Laboratory for Applied Ontology

Institute of Cognitive Science and Technology
Italian National Research Council

Ontology and Terminology

how can formal ontology help concept modeling and terminology?

Nicola Guarino

Laboratory for Applied Ontology (LOA)

Institute for Cognitive Sciences and Technology,

National Research Council

Trento, Italy

Thanks to all LOA people!

www.loa-cnr.it

Summary

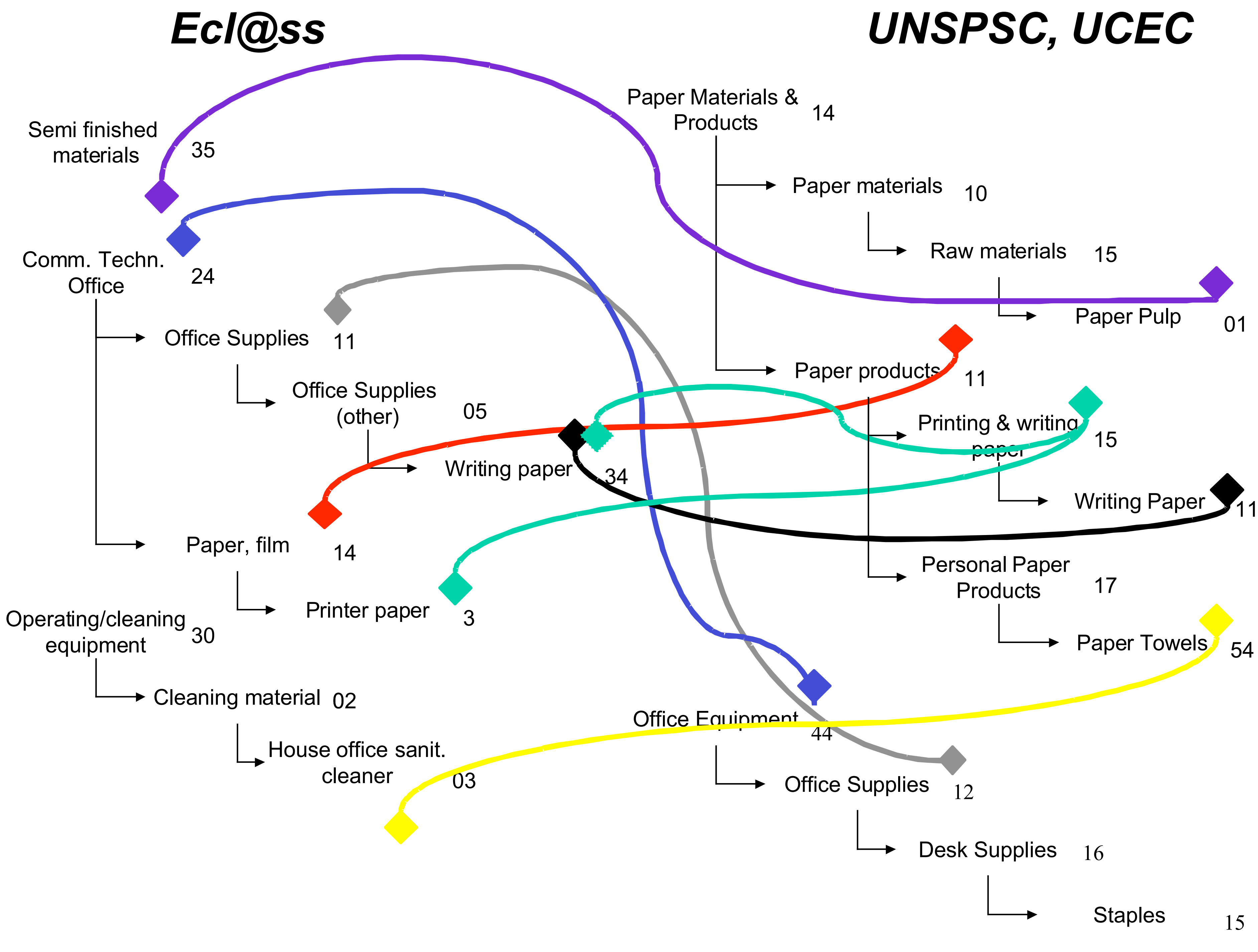
1. **What is an ontology**
2. **Ontology and natural language semantics**
3. **Ontology: an overloaded term**
4. **The role of Formal Ontology**
5. **A glimpse to OntoClean, DOLCE, OntoWordnet**

