

Lexical Semantics and Discourse Processing

Lecture 1: Introduction to concepts and problems

MPhil in Advanced Computer Science



Simone Teufel

Natural Language and Information Processing (NLIP) Group

sh.t25@c1.cam.ac.uk

Overview of Course: Lexical Semantics Part

2

- L1: Concepts in Lexical Semantics
- L2: Word Senses, Lexical Relations and Taxonomies
- L3: Supervised Word Sense Disambiguation
- L4: Unsupervised Word Sense Disambiguation
- L5: Sentiment
- L6: Semantic Spaces
- L7: Models of Semantic Similarity
- L8: Verb Classes and Semantic Role Labelling
- L9: Figurative Language

Direct translation: **meaning of words**; more specifically: **open set words** (closed set words and suffixes are part of syntax and morphology, respectively)

- Semantic anomalies and ambiguities
- Word senses and semantic relations, e.g., semantic similarity
- Figurative use of words

Example:

- *We smashed/unhinged the window/door.*
- *We climbed/walked through the window/door.*

door, window = $\left\{ \begin{array}{l} \text{opening in wall} \\ \text{cover for opening} \end{array} \right.$

- Ontologies (dictionaries for recording word meaning)
- Word sense disambiguation
- Models of similarity between word meanings, incl. clustering
- Frame Semantics and semantic role labelling

- L10: Coherence
- L11: Anaphora
- L12: Coreference Resolution
- L13: Dialogue Structure
- L14: Rhetorical Discourse Structure
- L15: Discourse Structure in Scientific Texts
- L16: Applications

Discourse Phenomena

Definition: Any linguistic phenomena of cohesion (“belonging together”) that acts across sentence boundaries.

- What makes sentences “hang together”?
- Phenomena **within** one sentence handled by syntax, semantics, and, to a certain degree, pragmatics
- 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 ...*

- Resolution of coreference/anaphora
- Generation of referring expressions
- Lexical chains and models of lexical coherence
- Computational theories of how text is logically structured

Syntagmatic vs paradigmatic affinities

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* an utterance:
 - *dog ... barked ...*
- **Paradigmatic**: semantic affinities between two grammatically identical words which can **replace** each other in an utterance:
 - *I haven't yet fed the [dog/cat/*lamppost].*

Derive information about a word's meaning from its relation with actual and potential linguistic contexts.

Possible sources of decisions about word meaning:

- Natural utterances
- Intuitive semantic judgements about some linguistic material, by native speakers

Introspection and Intuition

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.
- Diagnostic Methodology: Construct a linguistic context for a word, then judge normality/truthconditional conditions of that utterance.
- Alternative methods: Rank or compare normality of a sentence

- 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*

Semantic Traits

The meaning of a word can be seen as made up from the meaning of other words. Semantic traits can be:

- **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.*

- **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.*

Semantic traits and affinity highlight different aspects of meaning:

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

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*

- *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.*

Syntactic vs. Semantic Anomalies

1. *It is too light for me to lift.*
2. *Them yesterday goed home.*

Replacement Test: Choose appropriate replacement that makes the sentence normal.

- Open-class item replacements are called lexical and point to a semantic anomaly.
 - Closed-class item replacements are called syntactic and point to a syntactic anomaly.
1. *light* → *heavy*
 2. *them* → *they*; *goed* → *went*

Deviances cannot always be cured by either lexical or grammatical replacement:

1. *The table saw Arthur.* (rhinoceros/ was seen by)
2. *The cake was baked.* (baked/ taken, shaken, forsaken).

Full replacement test:

1. If lexical substitution is the **only** way to make the sentence normal, then the anomaly is semantic.
2. If syntactic substitution is the **only** way to make the sentence normal, then the anomaly is syntactic.
3. If both changes are possible, the anomaly is semantic if the open set substitutions share semantic properties, otherwise syntactic.

Test 2: Context Manipulation

If you can create a context which makes the sentence normal, then the anomaly is syntactic.

- E.g., Arthur is paranoid and believes that tables follow him.
- Quoting does not count: *As my Portuguese plumber says...*

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

Compound nouns:

- *A black bird sang softly in the willow tree.*
- *A blackbird sang softly in the willow tree.*

Idioms:

- *Arthur kicked the bucket.*
- *This cooked Kim’s goose.*

Judging Semantic Contrast

Humans can feel a difference when comparing semantic contrast:

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

Which semantic contrasts are equivalent?

- 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 difference is the same, then the item is a semantic constituent:

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

His remarks are $\left\{ \begin{array}{l} im- \\ - \end{array} \right\}$ *pertinent*. \neq What you suggest is $\left\{ \begin{array}{l} im- \\ - \end{array} \right\}$ *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 $m \left\{ \begin{array}{l} -at \\ -oss \end{array} \right\}$. \neq He does not like his new $b \left\{ \begin{array}{l} -at \\ -oss \end{array} \right\}$.

