

Semantic Theory

Summer 2005

Lexical Semantics

M. Pinkal / A. Koller

Technical Stuff

- Update Course Schedule:

- | | |
|-------------|---|
| Ø Thu 23.6. | L Lexical Semantics I (MP) |
| Ø Tue 28.6. | L Lexical Semantics II (Katrin Erk) |
| Ø Thu 30.6. | E Discourse Semantics |
| Ø Tue 5.7. | L Lexical Semantics III (MP) |
| Ø Thu 7.7. | E Lexical Semantics |
| Ø Tue 12.7. | --- (Accreditation) |
| Ø Thu 14.7. | Question Time,
Discussing "Sample Exam" (MP, AK) |
| Ø Tue 19.7. | Final Exam 11:00-13:00 (120 min.!!) |

- Registration for Final Exam, Deadline: 1.7.

What should a semantic theory provide?

- A framework to specify **word meaning**
- The **composition** process leading from word meanings to sentence information
- The building of a semantic **discourse representation** from a sequence of sentences in a text (or piece of dialogue)
- Disambiguation/ **resolution** mechanisms selecting the intended information of an utterance from the large number of linguistically possible interpretations
- **Inference** mechanisms leading from the given utterance information to other relevant information

Semantic Theory 2005 © M. Pinkal/A.Koller UdS Computerlinguistik

3

How to Catch a Dolphin with a Wordnet



Semantic Theory 2005 © M. Pinkal/A.Koller UdS Computerlinguistik

4

How to Catch a Dolphin in a Wordnet

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.

Content Words: One-place predicates

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.**

Content Words: One-place predicates

Common Nouns + adjectives

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.

Content Words: Relational expressions

Verbs (+ deverbal/relational nouns + relational adjectives)

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.

Other

Numbers

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.

Multi-word expressions

Compounds, idioms, (metaphors, metonymies)

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.

Proper names / Named Entities

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.

There was once a dolphin called Flipper in the Atlantic Ocean close to Miami Beach in the U.S. state Florida.

Lexical Semantics

- Function words:
 - Ø Connectives and quantifiers
 - Ø Modal verbs and particles
 - Ø Anaphoric pronouns and adverbs
 - Ø Degree modifiers, Copula, ...
 - Ø Prepositions (?)
- Content words
 - Ø Standard one-place predicates: Common nouns, adjectives, (intrans. verbs)
 - Ø Relational concepts with overt argument: Verbs, nouns, adjectives (prepositions?)
- Other
 - Ø Named Entities (Person, Company, Institution, Geographic names, Dates, ...)
 - Ø Numbers
 - Ø ...

Semantic Relations

- Synonymy
 - Ø case – bag
- Hyponymy/Hypernymy (ISA relation)
 - Ø dolphin – mammal
- Meronymy/Holonymy
 - Ø Part – Whole : branch – tree
 - Ø Member – Group: tree – forest
 - Ø Matter – Object: wood – tree
- Contrast:
 - Ø Complementarity: boy – girl
 - Ø Antonymy: long - short

Semantic Theory 2005 © M. Pinkal/A.Koller UdS Computerlinguistik

13

Monolingual Dictionaries, Thesauri

- Monolingual dictionaries, alphabetically ordered lemmas with enumeration and informal description of readings
 - Ø Oxford English Dictionary
 - Ø Webster's
 - Ø Wahrig /Duden
- A thesaurus is a hierarchically structured lexicon of a language:
 - Ø Roget's Thesaurus (English, since 1805)
 - Ø Dornseiff's "Deutscher Wortschatz nach Sachgruppen" (German, 1910)

Semantic Theory 2005 © M. Pinkal/A.Koller UdS Computerlinguistik

14

Ontologies

- In philosophy, **ontology** (from the Greek *on* = *being* and *λόγος* = *word/speech*) is the most fundamental branch of metaphysics. It studies being or existence as well as the basic categories thereof -- trying to find out what entities and what types of entities exist. Ontology has strong implications for the conceptions of reality.
- Basic Aristotelian categories:
 - Ø Substance, Quantity, Quality, Relation, Place, Time, Posture, State, Action, and Passion

