Semantic Theory: Lexical Semantics II

Summer 2007

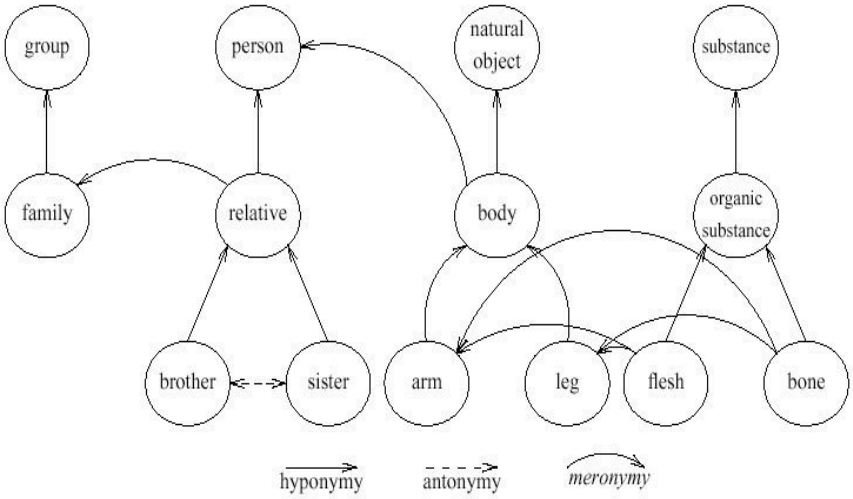
M.Pinkal/ S. Thater

Ontologies

- Hierarchical data structures
- Providing formally rigorous information about concepts and relation
- Within a specific domain (domain ontologies)
- Or concepts and relation of foundational, domain-independent relevance (upper ontologies)

Is WordNet an Ontology?

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



WordNet Relations in FOL

$\forall x(family(x) \rightarrow group(x))$

$\forall x(person(x) \rightarrow \exists y(substance\_m(y,x) \wedge body(y))$

$\forall x(body(x) \rightarrow \exists y(part\_m(y,x) \wedge leg(y))$

$\forall x(body(x) \rightarrow \exists y(part\_m(y,x) \wedge arm(y))$

WordNet Relations in Description Logic

Body ⊑ Natural_object
Relative ⊑ Person
Sister ⊑ Relative
Bone ⊑ Organic_substance

Family ⊑ Group
Brother ⊑ Relative
Flesh ⊑ Organic_substance
Organic_substance ⊑ Substance

Arm ⊑ ∃Substance_m.Flesh
Body ⊑ ∃Part_m.Arm
Person ⊑ ∃Substance_m.Body

Arm ⊑ ∃Substance_m.Bone
Body ⊑ ∃Part_m.Leg
Relative ⊑ ∃Member_m.Family

Description Logic: Terms (1)

Atomic Concepts:

- Concepts A ≈ unary predicates in FOL
- Roles R ≈ binary relations in FOL

Complex concepts:

- conjunction and disjunction of concepts: $C1 \sqcap C2$, $C1 \sqcup C2$
- negation (complementary concept): $\neg C$
- existential restriction: $\exists R.C$
  (set of all a s.t. there is x R(a,x) & C(x): “something that has an R which is a C”)
- value restriction: $\forall R.C$
  (set of all a s.t. for all x s.t. R(a,x), C(x) holds: “something all of whose R’s (if any) are C”)

Description Logic: Terms (2)

- Empty and universal concept: $\bot$, $\top$
- inverse roles R^{-1}
- Number or cardinality restrictions:
  $\exists_{\leq m}R$: Set of all a s.t. there are at most m different *x* for which R(a,x) holds
  $\exists_{\geq m}R$: Set of all a s.t. there are at least m different *x* for which R(a,x) holds

Formulas in Description Logic

1. Axioms or Rules (forming the TBox):
   - Inclusion $C \sqsubseteq D$, $R \sqsubseteq S$
   - Equality $C \equiv D$, $R \equiv S$
   - If the first concept of an equality axiom is atomic, the axiom is called a definition.

2. Assertions (forming the ABox):
   - Assertions: C(a), R(a,b)
     where a, b, c, … are individual constants

An example

A T-BOX

bachelor ≡ ¬ ∃married. ⊤ ⊓ man	*„bachelors are unmarried men“*
married ≡ married^{-1}	(being married to so. is reflexive)
∃ married. ⊤ ⊑ happy	*„all married people are happy“*
$\exists_{\geq 2}$ love ⊑ ⊥	*„you can love at most one person“*
∃ married.woman ⊑ ∃ love.woman	*„someone married to a woman also loves a woman“*

An A-BOX

woman(mary)	man(john)
man(sam)	woman(sue)
loves(john,mary)	loves(mary,sam)
married(sam,sue)	happy(sam)

Facts about Description Logic

- All versions of description logic are FOL fragments.
- Major reasoning tasks in description logic:
  - Subsumption check (Is C sub-concept of D?)
  - Satisfiability check (Are C and D compatible?)
- DL reasoning is much more efficient than FOL deduction.
- There are different versions of description logic, including or exluding, e.g., full term negation, intersection, number restrictions.
- DL reasoners: FaCT, Racer, Protégé, supporting different reasoning tasks for different DL versions.
- Description Logics form the core or backbone of Semantic Markup Languages for the Web (e.g., OWL)

The SUMO Ontology

- SUMO (The Suggested Upper Merged Ontology), is a system of ontologies, including MILO (mid-level ontology) and several omain ontologies, plus SUMO in the narrower sense.
- 20.000 concepts and 80.000 rules with all sub-ontologies
- Size of SUMO itself: 2.600 concepts, 6.000 relations, 2.000 rules

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)

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"

Comparison WordNet - SUMO

- Language-independence of SUMO
  - SUMO-WordNet interface
- Coverage
- Types of information contained:
  - (Full) FOL inference rules in SUMO
  - Something comparable in Extended WordNet

Extended WordNet

- The key idea: exploit the rich information contained in the definitional glosses.
- Intent is to automatically (1) syntactically parse the glosses, (2) transform glosses into logical forms and (3) tag semantically the nouns, verbs, adjectives and adverbs of the glosses.
- Example:
  - Excellent
  - Gloss: „of highest quality“
  - Logical form: excellent(x1) → of(x1,x2)& highest(x2)& quality(x2)
- http://xwn.hlt.utdallas.edu/

Comparison WordNet - SUMO

- Language-independence of SUMO
  - SUMO-WordNet interface
- Coverage
- Types of information contained
  - (Full) FOL inference rules in SUMO
  - Something comparable in Extended WordNet
- Formalisation
  - In Principle, WordNet can be formalised in DL, but:
  - Underspecification of WN relations (e.g., part-of)
  - No consistency control
  - Advantage or disadvantage?

Predicate-argument Structure

– *John is taller than Bill.*
– *Bill is shorter than John.*

taller_than(x,y) ↔ shorter_than(y,x)

Inverse („converse“) relations

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

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

Verb alternations

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

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

Verb alternations

– *The book sells for $39.95*
– *If the book sells, new editions may highlight particularly popular themes on the cover*
– *the book reads easy enough and is mildly entertaining*

Verb alternation pattern (Beth Levin, „Levin Classes“)

More complex lexical relations

– *Mary likes John*
– *John pleases Mary*

– *Mary gave Peter the book*
– *Peter received the book from Mary*

Predicate-argument structure correspondences

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

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.

An Inventory of Thematic Roles

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

Examples Annotated with Thematic Roles

– *[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]$_{ag}$*