Lexical Semantics and Discourse Processing

Lecture 2: Word Senses and Lexical Relations

MPhil in Advanced Computer Science

UNIVERSITY OF CAMBRIDGE

Simone Teufel

Natural Language and Information Processing (NLIP) Group

sht25@cl.cam.ac.uk

Last Time

- Normality Judgements and Linguistic Tests
- Paradigmatic vs. Syntagmatic Affinity
- Aspects of Semantic Infelicity (Pleonasm, Dissonance, Improbability, Zeugma)
- Semantic Traits
- Semantic Constituency

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).
- If you can find at least two contexts where the semantic difference between the two contexts is comparable, then the item is a semantic constituent:

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

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

This means that *im-* is a semantic constituent in *import* but not in *impertinent.*

More on the Recurrent Contrast Test

This does not work for random parts of words:

The cat sat on the m $\left\{ \begin{matrix} \text{-at} \\ \text{-oss} \end{matrix} \right\}$ *. ≠ He does not like his new b* $\left\{ \begin{matrix} \text{-at} \\ \text{-oss} \end{matrix} \right\}$ *.*

Now let's look at compounds again:

I saw a $\left\{ \begin{matrix} black\text{-} \\ blue\text{-} \end{matrix} \right\}$ *bird in the garden. ≠ Cynthia wore* $\left\{ \begin{matrix} black \\ blue \end{matrix} \right\}$ *stockings.*

and *blue* in *bluebird* and *blackbird* are not semantic constituents.

Modulation vs. Sense selection

- *Arthur poured the butter into a dish.*

→ the butter must be liquid.

- *Let me cash the cheque at the bank before we go.*

→ he must be talking about the financial institution.

- In each context, there is some semantic information contained in the lexical item, and some comes from the context

- But: different mechanisms in place: balance of information

- Sense selection: Rich bundles of semantic traits, one is chosen on the basis of the context; context acts as trigger.

- Contextual Modulation: One semantic trait (e.g., solid vs. liquid state) which is not explicitly mentioned is inferred from context

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.*

If a word form is underspecified, then the context totally conditions its interpretation. If a word form is ambiguous, then its senses should not in every case be totally conditioned by their contexts.

Recipe:

- Generate two contexts in which the item has different interpretations.

- Replace item by synonym or hypernym which covers both interpretations.

- If no information loss occurs, then the distinguishing interpretation is entirely the result of contextual modulation.

- Is *monarch* ambiguous between male and female interpretation, or is it underspecified?

- Construct contexts for the two interpretations:

  – *The Ruritanian monarch is expecting her second baby.*
  – *The child's father is the reigning monarch.*

- Replace with synonym (e.g., *crowned head, sovereign*)

- Does it result in information loss? Here, no.

- Ergo: the information that allows us to distinguish between the interpretations is entirely derived by contextual modulation.

- Ergo: *monarch* is not ambiguous, but underspecified.

First Ambiguity Test; ambiguous lexical item

- Now: *bank*
- Construct contexts:
  - *His wife is the manager of the local bank.*
  - *At this point, the bank was covered with brambles.*
- Replace with common hyponym:
  - *His wife is the manager of the local place.*
  - *At this point, the place was covered with brambles.*
- This results in information loss.
- So we have failed to show that the interpretation of *bank* is entirely the result of contextual modulation.
- Ergo: *bank* is ambiguous.

Another example

dog has two senses: canine/male dog

- *John prefers bitches to dogs/?canines.*
- *Arthur breeds dogs.*

The "male dog" sense is not the result of context modulation:

- *Incredibly, John prefers an aged, half-blind bitch to a dog, as his canine companion.*
- *? Mary prefers mares to horses.*

Second Test for Ambiguity: Yes/No Test

If a word form is ambiguous, then both of its senses must be independently maximisable (i.e., the interpretation is forced to cover all possible referents).

Recipe:

- Construct a situation including both interpretations of the word form, where one interpretation is false and the other correct.

- Show that this is so with a question concerning the word form which can be answered both yes and no, depending on the interpretation.

- Then the word form is ambiguous.

Second Test for Ambiguity

- *Is that a dog?*
  - *Yes, it's a Spaniel.*
  - *No, it's a bitch.*

- *Did Arthur make it to the bank?*
  - *Yes, he's a strong swimmer.*
  - *No, he was arrested as soon as he came out of the water.*

Second Test for Ambiguity: Underspecification

In contrast:

- *Is the subject of this poem a monarch?*

    – *Yes, it's a queen.*
    – *? No, it's a king.*

Third Test for Ambiguity: Zeugma Test

Contexts which activate more than one sense of an ambiguous word form give rise to the oddness called zeugma:

- *? 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.*

Fourth Test for Ambiguity: Crossed Interpretations Test

- For underspecified items, we can get a crossed interpretation:

  – *Mary has adopted a child; so has Sue.*

- I.e., there is the possibility that one has adopted a boy and the other a girl.