Now let's look at compounds again:

I saw a $\left\{ \begin{array}{l} black- \\ blue- \end{array} \right\}$ *bird in the garden*. \neq Cynthia wore $\left\{ \begin{array}{l} black \\ blue \end{array} \right\}$ *stockings*.

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

A modifier in a compound noun is its own semantic constituent if you can coordinate the modifier in a compound noun with another modifier in second position. The relationship between the second coordinated modifier and the semantic head must be clearly compositional.

- ? *blue and carnivorous whales*

This means that *blue whale* is one semantic component. *blue* is not (in this context).

Idioms

- Minimal semantic constituents which consist of more than one word.
- Definition: the meaning of an idiom cannot be inferred as a compositional function of the meaning of its parts.
- (non-idiomatic expressions are called **semantically transparent**).
- *pull somebody's leg*
- *be off one's rocker*

Syntactic Variability Tests:

- ? *Arthur has a bee, apparently, in his bonnet.* (insertion)
- ? *Arthur kicked the large bucket.* (modification)

- **Collocation:** lexical items which are semantic constituents but habitually co-occur.
- He's a $\left. \begin{array}{l} \text{heavy} \\ \text{light} \end{array} \right\} \text{drinker.} = \text{They are } \left\{ \begin{array}{l} \text{heavy} \\ \text{light} \end{array} \right\} \text{smokers.}$
- **Metaphor:** figure of speech in which an entity is interpreted as something else.
- **Dead metaphor:** the image has been established in the language, i.e., creative, situational figurative images are excluded.
 - *The locomotive snorted and belched its way across the plain.*
 - “*donner sa langue au chat*” (give your tongue to the cat)
 - “*appeller un chat un chat*” (call a cat a cat)

Rephrasing Test

If rephrasing results in similar semantics, the multi-word entity is not a semantic constituent (thus a dead metaphor, not an idiom).

Dead metaphors:

- *They tried to sweeten the pill.* \approx *They tried to sugar the medicine.*
- *We shall leave no stone unturned in our search for the culprit.* \approx *We shall look under every stone in our search for the culprit.*

Idioms:

- *John pulled his sister's leg* $\not\approx$ *John tugged at his sister's leg*
- *Arthur kicked the bucket* $\not\approx$ *Arthur tipped over the water receptacle*

- *Sue visited her cousin.*

cousin is underspecified wrt [male/female]

- *We finally reached the bank.*

bank has two distinct senses, with no general meaning covering both [financial institution, side of river]

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.

Typically there is more than one ambiguous word in the sentence.

- *Several rare ferns grow on the steep banks of the burn where it runs into the lake.*

Ambiguous: *rare, steep, bank, burn, run*

Hearer selects that combination of lexical readings which leads to the most normal possible utterance-in-context. [Assumption of cooperation in communication Grice]

Word form X is ambiguous if it stands in relation Y with other word forms Z₁ and Z₂ 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.

First Test for Ambiguity

If a word form is ambiguous, then its senses should not in every case be totally conditioned by their contexts (unlike in contextual modulation).

Contextual modulation:

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

Replacement with synonym (“crowned head”, “sovereign”) loses no information → all information is derived by contextual modulation.

Ambiguity:

- *His wife is the manager of the local bank.*
- *At this point, the bank was covered with brambles.*

Replacement with synonym for both (“place”) loses information. So we have failed to show that the interpretation of “bank” is the result of contextual modulation of an underspecified meaning.

dog has two senses: canine/male dog

- *Incredibly, John prefers an aged, half-blind bitch to a dog, as his canine companion.*
- *John prefers bitches to dogs/?canines.*

The context of 'dog' acts like a trigger and chooses one of two pre-existing, rich bundles of properties, unlike in modulation.

Second Test for Ambiguity: Yes/No Test

32

If you can construct a situation where a question containing a potentially ambiguous word can be answered both yes and no, then the word form is ambiguous.

- *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.*

Contrast to underspecification:

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

Contexts which activate more than one sense of a word 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.*

Identity test (Crossed Interpretation test) for Ambiguity

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

This includes the possibility that one has adopted a boy and the other a girl.

- *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.

We cannot coordinate the two senses of mouth:

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

But we can demonstrate a sense spectrum:

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!*

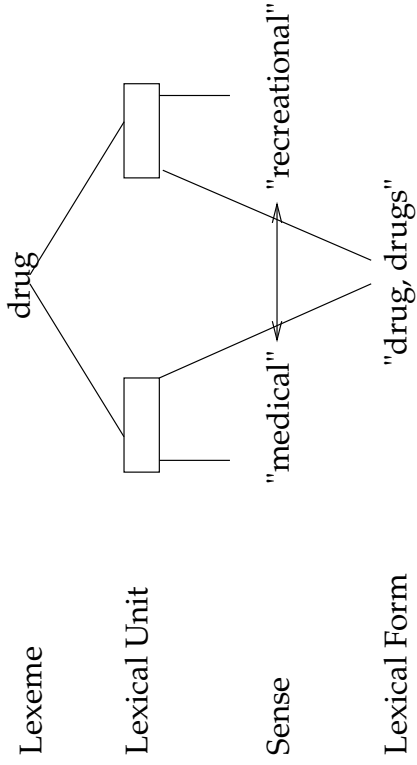
Lexical units and lexemes

Lexical unit: a form-meaning complex with relatively stable and discrete semantic properties which stand in meaning relations such as antonymy (long:short) and hyponymy (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.

