Semantic Theory

Lexical Semantics 1

Manfred Pinkal/ Stefan Thater
Saarland University
Summer 2009

Course Overview

- Inference in natural language; representing and processing meaning information in predicate logic
- Lexical semantics: Lexical-semantic resources; ontologies; Description Logic; Thematic roles; frames; event structure
- Sentence semantics: Montague-style type-theoretical representations; semantics construction; scope ambiguities; underspecification
- Discourse semantics: Anaphora; Discourse Representation Theory (DRT); presuppositions

Today's Agenda

- FLST Recap
  - Diversity of word meaning
  - Semantic Relations
  - WordNet
  - Distributional Similarity
- Predicate-Argument Structure
  - Inversion, Verb Alternations, Semantic Role
  - PropBank
  - FrameNet

Diversity of word meaning

Diversity of word meaning

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 or conceptual information – can be defined or paraphrased in language, represented in a logical or terminological framework („ontology“)
  - Visual (or other sensory) prototypical information
  - Action schemes (simple motoric actions and action plans)
  - Stereotypical information – valid in the „normal“, default case
- Word meaning: Multi-layered, multi-dimensional and open (no borderline between lexical meaning and world knowledge)

Definition of word meanings

- ... is not a good idea, for several reasons.
- Semantic Decomposition: Defining Word Meaning through
  - Feature sets:
    bachelor --> [+male, +adult, - married]
  - Structured representations, made up of „primitive concepts“:
    kill(X,Y) := CAUSE(X, BECOME(NOT(ALIVE(Y))))
- The alternative: partial specification of word meaning using semantic relations.

Basic Semantic Relations

- Synonymy: car - auto - automobile - machine
- Hypernymy / hyponymy, the sub-/superconcept relation:
  - *car - truck, dog - animal, kill - murder*
- 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

- WordNet provides a network of semantic relations between word senses, with the hyponymy relation as its backbone.
- Word senses are represented as „synsets“: Sets of synonymous words.
- 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.

Synsets + glosses + examples

- **car**
  - { 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"*

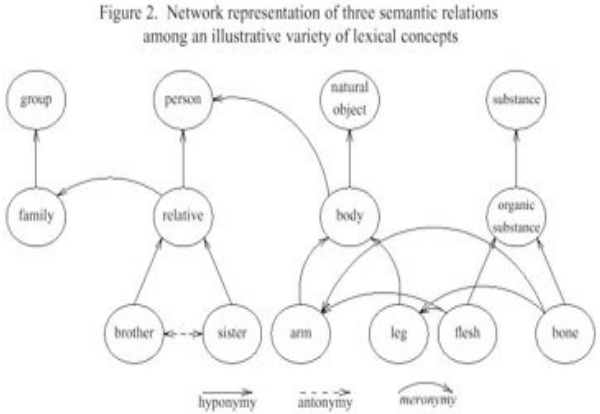
Senses of *car*

- 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**

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)

A small fragment of the WN graph



Basic Uses of WordNet for NLP

- WordNet as a database of axioms feeding logical inference
  - ∀x(family(x)→group(x))
  - ∀x(person(x) → ∃y(substance_m(y,x) ∧ body(y))
  - ∀x(body(x) → ∃y(part_m(y,x) ∧ leg(y))
  - ∀x(body(x) → ∃y(part_m(y,x) ∧ arm(y))
- Query expansion with WordNet synonyms/hyponyms
- Basis for semantic distance measures (normalised path length)

Distributional similarity

- Harris' Distributional Hypothesis:
  „Words that occur in the same contexts tend to have similar meanings"
- Very general, uniform, and inexpensive method to model lexical-semantic relatedness.

Distributional similarity measures

d1: About Dolphins

Dolphins are mammals, not fish. They are warm-blooded like humans, 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.

d2: About Whales

Whales are marine mammals of order Cetacea which are neither dolphins - members, in other words, of the families delphinidae or platanistoidae - nor porpoises. They include the blue whale, the largest animal ever to have lived. Like all mammals, whales breathe air into lungs, are warm-blooded, feed their young milk from mammary glands, and have some (although very little) hair.

