



中央研究院  
Academia Sinica



# Ontology for computational linguists

— Introduction —

Chu-Ren Huang and Laurent Prévot  
Institute of Linguistics, Academia Sinica

TIGP, Computational Linguistics and Chinese Language Processing

# Outline for today

- (1) Justification
- (2) Definitions
- (3) Ontological diversity
- (4) Languages for ontologies
- (5) Sum-up and schedule for the class



# Justification

## 1 Fundamental justifications

- Interaction between ontology and NL semantics
- Importance of ontology in linguistics
- Importance of linguistics for ontology

## 2 Practical justifications

- Ontologies and lexical resources
- Semantic Web
- Human computer interaction

# Interaction between ontology and NL semantics

## Ontology:

- Ontology is about **what there is**.
- Ontology is about general categories of entities.
- Ontology studies what are persons, processes, abstract objects...

## NL semantics:

- NL semantics is about **referring** to what there is.
- NL semantics studies how NL language refers to persons, objects, abstract object...

## Dependence recognized

- Ontological considerations in NL semantics
- NL semantics used as clues for ontological analysis



Big question: Which ontology for NL?

- Should every common noun correspond to an ontological category?

### Example

laptop vs. notebook

- Should ordinary NL language sentences be systematically reformulated into "ontologically" valid statements?

### Example

Ordinary language: There is a hole in that piece of cheese.

"Valid paraphrase": That piece of cheese is perforated.

Answers: Depends on your philosophical approach.

## The rich ontology of natural language

Cannot assume a simple physical ontology of what exists.  
Natural language can handle sophisticated entities.

- Co-located events

### Example

Nicola is singing while looking for mushrooms.

- Co-located entities

### Example

- There is a fly on the nose of this statue.
- ? There is a fly on the nose of this piece of marble.
- There is a fly on this piece of marble.

## Research of adequate ontological theories for linguistic problems

- Universals vs. particulars, abstract propositions [Quine, 1960]
- Time (Tense and Aspect): Events, Intervals, Instants [Davidson, 1967, Dowty, 1977]
- Mass Terms, Plurals, Generics: mereologic structures [Quine, 1960, Link, 1983]
- Plurals and Collections: mereologic structures [Link, 1983]





## Ontologies and lexical resources

- Improvement of existing resources [Gangemi et al., 2003a]
  - Identifying ontological problems (*e.g hyperonymy relation*)
  - Re-structuring the resource
- Multilingual lexical resources (EuroWordNet [Vossen, 1998] , SinicaBOW [Huang et al., 2004], NEDO Project)

## Semantic Web

Ontologies support semantic interoperability

- Ontologies as **shared understandings of a domain**  
(between systems, applications, artificial agents)
- Necessary for overcoming the terminological differences.

Ontological structuring helps for information retrieval  
[Welty et al., 2003]

- More accurate searches
- Easier to extend searches in case of failure

# Human computer interaction

## Ontology based dialogue systems

- Interpretation (Word-Sense Disambiguation)
- Models for sentence generation (Bateman)
- Dialogue management [Milward and Beveridge, 2003]
  - Use of taxonomy structure for handling question and answers (specificity)
  - Use of meronymy structure (accuracy)
  - Issue-based dialogue management for improving coherence





## Lexical resources and NLP for ontologies

Semi-automatic ontology creation from corpus

- Text mining: terms and relations extraction
- Tagging, Parsing, Disambiguation
- Term grouping

Lexical Resources for:

- Checking ontologies
- Deepening the coverage of upper/mid level ontologies  
[Niles and Pease, 2003]

Axiomatizing lexical resources gloss (NLP techniques)  
([Harabagiu et al., 1999, Gangemi et al., 2003b])

# Definitions

# Definitions

- 3 Ontology vs. ontologies
- 4 Ontology
- 5 Conceptualization

# Ontology vs. ontologies

- **Ontology**
  - Philosophical discipline
  - The study of "what there is"
  - The study of nature and structure of possible entities
- **ontology**
  - Theoretical or computational artifact
  - Created for expressing the intended meaning of a vocabulary



An ontology (among others):

- is an explicit and formal specification of a conceptualization. [Gruber, 1995]
- is designed to account for the commitment of a vocabulary to a conceptualization.
- constrains the interpretation of a vocabulary.

