proof, fox-proof, frost-proof, fume-proof, gas-proof, germ-proof, glare-proof, goof-proof, gorilla-proof, grease-proof, hail-proof, heat-proof, high-proof (110-proof, 80-proof), hurricane-proof, ice-proof, idiot-proof, jam-proof, leak-proof, leopard-proof, lice-proof, light-proof, mole-proof, moth-proof, mouse-proof, nematode-proof, noise-proof, oil-proof, oven-proof, pet-proof, pilfer-proof, porcupine-proof, possum-proof, puncture-proof, quake-proof, rabbit-proof, raccoon-proof, radiation-proof, rain-proof, rat-proof, rattle-proof, recession-proof, rip-proof, roach-proof, rub-proof, rust-proof, sand-proof, scatter-proof, scratch-proof, shark-proof, shatter-proof, shell-proof, shock-proof, shot-proof, skid-proof, slash-proof, sleet-proof, slip-proof, smear-proof, smell-proof, smudge-proof, snag-proof, snail-proof, snake-proof, snow-proof, sound-proof, stain-proof, steam-proof, sun-proof, tamper-proof, tear-proof, teenager-proof, tick-proof, tornado-proof, trample-proof, varmint-proof, veto-proof, vibration-proof, water-proof, weasel-proof, weather-proof, wind-proof, wolf-proof, wrinkle-proof, x-ray-proof, zap-proof

source: www.wordnik.com/lists/heres-your-proof

Units of Linguistic Meaning



The second part of L113 covers **Discourse Understanding**, which looks at effects of meaning **across** sentences, not within one sentence.

Discourse Understanding

Discourse Phenomena, Def: Any linguistic phenomena of cohesion ("belonging together") that acts across sentence boundaries.

Example: interpretation of anaphora

- If your baby cries, give it some hot milk.
 - If *it* does not stop crying, try...
 - If *it* boils too quickly, put some bicarbonate ...
 - If *it* is inconvenient to buy fresh milk ...

End of Overview of L113

Beginning of "Word Meaning" part of L113

Judging "normality" and semantic differences

Diagnostic Methodology: Construct a linguistic context for a word, then judge normality/truthconditional status of that utterance.
(Does it sound odd? Could it be true?)

- *a female mother.*
BUT: *a female pianist.*
- *Kate was very married.*
BUT: *Kate was married.*
- *The kitten drank a bottle of claret.*
BUT: *The undergraduate drank a bottle of claret.*
- *Arthur and his driving licence expired last Thursday*
BUT: *Arthur's passport and his driving licence expired last Thursday*

The coloured sentences are odd (for different reasons); the black ones are not. Why?

Aspects of semantic oddness

- **Pleonasm:** Tautologies; redundant information
 - *a female mother*
- **Dissonance:** "Selectional restrictions" are violated.
 - *Kate was very married*
- **Improbability:** The truthconditional conditions of the utterance are untrue/unlikely in most possible worlds, but one can imagine the situation under special circumstances.
 - *The kitten drank a bottle of claret*
- **Zeugma:** Two senses of a word are activated simultaneously.
 - *Arthur and his driving licence expired last Thursday*

Over to you (solutions in textbook)

- *He was wearing a scarf, a pair of boots, and a look of considerable embarrassment.*
- *Let us drink time.*
- *He was murdered illegally.*
- *Kick it with one of your feet.*
- *The throne is occupied by a chain-smoking alligator.*
- *They took the door off its hinges and walked through it.*
- *We smashed the window then climbed through it.*

But careful...

Do the following sentences mean the same thing?

- *He watched it with intense concentration for a few moments, then left the room.*
- *He looked at it with intense concentration for a few moments, then left the room.*
- Intuitions need to be disciplined.
- Reliability and accuracy.

Now we need some terminology

- **Sentence:** a linguistic object only; no truthconditional content.
 - *John saw Mary.*
 - *John's sister was seen by Peter's uncle.*
- **Proposition:** fact in the world
 - only two kinds:
 - relation between two or more entities
 - attribution of property to an entity
 - has truthconditional content (once instantiated)
 - is independent of linguistic form
- **Statement:**
 - Sentence + Proposition + Situation (reference) → linguistic form and truth-conditional content

Entailment

A proposition P is said to **entail** another proposition Q if the truth of Q is a logically necessary consequence of the truth of P (and the falsity of P is a necessary consequence of the falsity of Q).

- Sentence P “*That is a dog*” entails sentence Q “*That is an animal*”.
- This means in every situation where I can say P I can also say Q:
 - *It can't possibly be a dog and not an animal.*
 - *It's a dog therefore it's an animal.*
 - *If it is not an animal, then it follows that it's not a dog.*
 - *? It's a dog, so it must be a cat.*
 - *? It's not an animal, but it's just possible that it's a dog.*
 - *? It's a dog, so it might be an animal*

Types of entailment

Unilateral entailment:

- *It's a dog* unilaterally entails *It's an animal*

Mutual entailment (logical equivalence):

- *The meeting began at 10am* entails and is entailed by *The meeting commenced at 10am*.