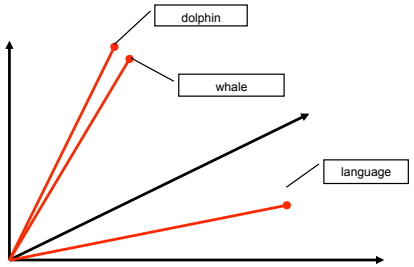
d3: About Language Technology

Applied CL focusses on the practical outcome of modelling human language use. The methods, techniques, tools and applications in this area are often subsumed under the term language engineering or (human) language technology. Although existing CL systems are far from achieving human ability, they have numerous possible applications. The goal is to create software products that have some knowledge of human language.

Distributional similarity measures

	dolphin	human	language	like	technol.	warm-bl.	whale
dolphin	5	1	0	2	0	2	3
human	3	4	5	1	1	1	0
language	0	3	5	0	1	0	0
like	1	1	0	2	0	2	3
technology	0	3	5	0	1	0	0
warm-blooded	5	1	0	2	0	2	3
whale	2	0	0	1	0	1	3

Vector-Space Models of Word Meaning



Cosine Similarity

- Standard measure for similarity between word vectors is cosine again:

$$sim_{\cos ine}(\vec{x},\vec{y}) = \frac{\vec{x}\cdot\vec{y}}{|\vec{x}||\vec{y}|} = \frac{\sum_{i=1}^{n} x_i \times y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

Similarity and association strength

Psycholinguistic word association experiments with human subjects form a third source for information about the semantic lexicon

Example for output:

- purchase: buy 86, clothes 12, money 8, item 6, sale 3, sell 3, acquire 2, car 2, mall 2, price 2, shopping 2, store 2
- sword: knife 47, saber 13, blade 11, sharp 9, ght 7, knight 6, cut 5, shield 5, stab 5, dagger 4, pirate 3, weapon 3, kill 2, point 2, stone 2
- large: small 77, big 24, giant 5, elephant 3, fat 3, huge 3

Distributional similarity

- Very general, uniform, inexpensive, scalable method to model lexical-semantic relatedness.
- Broader, but less specific meaning concept (than the one modelled by WordNet):
  - Paratactic and syntactic relations
  - Logical relations and stereotypical properties
  - Semantic and topical relatedness
  - …
- Similarity is symmetric, entailment is asymmetric.

Predicate-argument structure

- The inverse relation (covered by WordNet's heterogeneous „antonymy" relation) implicitly states a regularity about of the respective concepts involved to their arguments:

taller_than(x,y) ↔ shorter_than(y,x)
parent_of(x,y) ↔ child_of(y,x)
like(x,y) ↔ please(y,x)

Predicate-argument structure

- (Quasi-)Equivalent sentences with different realization of "the same" semantic argument positions.
- Verbs with varying number of explicit argument positions, and alternating syntactic realization of "the same argument".

More complex inversion-like correspondences

Mary gave Peter the book
Peter received the book from Mary

John sold the car to Bill for 3,000€
Bill bought the car from John for 3,000€

Verb alternations

John sells the book.
The book sells for 19.95€.

Mary reads the book
The book reads easy.

Verb alternations

The window broke
A rock broke the window
John broke the window with a rock

$break_3(x,y,z) \models break_2(z,y) \models break_1(y)$

The plane flew to Frankfurt
John flew the plane to Frankfurt
John flew Bill with the plane to Frankfurt

$fly_4(x,y,z,u) \models fly_3(z,y,u) \models fly_2(y,u)$

Thematic Roles (Fillmore 1968)

- Thematic roles describe the conceptual participants in a situation in a generic way, independent from their grammatical realization.
- Thematic roles form a small, closed, and universally applicable inventory conceptual argument types.
- A typical role inventory might consit of the roles: Agent, Theme (Patient, Object),Recipient, Instrument, Source, Goal, Beneficiary, Experiencer

Role Annotation Examples

- *[The window]$_{pat}$ broke*
- *[A rock]$_{inst}$ broke [the window]$_{pat}$*
- *[John]$_{ag}$ broke [the window]$_{pat}$ [with a rock]$_{inst}$*

- *[Peter]$_{ag}$ gave [Mary]$_{rec}$ [the book]$_{pat}$*
- *[Mary]$_{rec}$ received [the book]$_{pat}$ [from Peter]$_{ag}$*

Role linking, example

- Linking information, provided in the lexicon, maps syntactic functions to semantic roles
- An example:

give: SB → *Agent*
OA → Theme
OD → Recipient

receive: *SB → Recipient*
OA → Theme
OP-from → Agent

Thematic Roles

- Allow to represent the semantic correspondence between (uses of) relational concepts in a systematic way – thereby supporting basic lexical-semantic inference.
- Support a systematic representation of the mapping between syntactic complements and semantic argument positions (role-linking).
- Support the systematic description of selectional preferences and constraints (e.g.: Agent is animate, Source and Goal are locations)
- Support the encoding and application of additional inference rules.