## Examples

- *An ontology of "Lord of the Rings" specifies and makes explicit my understanding of Tolkien "world".*
- *An ontology of digital cameras, an ontology of meteorology...*

A conceptualization is an abstraction over the reality performed by an agent.

## The components of an ontology

An ontology is a **shared vocabulary** (arbitrary) plus a formal characterization of its **intended meaning**:

- Vocabulary: concepts (categories) and relations
- Characterization of meaning: axioms (logical formulas)

No axioms, no meaning.

Glosses and lexical relations may play the roles of axioms in linguistic ontologies, at the price of the formal rigor.

## Example

- Vocabulary: Bird, White, Wing, Leg, Fish, Animal...
- Characterization (informal version):
  - A bird is an animal. (relational)
  - Birds have wings. (intrinsic)
  - Fish are not birds.

## The content of an ontology

An ontology is a structure of **necessarily true knowledge** (or meant to be)

### Example

Birds have wings.

An ontology do **not** includes contingent situations or assertional data.

### Example

This bird is white.

Because an ontology specifies a **conceptualization**

## What is a conceptualization?


A formal structure for a piece of reality (not necessarily the true one) as perceived and organized by an agent, independently of:

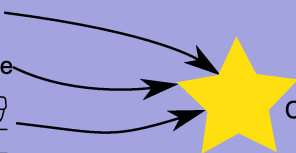
- the vocabulary used
- the actual occurrence of a specific situation

Different situations involving same objects, described by different vocabularies may share the same conceptualization.

V1: Star

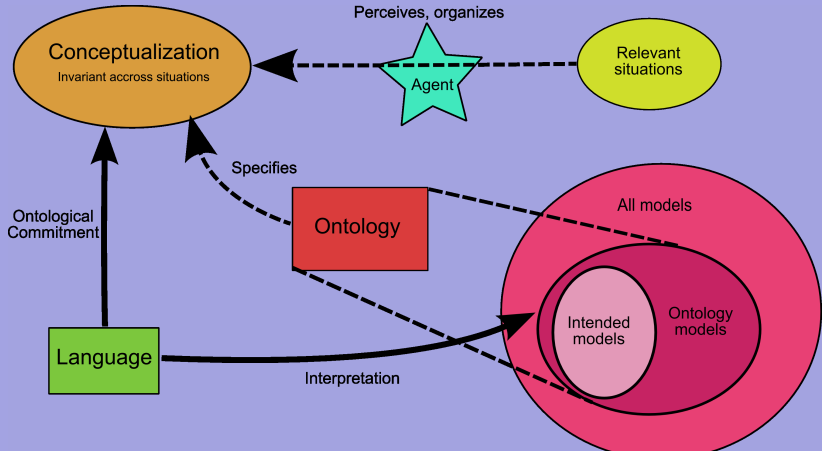
V2: Etoile

V3: 



Conceptualization

## Conceptualization: intended meaning



## Sum-up

An ontology is a

- shared vocabulary
- + definition
  - relation with other entities
  - essential properties

# Ontological diversity



## Different Approaches

- Linguistic "ontologies"
  - Glossary/Lexicon
  - Controlled vocabulary
  - Taxonomy
  - Thesaurus
- Computer Science "ontologies"
  - Conceptual Schema
  - Knowledge base
- Formal ontologies
  - Domain ontology
  - Core (reference) ontology
  - Foundational ontology

# Linguistic ontologies

- Word list
- Glossary: list of NL definitions
- Taxonomy: classification with a **is-a** or **is-a-kind-of** relation
- Thesaurus: Taxonomy + synonyms + related terms (generally as part as a standardization)

Important characteristics of linguistic ontologies, generally:

- Each item is generally lexicalized
- The relations and constraints are lexical ones (synonymy, antonymy, hyperonymy)

# Knowledge bases

Knowledge base:

- Mainly assertions
- Reflects specific states of affair
- Use descriptors provided by the ontology

Ontology:

- Mainly definitions
- Essential properties of a domain
- Independent of particular state of affairs
- Provides vocabulary and structure for the KB
- Ontological formulas are **necessarily true**

## Conceptual schemas

- Motivation similar to ontologies
- BUT traditionnally ad-hoc solutions (old style computer science)
- Influenced by the extension of the ontology/scheme in development framework (Programming)
- Getting ontologically better

For a in depth disscussion see [Guizzardi, 2005]

## Formal ontologies

An ontology is "*a specification of a conceptualization.*"

Formal ontology requires in more:

- clear semantics for the language (formal semantics)
- clear motivations for the adopted distinctions (philosophical analysis)
- strict rules on how to specify terms and relationships (formal language)

## Approaches sum-up

- Different motivations brought different kind of ontologies
- In the course we will be mainly interested in:
  - linguistic ontologies
  - formal ontologies
  - their interface

## Ontological precision: Examples, Wordlist

- football
- ping-pong
- indoor game
- ball
- chess
- go
- board game
- athletic game
- player

## Glossary

- athletic game: game requiring physical effort.
- board game: game played on a board.
- chess: board game for two players.
- football: game played with a ball on a rectangular field in which two teams try to kick a ball into each other's goal
- ...

# Taxonomy

- Game
  - athletic game
    - Football
    - Tennis
    - Rugby
  - board game
    - Go
    - Chess
- Sport equipment
  - paddle
  - ball

# Thesaurus

chess: a game for two players who move their 16 pieces according to specific rules; the object is to checkmate the opponent's king

- **synonyms: chess game**
- **related terms:**
  - chess move - the act of moving a chess piece
  - exchange - (chess) the capture by both players of pieces of equal value
  - bishop - (chess) a piece that can be moved diagonally over unoccupied squares of the same color
  - shogi - a form of chess played on a board of 81 squares; each player has 20 pieces
- **antonyms: none**

# Wordnet

- synset: chess, chess game
- gloss: a board game for two players [...]; the object is to checkmate the opponent's king
- Coordinate terms: [...] => backgammon – (a board game for two players; pieces move according to throws of the dice) [...]
- Hypernyms: chess, chess game => board game => parlor game, parlour game => game => activity [...]
- Hyponyms: shogi – (a form of chess played on a board of 81 squares; each player has 20 pieces)
- Topic terms: (noun) check#13, opening#12, chess opening#1, bishop#3 (verb) develop#18, promote#5

## Axiomatic theory

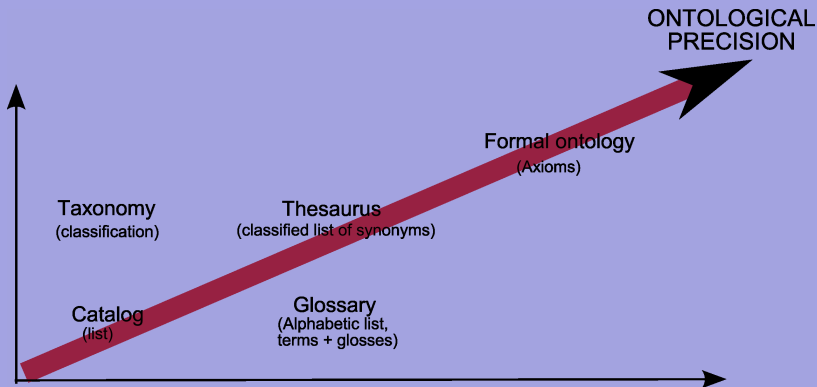
$$\textit{game}(x) \rightarrow \textit{activity}(x)$$

$$\textit{athleticGame}(x) \rightarrow \textit{game}(x)$$

$$\textit{boardGame}(x) \leftrightarrow \textit{game}(x) \wedge \exists y \textit{board}(y) \wedge \textit{playedOn}(y, x)$$

$$\textit{indoorGame}(x) \leftrightarrow \textit{game}(x) \wedge \exists y \textit{building}(y) \wedge \textit{playedIn}(y, x)$$

## Ontological precision



## Domain and coverage:

- Foundational ontologies
- Core ontologies
- Domain ontologies

# Foundational ontologies and upper levels

Foundational vs. upper level:

- Upper level is about the more general categories
- Foundational ontologies requires a rich characterization

Their content:

- They characterize **general terms**
- *e.g entity, event, process, spatial location*
- and **basic relations**
- *e.g part-of, quality-of, participation, constitution*

Their purpose:

- Provide a formal description of entities/relationships common to all domains
- Provide a consistent and unifying view

# What a foundational ontologies can say about ping-pong?

Almost nothing! Basic entities concerned:

- NON-AGENTIVE-PHYSICAL-OBJECT
- AGENTIVE-PHYSICAL-OBJECT
- PROCESS

But very fundamental things:

PHYSICAL-OBJECT PARTICIPATES-IN PROCESS

## Foundational ontologies and upper levels

Their domain:

- Everything
- Common to all domains

Their coverage (the number of terms it includes):

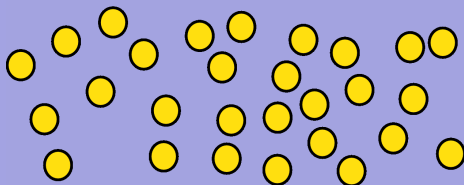
- Extremely broad but shallow or even null
- Talk about only very general things
- Have to be mapped to get some coverage

DOLCE: Roughly 40 categories, 100 relations, 80 axioms

## Coverage foundational ontologies



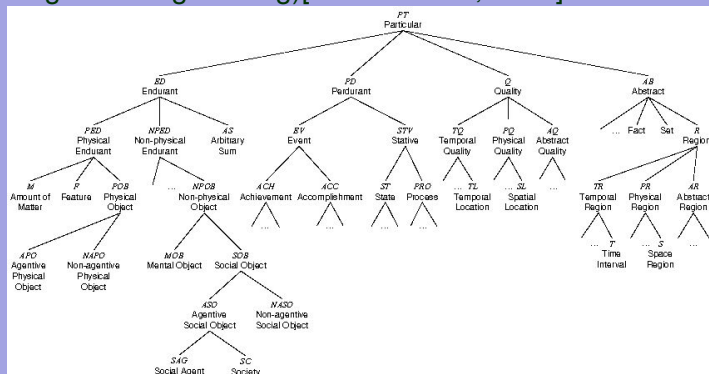
Top-level



Terms

## Example, DOLCE

DOLCE taxonomy(Descriptive Ontology for Linguistic and Cognitive Engineering)[Masolo et al., 2003]



## Domain ontologies

- All ontological information relevant for a given domain
- Provides a deep technical vocabulary

Their domain:

- only one
- usually not a broad one

Their coverage:

- Narrow coverage but very deep
- Ideally no gap within the domain
- Covers all the terms (and specific usage) of the domain
- It is the maximal coverage.

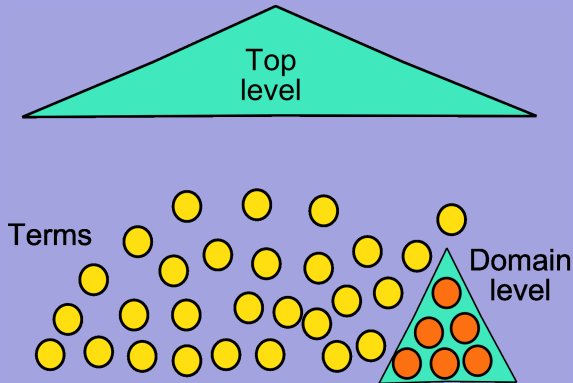
# What foundtational ontology of ping-pong can say about ping-pong?

A lots of things!

- All kind of paddles: maybe the different brands and models
- All kind of ball: 1STARBALL, 2STARBALL,...
- All kind of strokes: TOPSPIN-STROKE, LIFT-STROKE

But unrelated to general things.