- **Research activities at LOA**
- **A new journal: Applied Ontology** (www. applied-ontology.org)
- **An interdisciplinary conference: FOIS** (www.formalontology.org)

The importance of *subtle distinctions*

"Trying to engage with ***too many partners too fast*** is one of the main reasons that ***so many online market makers have foundered***. The transactions they had viewed as simple and routine actually involved many ***subtle distinctions in terminology and meaning***"

Harvard Business Review, October 2001



Ecl@ss
UNSPSC, UCEC
Semi finished materials 35
Comm. Techn. Office 24
Office Supplies 11
Office Supplies (other) 05
Writing paper 34
Paper, film 14
Printer paper 3
Operating/cleaning equipment 30
Cleaning material 02
House office sanit. cleaner 03
Paper Materials & Products 14
Paper materials 10
Raw materials 15
Paper Pulp 01
Paper products 11
Printing & writing paper 15
Writing Paper 11
Personal Paper Products 17
Paper Towels 54
Office Equipment 44
Office Supplies 12
Desk Supplies 16
Staples 15

Where subtle distinctions in meaning are important

- 2000 US Presidential elections: is there a ***hole***?
- 2001 twin towers catastrophe: how many ***events***?

…only ***ontological analysis*** solves these problems!!

A common alphabet is not enough…

- “XML is only the first step to ensuring that computers can communicate freely. ***XML is an alphabet for computers*** and as everyone who travels in Europe knows, knowing the alphabet doesn’t mean you can speak Italian or French”

Business Week, March 18, 2002

Standard glossaries can help, but...

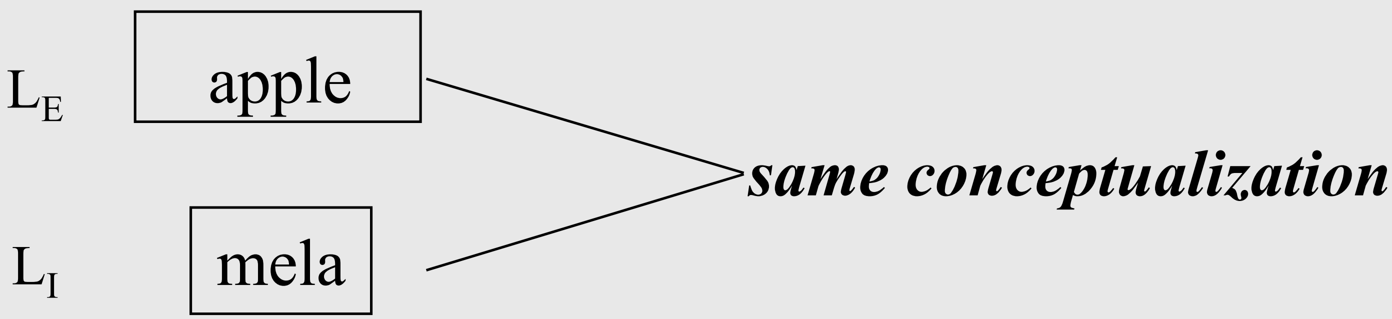
- Defining standard vocabularies is ***difficult and time-consuming***
- Once defined, standards ***don't adapt well***
- Heterogeneous domains need a ***broad-coverage vocabulary***
- People don't implement standards correctly anyway
- Vocabulary definitions are often ***ambiguous or circular***