Ontologies in Information Technology

- 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**). The computer science usage of the term *ontology* is derived from the much older usage of the term ontology in philosophy.
- 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, the whole article is worth reading)

Ontologies and Logic

- Ontologies are typically "formal ontologies": They use a logical representation formalism to encode conceptual knowledge:

Ø Versions of Description Logic (à OWL)
Ø Predicate /modal logic

General-purpose Ontologies

- SUMO (Suggested Upper Merged Ontology)
- MILO (Mid-level Ontology)
- Web interface for SUMO and MILO:
Ø <http://berkelium.tekknowledge.com:8080/sigma/home.jsp>
- Size: 2.600 concepts, 6.000 relations, 2.000 rules

The Concept "Fish" in SUMO [1]

- Description of Concept:
- (documentation **Fish** "A cold-blooded aquatic **Vertebrate** characterized by fins and breathing by gills. Included here are **Fish** having either a bony skeleton, such as a perch, or a cartilaginous skeleton, such as a shark. Also included are those **Fish** lacking a jaw, such as a lamprey or hagfish.")
- Relationship to other concepts:
 - (subclass **Fish** **ColdBloodedVertebrate**)
 - (disjointDecomposition **ColdBloodedVertebrate** **Amphibian** **Fish** **Reptile**)

Semantic Theory 2005 © M. Pinkal/A.Koller UdS Computerlinguistik

19

The Concept "Fish" in SUMO[2]

- A rule:
(=>
 (instance ?FISH Fish)
 (exists
 (?WATER)
 (and
 (inhabits ?FISH ?WATER)
 (instance ?WATER Water))))
- ... and its semi-colloquial paraphrase:
"if instance FISH **Fish**, then there exists WATER such that inhabits FISH WATER and instance WATER **Water**"

Semantic Theory 2005 © M. Pinkal/A.Koller UdS Computerlinguistik

20

WordNet

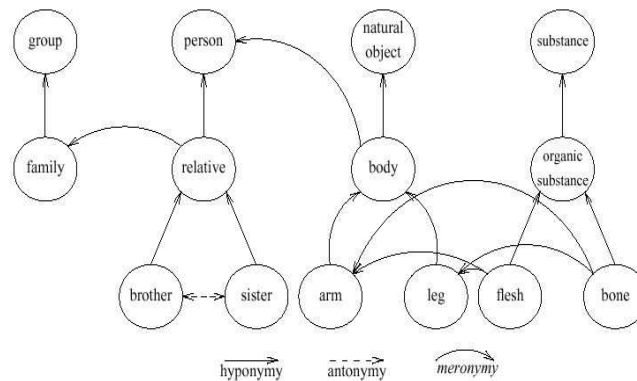
- WordNet is kind of an ontology, using language-specific natural vocabulary instead of concept labels
- A problem: There is no 1:1 relation between words and concepts:
 - Ø The same word can express different concepts (ambiguity)
 - Ø The same concept can be expressed by different words (synonymy)
- The WordNet solution: concepts are represented by „synsets“: Sets of synonymous words. „synsets“ form the basic units of WordNet
- Synsets are connected by different kinds of semantic relations (see above)

An example: *case*

- Ø {*case*, *carton*}
- Ø {*case*, *bag*, *suitcase*}
- Ø {*case*, *pillowcase*, *slip*}
- Ø {*case*, *cabinet*, *console*}
- Ø {*case*, *casing* (the enclosing frame around a door or window opening)}
- Ø {*case* (a small portable metal container)}

An example

Figure 2. Network representation of three semantic relations among an illustrative variety of lexical concepts



Semantic Theory 2005 © M. Pinkal/A.Koller UdS Computerlinguistik

23

WordNet

- English WordNet: about 150.000 lexical items
 Ø Web Interface: <http://wordnet.princeton.edu/cgi-bin/webwn2>
 Ø General Info: <http://wordnet.princeton.edu/>
- "GermaNet": a German WordNet version with about 90.000 lexical items
- Versions of WordNet for available for about 30 languages
- WordNet consists of different, basically unrelated databases for common nouns, verbs, adjectives (and adverbs)
- The respective hierarchies have a number of "unique beginners" each.