## Coverage domain ontologies

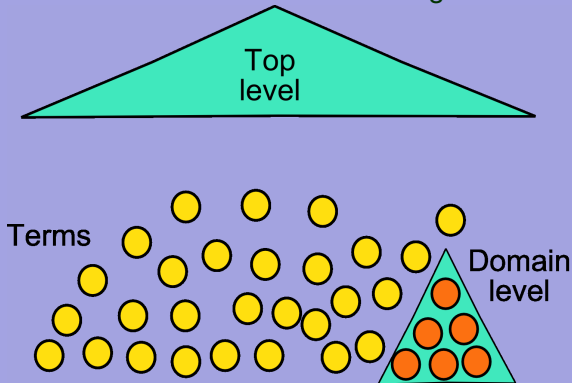


## Example

- factkb:EosinophilCountProcedure
- factkb:ErosionOfStomach
- factkb:ErosiveProcess
- factkb:ErythrocyteCount
- factkb:ErythrocyteCountProcedure
- factkb:ErythrocyteSedimentation
- factkb:ErythrocyteSedimentationRate
- factkb:EsophagealPathology
- factkb:EsophagitisProcess
- factkb:EssentialHypertension
- factkb:ExactlyPairedBodyStructure
- factkb:ExcretionOfUrine
- factkb:FacultativeAerobicMetabolicProcess
- factkb:FacultativeAnaerobicMetabolicProcess
- factkb:FailedAnaesthetic
- factkb:FailureOfCellGlucoseUptake
- factkb:FailureOfCellUptakeOfBloodGlucoseDueToCellInsulinResistance
- factkb:FailureOfCellUptakeOfBloodGlucoseDueToInsulinDeficiency
- factkb:FattyDegeneration
- factkb:FattyDegenerationOfHeart

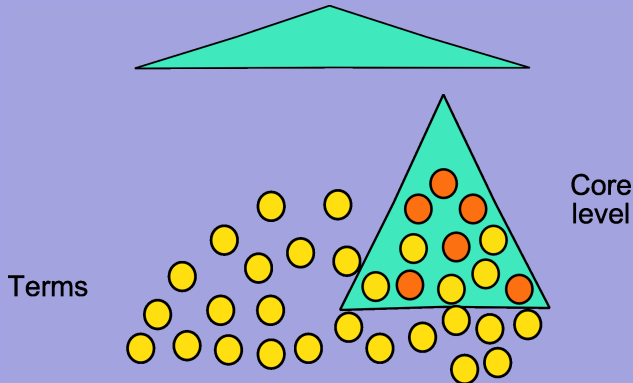
## Core ontologies

Domain/foundational distance might be to big.



## Core ontologies

Need an intermediate level: **Core** or **reference ontologies**.



## Core ontologies

Core ontologies include:

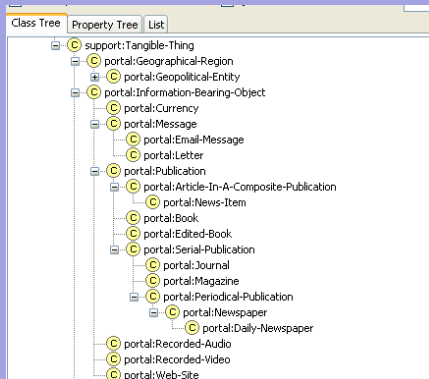
- The central concepts and relations of a domain
- Without entering into all the technical details
- May include rich conceptualization

Their coverage is balanced in scope and in depth.

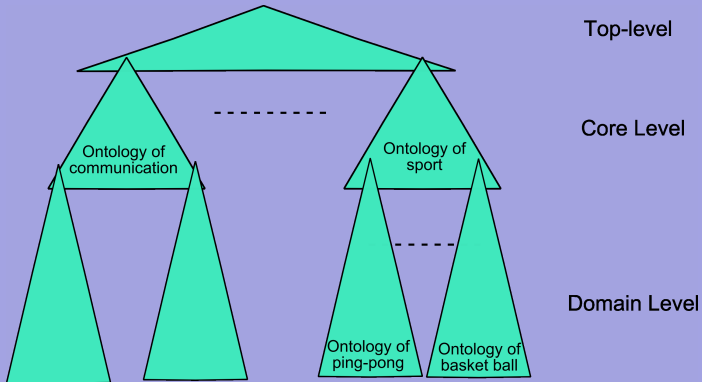
## What a core ontology of sports can say about ping-pong?