Ontology and Ontologies

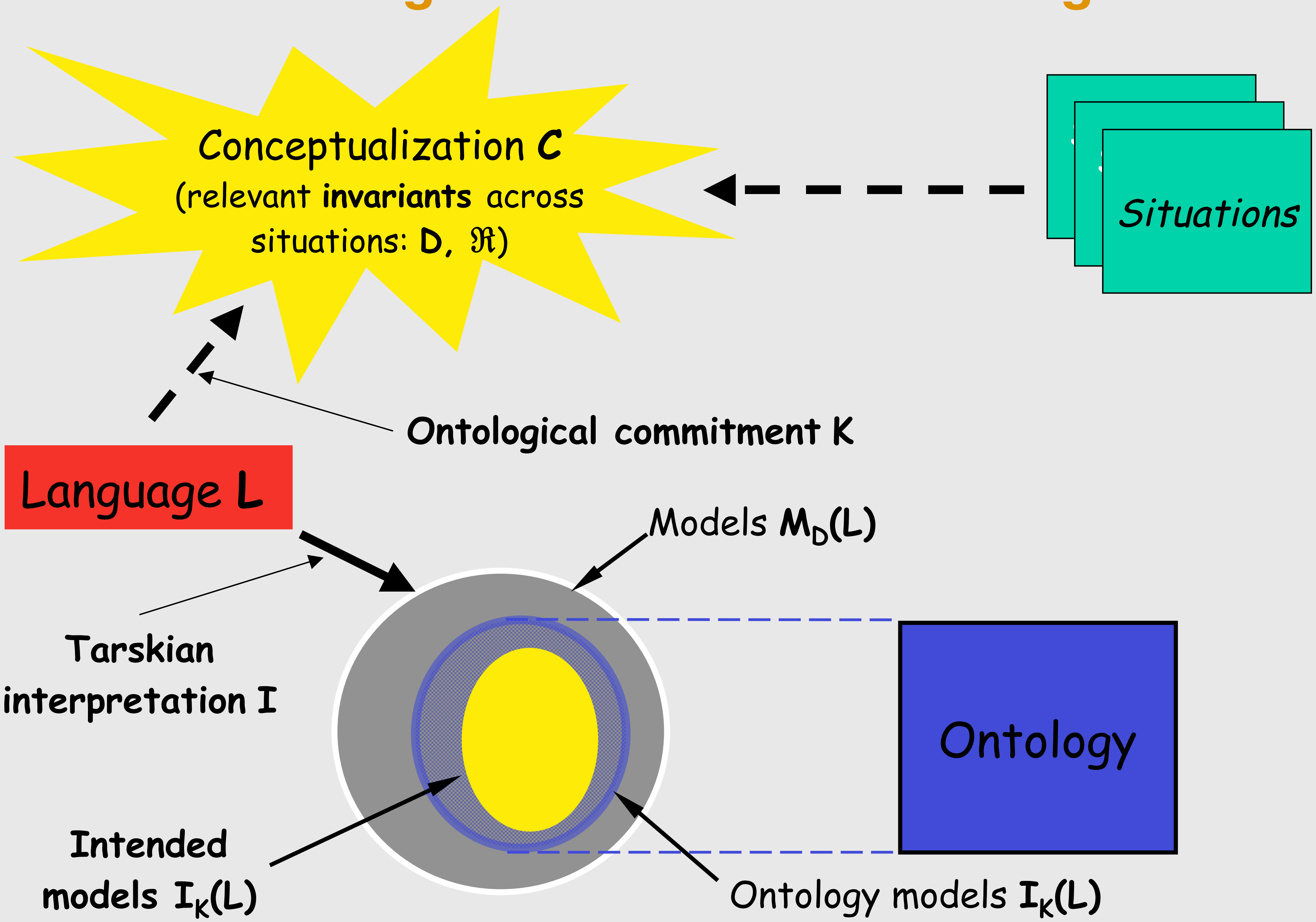
- Ontology: the philosophical discipline
  - Study of what there (possibly) is
  - Study of the nature and structure of reality
    - Domain of entities
    - Categories and relations
    - Characterizing properties
- *An* ontology: a theoretical or computational artifact
  - "An explicit and formal specification of a conceptualization" (Gruber)
  - A specific artifact expressing the intended meaning of a vocabulary in terms of the nature and structure of the entities it refers to