Contrariety:

- *It's a cat* entails *It's not a dog*.
- *It's a not a cat* does not entail *It's a dog*.

Contradiction:

- *It's dead* entails *It's not alive*.
- *It's alive* entails *It's not dead*.

Distributional definition of word meaning

Assumption: **The meaning of a word is fully reflected in its contextual relations.**

Words form two kinds of affinities:

- **Syntagmatic**: semantic associations between items **within** a sentence:

- *dog ... barked ...*

"semantic traits" of a target word.

- **Paradigmatic**: semantic affinities between two grammatically identical words which can **replace** each other in a sentence:

- *I haven't yet fed the [dog/cat/*lamppost].*

The meaning of a word is only defined by these two sets: its syntagmatic and its paradigmatic affinities.

Semantic Traits (Syntagmatic)

Semantic traits are syntagmatic properties attributed to an entity and can be of different types:

- **critical**: *animal-dog*
 - *It's a dog* logically entails *It's an animal*.
If it's a dog, it's necessarily an animal.
- **expected**: *bark-dog*:
 - *It's a dog, but it can't bark*.
 - *? It's a dog, but it can bark*.
- **possible**: *brown-dog*
 - *? It's a dog, but it is brown*.
 - *? It's a dog, but it isn't brown*.
- **unexpected**: *can sing-dog*
 - *It's a dog, but it can sing*.
 - *? It's a dog, but it can't sing*.

Semantic Traits, Ctd.

- **excluded**: *cat-dog*
 - *It's a dog* logically entails *It's not a cat*.
- **canonical**: *has four legs-dog*
 - A canonical trait is an expected trait whose absence is regarded as a defect.
 - *? The typical dog has 4 legs*.
 - *? A dog that does not have 4 legs is not necessarily defective*.
 - *The typical bird is adapted for flight*.

Syntagmatic and paradigmatic affinities highlight different aspects of similarity:

cat and *dog* have a high degree of paradigmatic affinity, but syntagmatically, they are excluded traits of each other.

Comparing Semantic Contrast

Does an affix, a word or a set of words form a "minimal semantic unit"?

- $I \left\{ \begin{smallmatrix} dis \\ - \end{smallmatrix} \right\}$ like him.
- $I \left\{ \begin{smallmatrix} dis \\ - \end{smallmatrix} \right\}$ approve of the idea.
- $We \left\{ \begin{smallmatrix} dis \\ - \end{smallmatrix} \right\}$ appointed her.
- $You \text{ must } \left\{ \begin{smallmatrix} dis \\ - \end{smallmatrix} \right\}$ embark here.

Which semantic contrasts are equivalent?

Word Senses

- Why do some word forms have more than one sense?
 - Random historic effects → homonymy
 - Senses evolve and are connected → polysemy
- How can we find out whether a word has more than one sense, or whether it is underspecified?
 - *child* could be a boy or a girl
 - *school* could be a building or a logical institution
- Several linguistic tests will be introduced after the break.

Recurrent Contrast Test

Construct two very different contexts in which the item can occur and can be replaced by a second item (all other words must be different).

The definition of a semantic constituent is that it appears in at least 2 such contexts, and creates a comparable semantic difference.

$John \left\{ \begin{smallmatrix} in- \\ ex- \end{smallmatrix} \right\} \text{haled.} = \text{They } \left\{ \begin{smallmatrix} im- \\ ex- \end{smallmatrix} \right\} \text{port textiles.}$

→ *im-* is a semantic constituent in *import* and *inhale*.

$His \text{ remarks are } \left\{ \begin{smallmatrix} im- \\ - \end{smallmatrix} \right\} \text{pertinent.} \neq \text{What you suggest is } \left\{ \begin{smallmatrix} im- \\ - \end{smallmatrix} \right\} \text{possible.}$

→ *im-* in *impertinent* is not a semantic constituent (more checks necessary).

"SHOWER" has more than one word sense

- Handout: corpora examples for "shower".
- After break – please tell me how many senses "shower" has.
- Please work in groups.
- Please disregard "shower" if it occurs in a compound noun (together with another noun, e.g., "shower curtain" or "hail shower")
- **Break!**

Underspecification vs. Ambiguity

Underspecification:

- Sue visited her cousin.

cousin is underspecified wrt [male/female]. Which interpretation applies is (sometimes) **inferred** from the context:

- Sue's cousin is pregnant.

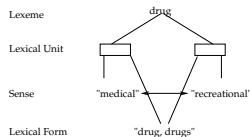
Ambiguity:

- We finally reached the bank.

bank has two distinct senses, with no general meaning covering both. Which sense applies is **sense selected** from the context:

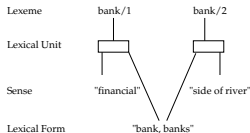
- The bank is steep and covered with brambles.

Polysemy (Word Senses)



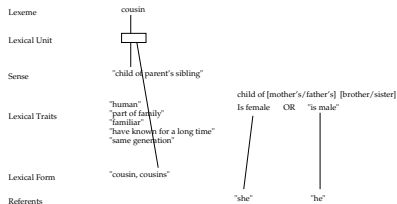
A lexeme which has a number of senses is **polysemous**.