- PLAYERS EXECUTE STROKES
- The SPORTEQUIPMENT used is PINGPONG-PADDLE and PINGPONG-BALL
- PLAYERS might be somehow related to AGENTIVE-PHYSICAL-OBJECT
- SPORTEQUIPMENT will be related to NON-AGENTIVE-PHYSICAL-OBJECT
- STROKES will be related to the TOPSPINSTROKE, LIFTSTROKE
- PINGPONG-BALL will be related to 1STARBALL,...

## Example



## Illustration, recap



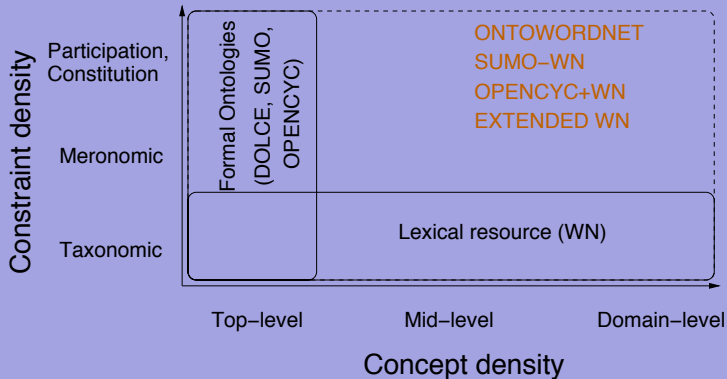
## Light-weight vs. Heavy weight

- Light-weight/on-line/shared ontologies [Guarino, 1998]
  - Not a lot of constraints/relations
  - (*e.g simple taxonomies*)
  - easy to handle by programs
- Heavy-weight/off-line/reference ontologies [Guarino, 1998]
  - Rich constraints system
  - (*e.g Meronomic, Participation, Dependence, Constitution, Causal relations*)
  - Much more difficult and costly to create
  - More complex to handle (computational issue)

## Light-weight vs. Heavy weight

- **Concept density: number of concepts or terms for a given coverage**
  - Low concept density: Upper levels, foundational ontologies
  - High concept density: Lexical resources (WN)
- **Constraint density: the number or relations/constraints for a given coverage**
  - Low constraint density: Taxonomies, catalogues, light-weight ontologies
  - High constraint density: Foundational and heavy-weight ontologies,

## Constraint and Concept densities





# Languages for ontologies



# Languages for ontologies

## Languages used

- Natural languages: philosophy sources and annotations
- First-Order Logic: Foundational works (foundational ontology)
- Description logic: More applied than FOL
- Semantic Web languages (XML, RDF, OWL): Current development of ontologies for semantic web

## Examples: Natural language

*The main characteristic of physical objects is that they are endurants with unity. However, they have no common unity criterion, since different subtypes of objects may have different unity criteria. Differently from aggregates, (most) physical objects change some of their parts while keeping their identity, they can have therefore temporary parts. (...)*

## Examples: First order logic

### 4.3.3 Constitution

#### *Argument restrictions*

$$(Ad20) \quad K(x, y, t) \rightarrow ((ED(x) \vee PD(x)) \wedge (ED(y) \vee PD(y)) \wedge T(t))$$

$$(Ad21) \quad K(x, y, t) \rightarrow (PED(x) \leftrightarrow PED(y))$$

$$(Ad22) \quad K(x, y, t) \rightarrow (NPED(x) \leftrightarrow NPED(y))$$

$$(Ad23) \quad K(x, y, t) \rightarrow (PD(x) \leftrightarrow PD(y))$$

#### *Ground Axioms*

$$(Ad24) \quad K(x, y, t) \rightarrow \neg K(y, x, t)$$

$$(Ad25) \quad (K(x, y, t) \wedge K(y, z, t)) \rightarrow K(x, z, t)$$

#### *Links with other Primitives*

$$(Ad26) \quad K(x, y, t) \rightarrow (PRE(x, t) \wedge PRE(y, t))$$

$$(Ad27) \quad K(x, y, t) \leftrightarrow \forall t' (P(t', t) \rightarrow K(x, y, t'))$$

$$(Ad28) \quad (K(x, y, t) \wedge PED(x)) \rightarrow x \approx_S < y, t >$$

$$(Ad29) \quad (K(x, y, t) \wedge P(y', y, t)) \rightarrow \exists x' (P(x', x, t) \wedge K(x', y', t))$$

