

# Semantic Theory: Lexical Semantics I

Summer 2007

M.Pinkal/ S. Thater



- 03-07-07 Lecture: Lexical Semantics I
- 05-07-07 Lecture: Lexical Semantics II
- 10-07-07 Lecture: Everything else
- 12-07-07 Exercise: lexical Semantics
- 17-07-07 Question time, Sample exam
- 19-07-07 Individual question time
- 25-07-07 **Final exam, 11:00 (s.t.)**



## Structure of this course

- Sentence semantics
- Discourse semantics
- Lexical semantics



## Dolphins in First-order Logic

*Dolphins are mammals, not fish.*

$\forall d (\text{dolphin}'(d) \rightarrow \text{mammal}'(d) \wedge \neg \text{fish}'(d))$

*Dolphins live-in pods.*

$\forall d (\text{dolphin}'(d) \rightarrow \exists x (\text{pod}'(p) \text{live-in}'(d,p))$

*Dolphins give birth to one baby at a time.*

$\forall d (\text{dolphin}'(d) \rightarrow$   
 $\quad \forall x \forall y \forall t (\text{give-birth-to}'(d,x,t) \text{give-birth-to}'(d,y,t)$   
 $\quad \quad \rightarrow x=y)$



## Dolphins in First-order Logic

*Dolphins are mammals, not fish.*

$\forall d (\text{dolphin}'(d) \rightarrow \text{mammal}'(d) \wedge \neg \text{fish}'(d))$

*Dolphins live-in pods.*

$\forall d (\text{dolphin}'(d) \rightarrow \exists x (\text{pod}'(p) \text{live-in}'(d,p))$

*Dolphins give birth to one baby at a time.*

$\forall d (\text{dolphin}'(d) \rightarrow$   
 $\quad \forall x \forall y \forall t (\text{give-birth-to}'(d,x,t) \text{give-birth-to}'(d,y,t)$   
 $\quad \quad \rightarrow x=y)$



## The dolphin text

Dolphins are mammals, not fish. They are warm blooded like man, and give birth to one baby called a calf at a time. At birth a bottlenose dolphin calf is about 90-130 cms long and will grow to approx. 4 metres, living up to 40 years. They are highly sociable animals, living in pods which are fairly fluid, with dolphins from other pods interacting with each other from time to time.



## Sentence Semantics (Predicate Logic)

Dolphins **are** mammals, **not** fish. **They are** warm blooded **like** man, **and** give birth to **one** baby called **a** calf at **a** time. At birth **a** bottlenose dolphin calf **is about** 90-130 cms long **and will** grow to **approx.** 4 metres, living up to 40 years. **They are** highly sociable animals, living in pods **which are fairly** fluid, with dolphins from **other** pods interacting with **each other** from time to time.



## Lexical semantics

**Dolphins** are **mammals**, not **fish**. They are **warm** blooded like **man**, and **give birth** to one **baby** called a **calf** at a **time**. At **birth** a **bottlenose dolphin calf** is about 90-130 **cms long** and will **grow** to approx. 4 **metres**, **living** up to 40 **years**. They are highly **sociable animals**, **living** in **pods** which are fairly **fluid**, with **dolphins** from other **pods interacting** with each other from **time** to **time**.



## Major word-semantic categories

- Function words:
  - Connectives and quantifiers
  - Auxiliary and modal verbs
  - Temporal and modal adverbials
  - Anaphoric pronouns, articles
  - Degree modifiers, Copula, ...
- Content words
  - Common nouns
  - Full verbs
  - Adjectives
- Other
  - Named Entities (Persons, institutions, geographic entities, dates)
  - Numbers
  - Etc.