What is a conceptualization

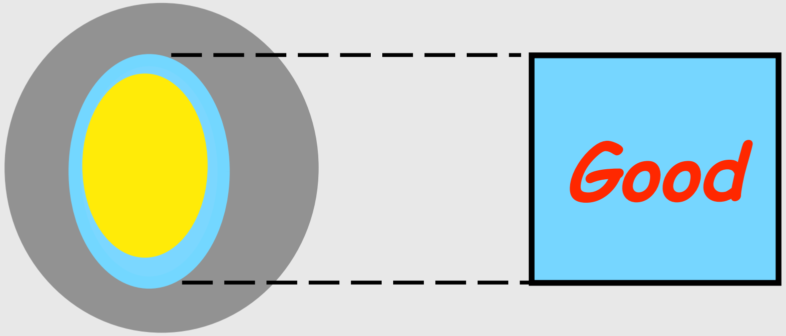
- Formal structure of (a piece of) reality *as perceived and organized by an agent*, *independently* of:
  - the ***vocabulary*** used
  - the actual occurence of a specific ***situation***
- Different situations involving same objects, described by different vocabularies, may share the same conceptualization.



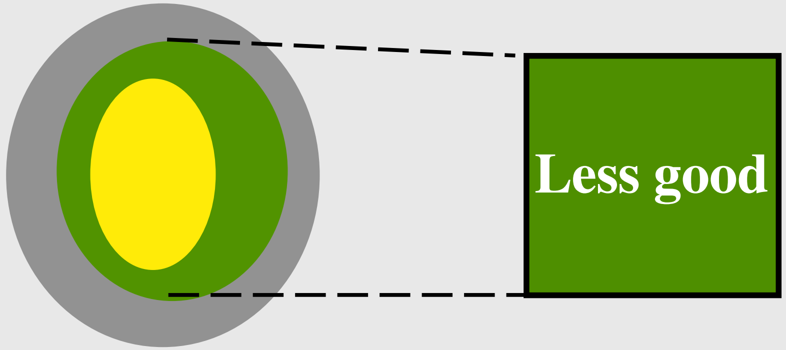
Ontologies and intended meaning



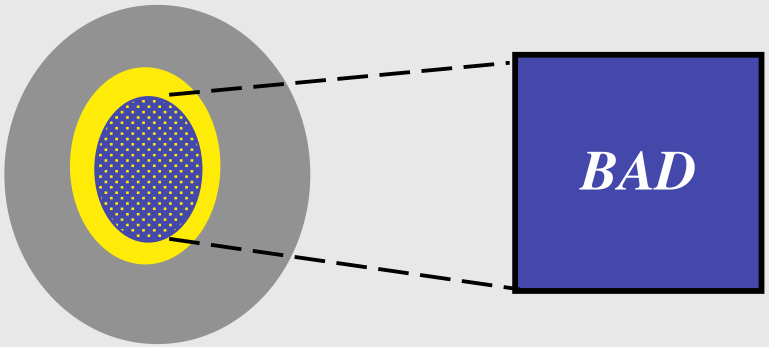
Ontology Quality: Precision and Coverage