## Examples: KIF

```
(subclass LandArea GeographicArea)\\  
(documentation LandArea "An area which is predominantly solid ground,  
e.g. a &\%Nation, a mountain, a desert, etc. [...])\  
(=>  
  (instance ?LAND1 LandArea)  
  (exists (?LAND2)  
    (and  
      (part ?LAND1 ?LAND2)  
      (or  
        (instance ?LAND2 Continent)  
        (instance ?LAND2 Island))))))
```

## Examples: OWL

(Web Ontology Language)

```
<owl:Class rdf:about="#physical-object">
  <rdfs:comment rdf:datatype="xsd:string">The main characteristic
  <rdfs:subClassOf>
    <owl:Class rdf:about="#physical-endurant">
      </owl:Class>
    </rdfs:subClassOf>
  </owl:Class>
<owl:Class rdf:about="#feature">
  <owl:disjointWith>
    <owl:Class rdf:about="#physical-object">
      </owl:Class>
    </owl:disjointWith>
  </owl:Class>
```

## Trade off between expressivity and decidability

- First-Order Logic: expressive but undecidable
- Description logics: Decidable fragments of FOL, KR oriented, closer to SW languages
- Semantic Web languages: Trade off between efficiency and expressivity
  - OWL-Lite: not expressive but "Lite"
  - OWL-DL: Restricted to a FOL fragment (decidable)
  - OWL-Full: Expressive but undecidable

## Trade off between expressivity and decidability

- An ontology expressed in a given language can be translated in another one **IF** expressivity allows.
- Many important characterizations are lost when foundational ontologies are translated in less expressive language such as OWL-Lite.

# Conclusion



## Sum-up

- Ontology vs. ontologies
- An ontology (among others) is:
  - a shared vocabulary + its characterization
- Different approaches to ontologies (Formal, Linguistic, IS)
- Different level for ontologies (Upper, Core, Domain)
- Different languages for ontologies (NL, FOL, SW)

Which approach, at which level and in which language an ontology should be chosen/developed is dictated by the main motivation for using/building the ontology.

## Future courses

- (1) Levels in ontologies, specially upper levels and foundational ontologies
- (2) Introduction to the ontology-lexical resource interface
- (3) Language and tools for ontologies
- (4) Methods and tools for building ontologies
- (5) Methods and tools for combining ontologies and lexical resources
- (6) Maintenance and improvement methodologies
- (7) Using an ontology for improving multilinguality resources
- (8) Ontology and formal semantics
- (9) Application: Ontology evaluation for NLP tasks
- (10) Application: Ontology and the Semantic Web

## Readings

For this lesson:

- Introduction, definitions: **[Guarino, 1998]**
- Philosophical background: [Quine, 1958, Quine, 1960], "Logic and Ontology" on Stanford Encyclopedia of Philosophy.
- Ontology and NL semantics: [Dölling, 1993, Carrara and Varzi, 2001]
- Ontology and NLP: [Poesio, 2005], [Bateman, 1997]
- Ontology and Information Science: [Guizzardi, 2005]

For next lesson (Foundational ontologies and upper levels):

**[Masolo et al., 2003], [Niles and Pease, 2001]**



# References I



Bach, E. (1986).

**The algebra of events.**

*Linguistics and Philosophy*, 9:5–16.



Bateman, J. (1997).

**The theoretical status of ontologies in natural language processing.**

In *Text Representation and Domain Modelling – ideas from linguistics and AI.*, pages 50–99. Technische Universität Berlin.

Also available from the Computation and Language E-print archive: [cmp-ig/9704010](http://cmp-ig/9704010).



Carrara, M. and Varzi, A. (2001).

**Ontological commitment and reconstructivism.**

*Erkenntnis*, 55:33–50.



Davidson, D. (1967).