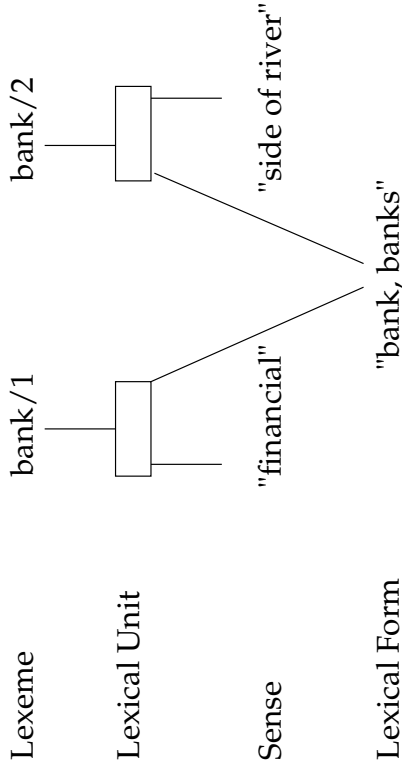
Lexical form: family of word forms differing only in inflectional morphology. Must be a semantic constituent, i.e., can include multi-word units.

Lexeme: contains one or more lexical units of the same POS, if either

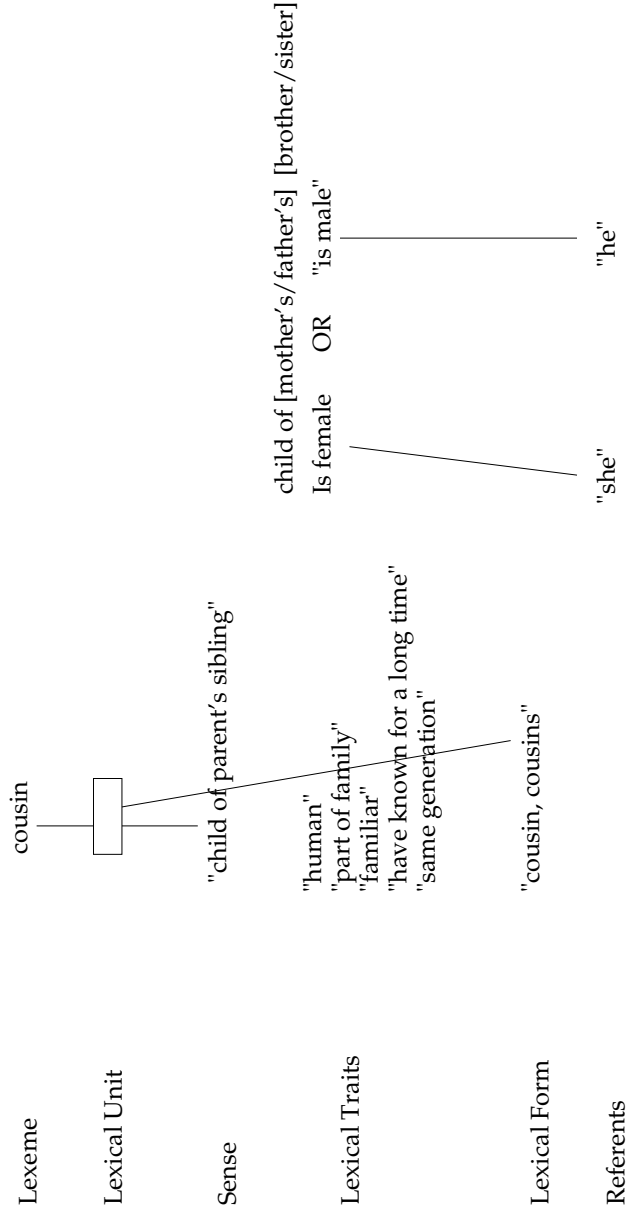
- there exists a lexical rule which permits the existence of the sense of one from the existence of the sense of the other. Recurrent semantic contrast between senses is evidence of a lexical rule (e.g., unit and type readings of pieces of clothing).
- the senses are local senses belonging to a sense spectrum (e.g., *mouth* of a river and *human mouth*)



A lexeme which has a number of senses is polysemous.



A lexical form which realises lexical units belonging to more than one lexeme are **homonymous**.



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

Summary of Today

- Semantic tests (normality judgements; comparison of semantic contrast)
- Syntactic vs. semantic ambiguities/anomalies (+ tests)
- Semantic traits (+ tests)
- Semantic constituency; idioms vs. dead metaphors (+ tests)
- Sense Ambiguity vs. Underspecification
- Definition of lexical item and lexeme

Background Reading:

- D. A. Cruse, *Lexical Semantics*. Cambridge Textbooks in Linguistics, 1986. Chapters 1-3