## Challenges in lexical semantics

- The multiplicity of senses: Lexical ambiguity
- The diversity of meaning information (in a given sense)
- The size of the lexicon



## The word-meaning-relation

- The relation between
  - phonological/ orthographic words and
  - senses/ word meanings/ conceptsis not one-to-one.
- One sense/ concept can be encoded in different phonological words: **Synonymy**
- One phonological word can be associated with several senses: **Lexical ambiguity**



## Lexical Ambiguity

- Ambiguity between unrelated senses: **Homonymy**  
(*bank as river bank or financial institution*)
- Ambiguity between semantically related concepts:  
**Polysemy**
- Homonyms are typically represented as different lexical entries (**lexemes, lemmas**), cases of polysemy as single entries with multiple sense descriptions.
- We distinguish
  - unsystematic cases of polysemy (e.g., *bank: financial institution - blood bank; case: carton, suitcase, pillowcase; to serve a meal / as a president*)
  - systematic polysemy (*rabbit, dear, chicken: animal – meat; fast: fast car, fast road, fast driver*)



# Lexical Ambiguity

- Ambiguity, in particular polysemy, is a pervasive feature of the lexicon. The number of senses increases with the frequency of a word (up to about 50, according to standard dictionaries and WordNet).
- There is no clear outer boundary for the set of readings of a lexical item, because of meaning extensions and figurative uses (metaphor, metonymy) that can range from fully conventionalized to completely novel
  - *to grasp an idea, the Wikipedia as a gold mine, data-mining; to wear rabbit; the (computer) desktop, mouse, folder, file*
- There is no clear inner criterion for the distinction between senses (vs. different usages of the same sense, collocations):
  - *onion (eating onions – growing onions)*



# What is a dolphin?





## Diversity of word meaning

- The concepts corresponding to single readings of a word are typically multi-layered, consisting of heterogeneous kinds of information (crossing modality), among other things:
  - **Propositional** information – can be paraphrased in language, symbolically represented in a logical framework
  - Visual (or other sensory) **prototypical** information
  - Stereotypical information – valid in the „normal“, default case
- No clear-cut boundary between word meaning and world knowledge.
- No clear-cut boundary between common-sense meaning and domain-specific „ontological“ information.



## Size and complexity of the lexicon

- The lexicon is very large (100 – 200K words in standard dictionaries or WordNet).
- No upper boundary to the size of the lexicon:
  - compounds, foreign words, special terminology (1.5 million new words in a 200 million word corpus of German)
  - subject to extreme application-dependent variation concerning extent and relevant dimensions
- The lexicon is heterogenous: multimodal and multi-dimensional