**The logical form of action sentences.**

In Rescher, N., editor, *Logic of Decision and Action*. Pittsburgh UP.



Dölling, J. (1993).

**Commonsense ontology and semantics of natural language.**

*Zeitschrift für Sprachtypologie und Universalienforschung*, 46(2):133–141.

## References II



Dowty, D. R. (1977).

Toward a semantic analysis of verb aspect and the english *is*quoimperfective*s*quo progressive.  
*Linguistics and Philosophy*, 1(1):45–77.



Gangemi, A., Guarino, N., Masolo, C., and Oltramari, A. (2003a).

Sweetening WordNet with DOLCE.  
*AI Magazine*, 3(24):13–24.



Gangemi, A., Navigli, R., and Velardi, P. (2003b).

The ontowordnet project: extension and axiomatisation of conceptual relations in wordnet.  
In *International Conference on Ontologies, Databases and Applications of SEMantics (ODBASE)*, Catania, (Italy).



Gruber, T. R. (1995).

Toward principles for the design of ontologies used for knowledge sharing.  
*International Journal of Human and Computer Studies*, 43(5-6).



Guarino, N. (1998).

Formal ontology in information systems.  
In Press, I., editor, *Proceedings of FOIS'98*, pages 3–15, Trento, Italy.

## References III



Guizzardi, G. (2005).  
*Ontological Foundations for Structural Conceptual Models*.  
Telematica Instituut Fundamental Research Series.



Harabagiu, S. M., Miller, G. A., and Moldovan, D. I. (1999).  
Wordnet 2 - a morphologically and semantically enhanced resource.  
In *SIGLEX 1999*.



Hofweber, T. (2004).  
Logic and ontology.  
<http://plato.stanford.edu/entries/logic-ontology/>.  
Stanford Encyclopedia of Philosophy.



Huang, C.-R., Chang, R.-Y., and Lee, S.-B. (2004).  
Sinica BOW (bilingual ontological wordnet): Integration of bilingual WordNet and SUMO.  
In *4th International Conference on Language Resources and Evaluation (LREC2004)*, Lisbon.



Kripke, S. A. (1982).  
*Wittgenstein on rules and private language: an elementary exposition*.  
Harvard University Press.

## References IV



Link, G. (1983).

**The logical analysis of plurals and mass terms: A lattice-theoretical approach.**  
*In Meaning, Use and Interpretation of Language*, pages 302–323. de Gruyter.



Masolo, C., Borgo, S., Gangemi, A., Guarino, N., and Oltramari, A. (2003).

**Wonderweb deliverable18, ontology library (final).**  
Technical report, LOA-ISTC, CNR.



Milward, D. and Beveridge, M. (2003).

**Ontology-based dialogue systems.**  
*In Workshop IJCAI on Knowledge and Reasoning in Practical Dialogue Systems.*



Niles, I. and Pease, A. (2001).

**Towards a standard upper ontology.**  
*In Welty, C. and Smith, B., editors, Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, Ogunquit, Maine.



Niles, I. and Pease, A. (2003).

**Linking lexicons and ontologies: Mapping wordnet to the suggested upper merged ontology.**  
*In Proceedings of the 2003 International Conference on Information and Knowledge Engineering (IKE 03)*, Las Vegas, Nevada.

## References V



Poesio, M. (2005).

Domain modelling and nlp: Formal ontologies? lexica? or a bit of both?  
*Applied Ontology*, 1(1):27–33.



Quine, W. O. (1958).

On what there is.  
*Review of metaphysics*, 2(1):21–28.  
Reprinted many times.



Quine, W. O. (1960).

*Word and Object*.  
MIT Press.



Strawson, P. F. (1959).

*Individuals. An Essay in Descriptive Metaphysics*.  
Routledge.



Talmy, L. (2000).

*Toward a cognitive semantics*.  
MIT.



## Acknowledgements

- Laure Vieu, Nicola Guarino, Stefano Borgo for their foundational courses and all the members of LOA for hours of ontological discussion.
- Nathalie Aussenac for kindly providing me her teaching material about semi-automatic tools and methodologies for ontology creation and management.