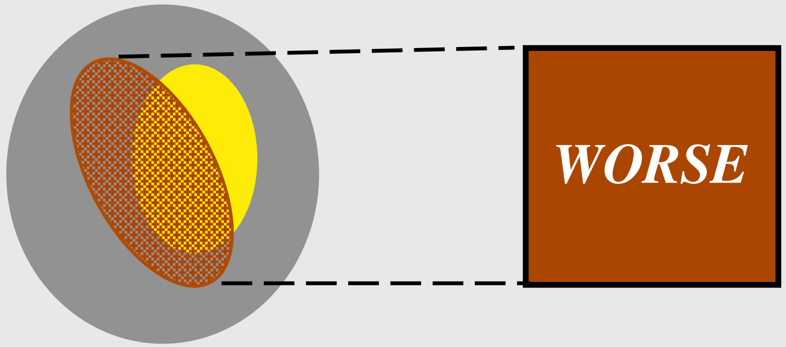
High precision, max coverage



Low precision, max coverage

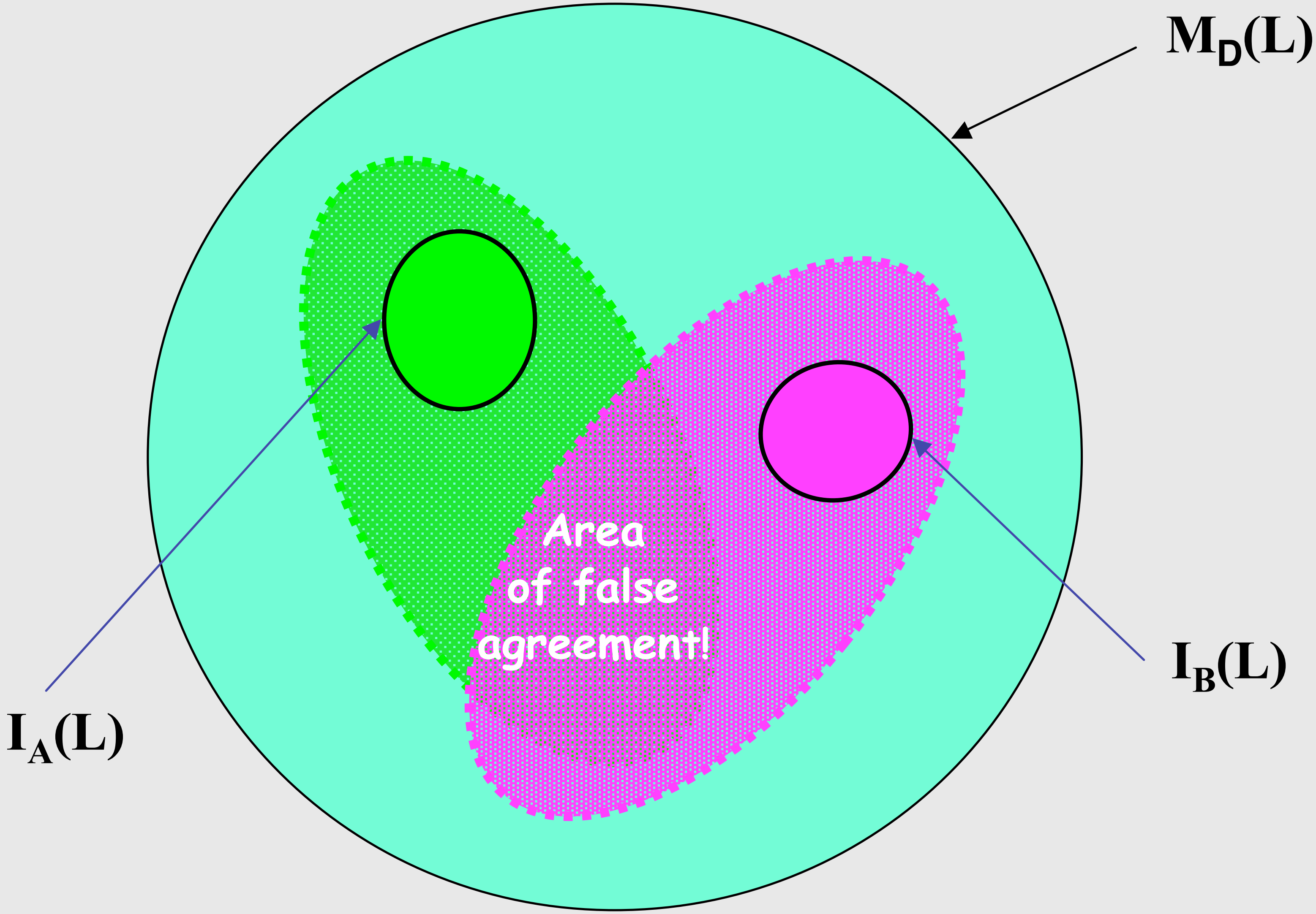


Max precision, limited coverage



Low precision, limited coverage

Why precision is important



Levels of Ontological Precision

tennis
football
game
field game
court game
athletic game
outdoor game

Catalog

Glossary

game
- athletic game
  - court game
    - tennis
  - outdoor game
    - field game
      - football

Taxonomy

game
NT athletic game
  NT court game
    RT court
    NT tennis
      RT double fault