- Not so for ambiguous items:

  – *Tom has reached the bank; so has Joe.*
  – *Tom wants to know if this is a dog; so does Joe.*

- This can only mean that both reached the same type of bank, and enquire either about the breed or the sex of the dog.

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 discrimiation by race. – racist

The race was won by Arthur – racing.

Systematic Relationships between senses

- Physical object – content:
  - *I was hit on the head by a novel.*
- Unit – type:
  - *I want that shirt.*
- Metaphor:
  - *Has Arthur changed his position?*
- …
- These systematic relationships are sometimes referred to as "lexical rules".

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 enormeous cave was also that of the underground river.*

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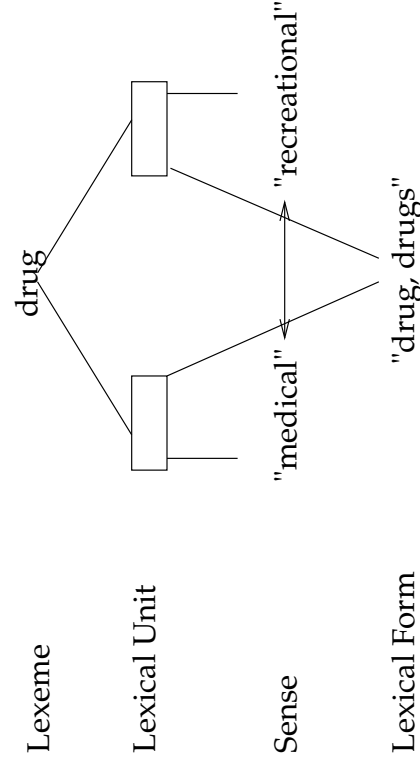
Lexical units and lexemes

Lexical unit: a form-meaning complex with relatively stable and discrete semantic properties which stand in meaning relations such as antonymy (long:short) and hyponomy (dog:animal). The meaning aspect of a lexical unit is called a sense. The form aspect of a lexical unit is called a lexical form.

Lexical form: family of word forms differing only in inflectional morphology. Must be a semantic constituent, i.e., can include multi-word units.
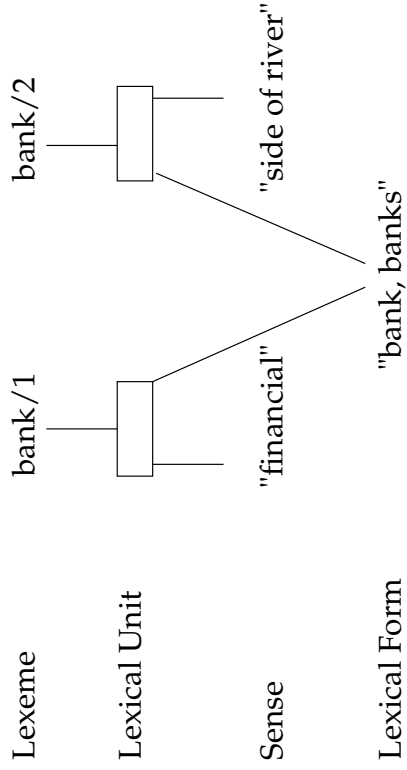
Lexeme: contains one or more lexical units of the same POS, if either

- there exists a lexical rule which permits the existence of the sense of one from the existence of the sense of the other. Recurrent semantic contrast between senses is evidence of a lexical rule (e.g., unit and type readings of pieces of clothing).