Homonymy (Word Senses)



A lexical form is **homonymous** if it realises lexical units belonging to more than one lexeme.

Underspecification



An **underspecified** lexical form has only one sense, but a (single) semantic trait that is left open/underspecified.

Crossed Interpretations Test for Ambiguity

If two items of the potentially ambiguous word form occur in a sentence, and if it is possible that two contrary interpretations hold, then the word form is underspecified, not ambiguous.

- *Mary has adopted a child; so has Sue.*
 - **YES** – one can be a boy, the other a girl → underspecification ("child" does NOT have two senses)
- *Tom has reached the bank; so has Joe.*
 - **NO** – This can only mean that both reached the same type of bank. → ambiguity ("bank" has two senses)
- *Tom wants to know if this is a dog; so does Joe.*
 - **NO** – This can only mean that both enquire either about the breed or the sex of the dog. → ambiguity

Yes/No Test for Ambiguity

- For an ambiguous word form, you can construct a single situation including both senses of the word form, where one of the senses is true, and the other false.
- Show that you have done so, by giving a question containing the word form which can be answered both **yes** and **no**, depending on the sense the speaker has in mind.
- You cannot do this for an underspecified word form.

Zeugma Test for Ambiguity

Ambiguous word forms give rise to zeugma, if more than one sense of an ambiguous word form is activated.

- ? *John and his driving licence expired last Thursday.*

Underspecified word forms don't give rise to zeugma:

- *My cousin, who is pregnant, was born on the same day as Arthur's, who is the father.*

Yes/No Test for Ambiguity

- *Is that a dog?* [species yes, male dog, no]
 - Yes, it's a *Spaniel*.
 - No, it's a *bitch*.
- *Did Arthur make it to the bank?* [riverbank yes, money bank no]
 - Yes, he's a *strong swimmer*.
 - No, he was *arrested as soon as he came out of the water*.

Yes/no Test fails for Underspecified item

In contrast:

- *Is the subject of this poem a monarch?* [queen yes, king no]
 - Yes, it's a queen.
 - ? No, it's a king.

Indirect Tests for Ambiguity

Word form X is ambiguous if it stands in relation Y with other word forms Z_1 and Z_2 in one occurrence context but not another (and the two contexts exemplify different senses).

Y=Synonymy

- *Guy struck the match.* – lucifer
- *The match was a draw.* – contest

Y=Antonymy

- *The room was painted in light colours.* – dark
- *Arthur has a light teaching load.* – heavy

Y=Paronymy

- *She complained about discrimination by race.* – racist
- *The race was won by Arthur* – racing.

A strange phenomenon: Sense Spectra

Zeugma test shows different senses for *mouth*:

- ? *The poisoned chocolate entered the Contessa's mouth at the same instant that the yacht entered that of the river.*

But there is a **sense spectrum** connecting the two:

- 1 *John keeps opening and shutting his mouth like that of a fish.*
- 2 *The parasite attaches itself to the mouths of fishes, sea squirts etc.*
- 3 *The mouth of a sea squirt resembles that of a bottle.*
- 4 *The mouth of a cave resembles that of a bottle.*
- 5 *The mouth of the enormous cave was also that of the underground river.*

We can't do the same with *expire* and *expire*!

Types of Polysemy

- Linear polysemy: two word forms are in a systematic semantic relation to each other, and one of them takes on the meaning of the other:
 - dog/canine–male
 - door/part–whole
 - man/male–humankind
 - wheat/plant–food
- Non-linear polysemy (mainly metaphor):
 - *Has Arthur changed his position?*
 - *The ham sandwich asked for the bill.*

Systematic Polysemy

- Apply to all members of a class by default
- Physical object – content:
 - *I was hit on the head by a novel.*
- Unit – type:
 - *I want that shirt.*
- Species – individual
- Animal – meat
- ...

Word Senses: Example *interest*

- *She pays 3% interest on the loan.*
- *He showed a lot of interest in the painting.*
- *Microsoft purchased a controlling interest in Google.*
- *He said nothing of great interest.*
- *It is in the national interest to invade the Bahamas.*
- *I only have your best interest in mind.*
- *Playing chess is one of my interests.*
- *Business interests lobbied for the legislation.*
- *Primary colours can add interest to a room.*

Multilingual aspect of word sense ambiguity

Example: *interest* translated into German

- Zins: financial charge paid for loan
- Anteil: stake in a company
- Anteilnahme: curiousness
- Interesse: all other senses

Summary of Today

- Proposition, sentence meaning, entailment, semantic traits: [chapter 2](#).
- Normality Judgements and sources of anomaly: [chapter 3](#).
- Semantic traits: [chapter 3](#).
- Recurrent contrast test (page 70ff).
- Sense Ambiguity Tests [chapter 6](#).
- D. A. Cruse, *Meaning in Language*. Oxford Linguistics Press, 2000. Chapter 6 (plus 1-3 + 5 for terminology and core concepts)