Semantic Theory 2005 © M. Pinkal/A.Koller UdS Computerlinguistik

24

Unique Beginners for WordNet Nouns

Table 1
List of 25 unique beginners for WordNet nouns

<i>{act, action, activity}</i>	<i>{natural object}</i>
<i>{animal, fauna}</i>	<i>{natural phenomenon}</i>
<i>{artifact}</i>	<i>{person, human being}</i>
<i>{attribute, property}</i>	<i>{plant, flora}</i>
<i>{body, corpus}</i>	<i>{possession}</i>
<i>{cognition, knowledge}</i>	<i>{process}</i>
<i>{communication}</i>	<i>{quantity, amount}</i>
<i>{event, happening}</i>	<i>{relation}</i>
<i>{feeling, emotion}</i>	<i>{shape}</i>
<i>{food}</i>	<i>{state, condition}</i>
<i>{group, collection}</i>	<i>{substance}</i>
<i>{location, place}</i>	<i>{time}</i>
<i>{motive}</i>	

WordNet: Advantages

- WordNet is big and has very large coverage (concerning both words and word senses (compared to SUMO/MILO))
- WordNet allows, among other things
 - Ø query expansion for Information Retrieval
 - Ø basic inferences via semantic relations
- The mapping from NL expressions to WordNet concepts (in a given language) is trivial (modulo ambiguity) (compared e.g. to CYC)
- CYC: a huge ontology which is very expensive, but not really useful for LT purposes. "Open CYC" is free, but has no coverage
- SUMO/MILO are supposed to be language-neutral – Are they really?

WordNet: Disadvantages

- Different parts of WordNet have different granularity for the description of word senses. In general, WordNet is too fine-granular for many purposes.
- There are WordNet versions for a large number of languages, but there is no real multi-lingual WordNet: The different WordNet differ in coverage, format, and availability.
- WordNet focusses on paratactic semantic relations between single words. It does not provide the core lexical information needed for composition:
 - Ø Predicate-argument structure /Semantic roles.

Thematic roles: Some observations

- (Quasi-)Equivalent sentences with different realization of "the same" semantic argument positions:

Ø *Mary likes John*

Ø *John pleases Mary*

Ø *Mary gave Peter the book*

Ø *Peter received the book from Mary*

Some observations [2]:

- Verbs that have alternative grammatical valency patterns for the realization of (roughly) the same proposition:

Ø Man kann den Wein gut trinken

Ø Der Wein trinkt sich gut

Ø Maria hat das Buch von Hans geliehen

Ø Hans hat Maria das Buch geliehen

Semantic Theory 2005 © M. Pinkal/A.Koller UdS Computerlinguistik

29

Some observations [3]:

- Verbs with varying number of explicit argument positions, and varying realization of "the same argument".

Ø The window broke

Ø A rock broke the window

Ø John broke the window with a rock

Ø The plane flew to Frankfurt

Ø John flew the plane to Frankfurt

Ø John flew Bill with the plane to Frankfurt

Semantic Theory 2005 © M. Pinkal/A.Koller UdS Computerlinguistik

30

Thematic Roles (Fillmore 1968)

- **Frames** are the units for the conceptual modelling of the world: structured schemata representing complex situations, events, and actions. The meaning of words in terms of the part which they play in frames.
- **Thematic roles** describe the conceptual participants in a situation in a generic way, independent from their grammatical realization.