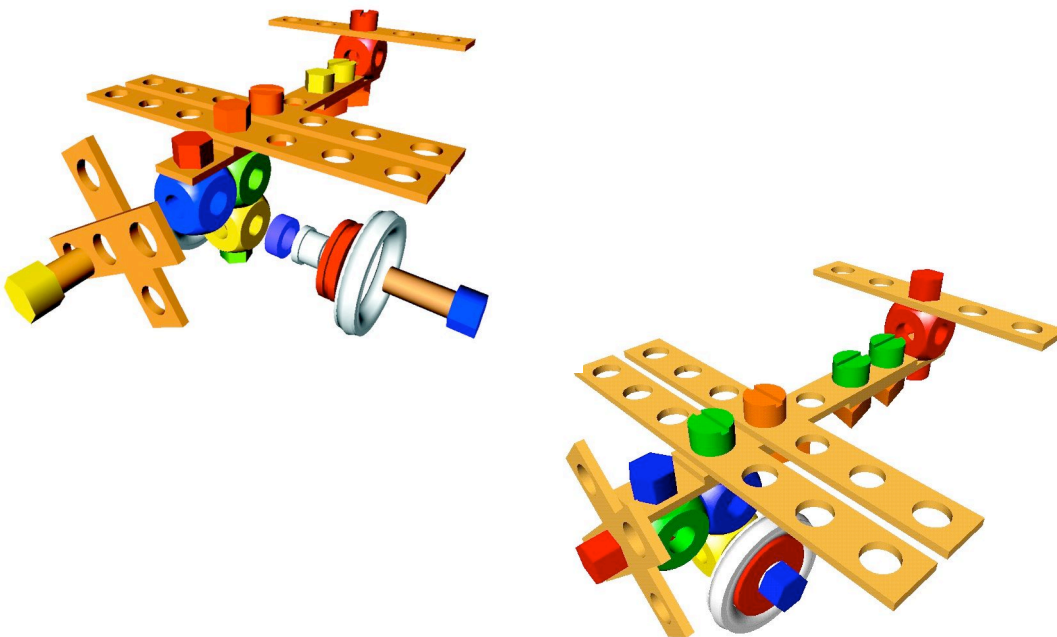


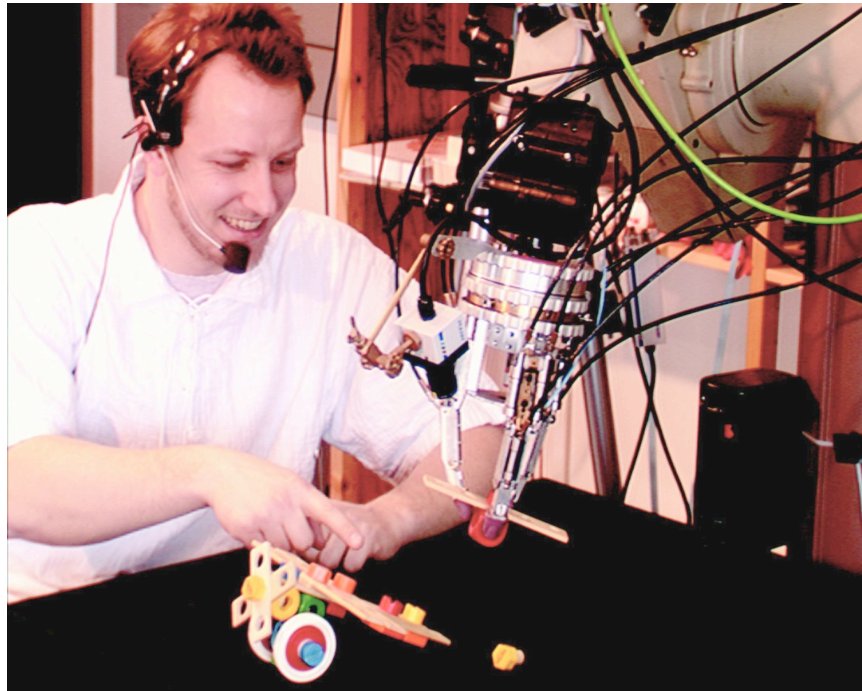
## Central questions

- How do we organise/ represent lexical semantic information?
- How do we provide lexical semantic resources?
- Which kind of lexical-semantic information is required – **given a (type of) application?**
- Example 1: Robotics
- Example 2: Information Access



## A robotics application





## An information search application

*Which companies sell motor vehicles?*

- Query Expansion with hyponyms:  
{company, sell, „motor vehicle“, car, bus, motorcycle, bike, truck}

Axioms for inference/ entailment checking:

- Axiom  $\forall x (\text{truck}(x) \rightarrow \text{motor\_vehicle}(x))$ , expressing hyponymy relation between truck and motor vehicle
- and text *Volvo sells trucks* :  $\exists x (\text{truck}(x) \wedge \text{sell}(\text{volvo}, x))$
- together entail:  $\exists x (\text{motor\_vehicle}(x) \wedge \text{sell}(\text{volvo}, x))$ ,
- which is a direct answer to the above question.



## WordNet

- WordNet represents a layer of the semantic lexicon of English as a **network of semantic relations**, with the **hyponymy** relation and its inverse relation, **hypernymy**, as its backbone.
- The nodes of the semantic network are „**synsets**“: Sets of synonymous words, which represent concepts/ word senses.
- Synsets directly provide synonymy information, and information about the word-concept mapping: A (orthographic) word has all those senses/synsets as readings, of which it is a member.
- In cases where no or too few synonyms are available for sense distinction, WordNet glosses and examples help to disambiguate.