The Role Dilemma

- A closed inventory of 8 or 12 or even 20 roles is not sufficient to describe the wealth of predicate-argument relations.

Options:

- Use role names in a more or less arbitrary way, or:
- Assume a much greater role inventory, e.g.: Use different roles for every verb (modulo Alternation)

PropBank

- PropBank: Annotation of Penn TreeBank with predicate-argument structure. Verbs come with individual roles.
- Generalisation over alternation patterns of single verbs (the *break* case).
- No generalisation across lexeme boundaries (the *give/receive* case).
- Efficient annotation process, high inter-annotator agreement

PropBank Example: expect

Roles:

Arg0: expecter
Arg1: thing expected

Example: Transitive, active:

Portfolio managers expect further declines in interest rates.

Arg0:	Portfolio managers
REL:	expect
Arg1:	further declines in interest rates

PropBank example: *give*

Roles:

Arg0: giver
Arg1: thing given
Arg2: entity given to

Example: double object

The executives gave the chefs a standing ovation.

Arg0:	*The executives*
REL:	*gave*
Arg2:	*the chefs*
Arg1:	*a standing ovation*

PropBank: Limitations

- Role assignment is to some part motivated by syntactic structure.
- No cross-lexical generalisations
- No cross-lingual generalisation
- This is illustrated by the following “Trends in argment numbering“, taken from annotators guidelines
  - Arg0 = agent
  - Arg1 = direct object / theme / patient
  - Arg2 = indirect object / benefactive / instrument / attribute / end state
  - Arg3 = start point / benefactive / instrument / attribute
  - Arg4 = end point

(3 Slides taken over from Baker/Hajic/Palmer/Pinkal, ACL 2004)

Relations not covered by PropBank

- Airbus sells five A380 planes to China Southern for 220 million Euro
- China Southern buys five A380 planes from Airbus for 220 million Euro
- Airbus arranged with China Southern for the sale of five A380 planes at a price of 220 million Euro
- Five A380 planes will go for 220 million Euro to China Southern

Frame Semantics

- Structured schemata representing complex prototypical situations, events, and actions are the basic inventory for the conceptual modelling of the world. These are called frames.
- Frames are „evoked“ by NL expressions,typically content words (also called frame-evoking elements (FEEs) or target words).
- Thematic roles are neither universal nor lemma-specific: Role specifications have local validity for the target words of a frame (therefore also called frame elements/ FEs).

- Frame: REQUEST

  Frame Elements: SPEAKER, ADDRESSEE, MESSAGE, MEDIUM, ...

  Lexical Units: *appeal.n, ask.v, beg.v, beseech.v, call.v, command.n, command.v, demand.n, demand.v, entreat.v, entreaty.n, implore.v, invite.v, order.n, order.v, petition.n, plea.n, plead.v, request.n, request.v, suggestion.n, summon.v, tell.v, urge.v*

- Frame: COMMERCE

  Frame Elements: BUYER, SELLER, GOODS, ...

  Lexical Units: *auction.v, retail.v, retailer.n, sale.n, sell.v, vend.v, vendor.n*

Frame-semantic Representation

Airbus sells five A380 planes to China Southern for 220 million Euro

China Southern buys five A380 planes from Airbus for 220 million Euro

Airbus arranged with China Southern for the sale of five A380 planes at a price of 220 million Euro

Five A380 planes will go for 220 million Euro to China Southern

Common frame-semantic representation:

Frame: COMMERCIAL_TRANSACTION
SELLER: Airbus
BUYER: China Southern
GOODS: five A380 superjumbo planes
PRICE: 220 million Euro

The Berkeley FrameNet Database

The FrameNet database consists of:

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

The Berkeley FrameNet Database

- Current release: 700 frames, about 10,000 lexical units (mostly verbs)
- Planned: A total of 15000 verb descriptions
- http://framenet.icsi.berkeley.edu/

Frame-to-Frame Relations