Examples for Thematic Roles

- Agent
- Theme/ Patient/ Object
- Recipient
- Instrument
- Source
- Goal
- Beneficient
- Experiencer

Examples Annotated with Thematic Roles

\emptyset [*The window*]_{pat} *broke*

\emptyset [*A rock*]_{inst} *broke* [*the window*]_{pat}

\emptyset [*John*]_{ag} *broke* [*the window*]_{pat} [*with a rock*]_{inst}

\emptyset [*Peter*]_{ag} *gave* [*Mary*]_{rec} [*the book*]_{pat}

\emptyset [*Mary*]_{rec} *received* [*the book*]_{pat} [*from Peter*]_{ag}

Thematic Roles

- allow more abstract/ generic semantic representations
- support the systematic description of selection preferences and constraints
- support the encoding and application of general inference rules
- support the semantic interpretation process (role linking)

Role linking, example

give: SB à Agent
OA à Theme
OD à Recipient

get: SB à Recipient
OA à Theme
OP-from à Agent

The „Role Dilemma“

- In Fillmore's original theory and in early KR research a small, closed, and universally applicable inventory of roles is postulated.
- This assumption is untenable, given the semantic richness of natural languages.

Extreme solutions

- Every verb comes with its individual roles:
Ø „Giver“, „Given“, „Givee“ (*HPSG, Situation Semantics*)
Ø $\theta_1, \theta_2, \theta_3$ (Theta Roles in Generative Grammar)
Problem: Generalization is lost.
- All concepts use the same role concepts:
Ø *arg1, arg2, arg3* (e.g., *Verbmobil*)
Problem: Roles do not have a meaning anymore.

Fillmores Frame-semantic Concept (1976)

- „...first identify the phenomena, experiences, or scenarios represented by the meanings of the *target words* ...“
- „...then identify labels to the parts or aspects of these which are associated with specific means of linguistic expression ...*frame elements* ...“

... implemented in the Berkeley FrameNet Database (since 1996)

- **Frames**: an inventory of conceptual structures modelling a prototypical situation like "COMMERCIAL_TRANSACTION", "COMMUNICATION_REQUEST", "SELF_MOTION"
- Semantic roles are **locally valid** only (and accordingly called "Frame Elements" (FE):
 - Ø FEs of the COMMUNICATION_REQUEST frame: SPEAKER, ADDRESSEE, MESSAGE, ...
 - Ø FEs of the COMMERCIAL_TRANSACTION frame: BUYER, SELLER, GOODS, PRICE, ...
- A set of "**target words**" associated with each frame: e.g., for COMMERCIAL_TRANSACTION:
 - Ø buy, sell, pay, spend, cost, charge,
 - Ø price, change, debt, credit, merchant, broker, shop
 - Ø tip, fee, honorarium, tuition

The Berkeley FrameNet Database

The FrameNet database consists of:

- A data-base of frames with
 - Ø Descriptions of frames with inventory of Roles/Frame elements and associated lemmas
 - Ø Frame-to-Frame Relations
- A lexicon with
 - Ø Frame information
 - Ø Grammatical realisation patterns (Role Linking)
 - Ø Annotations of example sentences (from BNC) for all use variants of words

The Berkeley FrameNet Database

- Current release: 700 frames, about 8000 lexical units (mostly verbs)
- Planned: A total of 15000 verb descriptions
- <http://framenet.icsi.berkeley.edu/>

FrameNet: Advantages

- A very deliberate and careful unified modeling of the core lexicon of English (relational) expressions, mostly verbs, but also deverbal nouns and relational adjectives, which supports
 - Ø semantic representation at an appropriate level of granularity and abstraction
 - Ø semantic construction via grammatical realization patterns
 - Ø inference
 - Ø Multi-lingual applications

FrameNet: Disadvantages

- Lack of coverage
- Too few Frame-to-Frame Relations
- No frame-internal distinctions with respect to semantic features other than role (Frame-Element) relations
- No powerful interfaces for language technology purposes

FrameNet going multi-lingual

- SALSA: The Saarbrücken Lexical Semantics Annotation and Analysis Project – A corpus-based, large, application-oriented lexical-semantic resource based on FrameNet
- Spanish and Japanese FrameNet under work.