## Senses of *car*

- **S:** (n) **car**, auto, automobile, machine, motorcar
- **S:** (n) **car**, railcar, railway car, railroad car
- **S:** (n) **car**, gondola
- **S:** (n) **car**, elevator car
- **S:** (n) cable car, **car**



## Synsets + glosses + examples

- **S:** (n) **car**, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) "*he needs a car to get to work*"
- **S:** (n) **car**, railcar, railway car, railroad car (a wheeled vehicle adapted to the rails of railroad) "*three cars had jumped the rails*"
- **S:** (n) **car**, gondola (the compartment that is suspended from an airship and that carries personnel and the cargo and the power plant)
- **S:** (n) **car**, elevator car (where passengers ride up and down) "*the car was on the top floor*"
- **S:** (n) cable car, **car** (a conveyance for passengers or freight on a cable railway) "*they took a cable car to the top of the mountain*"



## Hyponyms of *motor vehicle*

- **S:** (n) **motor vehicle**, automotive vehicle (a self-propelled wheeled vehicle that does not run on rails)
- **direct hyponym / full hyponym**
  - **S:** (n) amphibian, amphibious vehicle (a flat-bottomed motor vehicle that can travel on land or water)
  - **S:** (n) bloodmobile (a motor vehicle equipped to collect blood donations)
  - **S:** (n) car, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) "*he needs a car to get to work*"
  - **S:** (n) doodlebug (a small motor vehicle)
  - **S:** (n) four-wheel drive, 4WD (a motor vehicle with a four-wheel drive transmission system)
  - **S:** (n) go-kart (a small low motor vehicle with four wheels and an open framework; used for racing)
  - **S:** (n) golfcart, golf cart (a small motor vehicle in which golfers can ride between shots)
  - **S:** (n) hearse (a vehicle for carrying a coffin to a church or a cemetery; formerly drawn by horses but now usually a motor vehicle)
  - **S:** (n) motorcycle, bike (a motor vehicle with two wheels and a strong frame)
  - **S:** (n) snowplow, snowplough (a vehicle used to push snow from roads)
  - **S:** (n) truck, motortruck (an automotive vehicle suitable for hauling)



## WordNet: More Semantic Relations

- **Meronymy**, the part-of relation, and its inverse relation, **holonymy**, with three (well-motivated) sub-relations:
  - Physical Part – Whole relation: *branch – tree*
  - Member – Group relation: *tree – forest*
  - Substance – Object relation: *wood – tree*
- **Antonymy**, a general super-concept for opposition/contrast, comprising
  - Contrast (or antonymy in the narrower sense): *good – bad, expensive – cheap*
  - Complementarity: *man – woman, married – single*
  - Converse/ inverse relation: *buy – sell, ancestor - descendant* (according to Lyons 1979)



## WordNet

- English WordNet is by far the largest lexical-semantic resource:
  - 150.000 lexical items
  - 120.000 synsets
  - 200.000 word-sense pairs
- WordNet is extensively used in many Language technology applications.
- Versions of WordNet currently available for about 45 languages (with large differences in coverage, design, and availability)
- "GermaNet": a German WordNet version with about 100.000 lexical items.



## WordNet

- WordNet consists of different, basically unrelated databases for common nouns, verbs, adjectives (and adverbs). There are more semantic relations for the POS-specific databases (in particular for verbs).
- Different parts of WordNet differ in their granularity. In general, WordNet tends to be too fine-granular for many purposes (assuming sense distinctions and requiring word-sense disambiguations in cases of subtle variants of usage).
- WordNet focusses on paratactic semantic relations between single words. It lacks information which is necessary for building predicate-argument structure.



## Conventional lexical-semantic resources

- **Monolingual dictionaries**, alphabetically ordered, provide informal meaning information about the readings of a word informally, through synonyms, glosses, typical examples, etc.
- A **thesaurus** groups the lexicon of a language according to the semantically relatedness of the words.



# Ontologies

- An **ontology** is the product of an attempt to formulate an exhaustive and rigorous conceptual scheme about a domain. An ontology is typically a hierarchical data structure containing all the relevant entities and their relationships and rules within that domain (eg. a **domain ontology**).
- An ontology which is not tied to a particular problem domain but attempts to describe general entities is known as a **foundation ontology** or **upper ontology**.  
(Wikipedia)



# Ontologies, Overview

- Special Ontologies: Terminological information for certain subjects /areas of research and technology. Most wide-spread are bio-medical ontologies.
- "Upper-model ontologies" provide common-sense, general terminological knowledge.
- Ontologies are typically **formalised**, using a logical representation formalism to encode conceptual knowledge.