- the senses are local senses belonging to a sense spectrum (e.g., *mouth of a river* and *human mouth*)
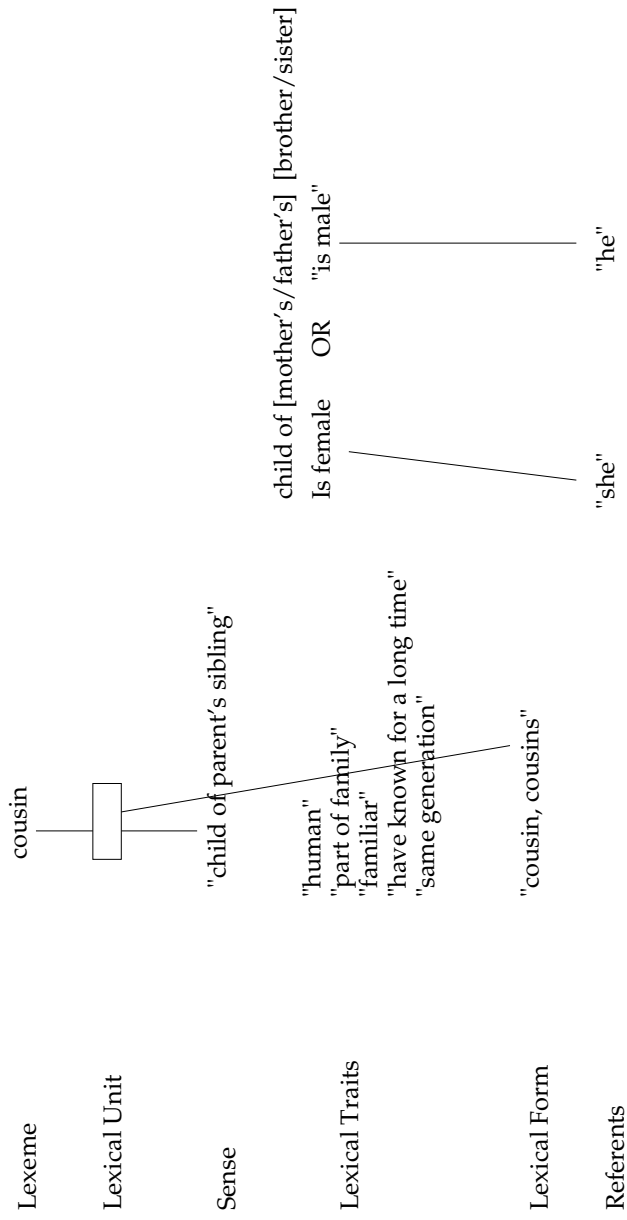
Polysemy

Lexeme

Lexical Unit

Sense

Lexical Form

drug

"medical"   "recreational"

"drug, drugs"

A lexeme which has a number of senses is polysemous.

Homonymy

Lexeme   bank/1   bank/2

Lexical Unit

Sense   "financial"   "side of river"

Lexical Form   "bank, banks"

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

Underspecification

Lexeme   cousin

Lexical Unit

Sense   "child of parent's sibling"   child of [mother's/father's] [brother/sister]

Lexical Traits   "human"   Is female   OR   "is male"
"part of family"
"familiar"
"have known for a long time"
"same generation"

Lexical Form   "cousin, cousins"

Referents   "she"   "he"

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

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.*

Breakout session:

How many clusters are there here, and why?

Lexical Relations: Congruence Relations

- Synonymy (*sofa/couch*): X is a cognitive synonym of Y if X and Y are syntactically identical, and any grammatical declararative sentence S containing X has equivalent truthconditional conditions to another sentence S_1, which is identical to S expect that X is replaced by Y.

- Hyponymy (*dog/animal*): X is a hyponym of Y if there is unilateral entailment $S \Rightarrow S_1$ Condition: S, S_1 are of form "This is X".

- Compatibility (*dog/pet*), (*husband/policeman*):
No systematic entailment relations hold, but a common superordinate exists. Some semantic traits are shared; difference concerns traits which do not clash

Incompatibility (*cat/dog*):

- *It's a X ⇒ It's not a Y*

- Not very interesting: *affix* and *volcano* are incompatibles.

- Normal definition includes the criterion that X and Y must fall under a single superordinate: *cat, dog, lion, elephant, aardvark*, etc.

- Close relationship to contrariness, but relationship not straightforward

- Items in a coordinated list are usually incompatibles:

  – *? I like fruit and bananas.*

Congruence Variants

- X is a congruent of Y:

$$\forall X \forall Y R(X,Y) \cap \forall Y \forall X R(X,Y)$$

- X is a hypo-R of Y; Y is a super-R of X:

$$\forall X \forall Y R(X,Y) \cup \exists Y \nexists X \; \mathcal{R}(X,Y)$$

- X and Y are semi-Rs: some but not all X and some but not all Y stand in relation R(X,Y).

- Y is a quasi-R of X: if X, Y in the right relationship exist, but do not agree in their POS: *Cutlery* is a quasi-hypernym of *knife, fork, spoon.*

- X is a para-R of Y: expectation rather than necessity holds. E.g., *student, bankmanager* are para-incompatibles:

  – *He's a student but he's a bank-manager.*