Thesaurus

DB/OO scheme

$\text{game}(x) \rightarrow \text{activity}(x)$
$\text{athletic game}(x) \rightarrow \text{game}(x)$
$\text{court game}(x) \leftrightarrow \text{athletic game}(x) \wedge \exists y.\ \text{played\_in}(x,y) \wedge \text{court}(y)$
$\text{tennis}(x) \rightarrow \text{court game}(x)$
$\text{double fault}(x) \rightarrow \text{fault}(x) \wedge \exists y.\ \text{part\_of}(x,y) \wedge \text{tennis}(y)$

Axiomatic theory

Ontological precision

Ontology and natural language semantics

Ontology and semantics

- Strictly intertwined: ontology is about ***what there is*, semantics is about *referring*** to what there is...
- Structural semantics vs. ***referential semantics***
- Different aspects of language, different roles of ontology
  - Complex sentences (conjunctions, conditionals...)
  - Primitive sentences (predication)
  - Quantifiers and modifiers
  - Prepositions
  - Nouns and verbs
  - Discourse structure

The ontological commitment of natural language

- Every natural language (or maybe every contextualized sentence) *commits* to some ontology, in two ways:
  - Through a *closed* system of grammatical features
  - Through an *open* system of lexemes
- "Ontological semantics" [Nirenburg & Raskin 2004]: the semantics is driven by an ontology.
  - Practical role of ontologies for NLP systems

Which ontology for NL?

- Quine: every (logical) theory commits to the class of entities it ***quantifies on***.

- Problems:

  - Should every common noun correspond to an ontological category?
    - Questionable entities: Events, features, qualities, fictional characters...

  - Should different linguistic behaviors mark/reflect different ontological categories?

Giving a flower, giving a kiss[1]

- *Descriptive approach:* semantic structure of sentences is preserved (as best as possible)
- *Revisionary approach:* ontological eliminativism based on ***paraphrasability***:
  - John gives a kiss to Mary (Mary is given a kiss by John)
  - John kisses Mary (Mary is kissed by John)

  - John gives a flower to Mary
  - *John flowers Mary

  - There is a hole in this wall
  - This wall is holed

  - This statue has a long nose
  - This statue is long-nosed

[1] Thanks to Achille Varzi

The traps of revisionism

- Is ***systematic*** paraphrasing really possible (also for complex sentences)?
  - There are 7 holes in this piece of cheese
- How to choose ***whether*** paraphrasing?
  - Mary makes a leap
  - Mary makes a cake
- Can we account for ***proper inferences***?
  - There are two things John gave to Mary: a kiss and a flower

- ***Where to stop*** while eliminating entities?
  - Should we paraphrase everything in terms of bunches of molecules moving around? [not very interesting for a linguist...]

The rich ontology of Natural Language

Multiple ***co-located events***

- ***John sings while taking a shower***

Multiple ***co-located objects***

- I am talking here
- *This bunch of molecules is talking
- *What’s here now is talking

- This statue is looking at me
- *This piece of marble is looking at me
- This statue has a strange nose
- *This piece of marble has a strange nose

Individual ***qualities***

- The temperature of this room is increasing
- I like the color of this rose
- The color of this rose turned from red to brown in one week

Ontology and polysemy

- Systematic polysemy [Pustejovsky]
  - Book: text/physical object
  - Window: opening/artifact
  - Apple: fruit/substance
  - ....

- A reason for not taking ontological semantics seriously? [Wliks]
- A reason for making clear the separation between lexicon and ontology? [Niremburg]
- A linguistic phenomenon *explained* by ontological dependence?

Ontology: an overloaded term

Ontologies vs. classifications

- Classifications focus on:
  - ***access***, based on pre-determined criteria (encoded by ***syntactic keys***)

- Ontologies focus on:
  - ***Meaning*** of terms
  - ***Nature*** and ***structure*** of a domain

Ontologies vs. Knowledge Bases

- Knowledge base
  - Assertional component
    - **reflects *specific (epistemic) states of affairs***
    - designed for ***problem-solving***
  - Terminological component (*ontology*)
    - ***independent*** of particular ***states of affairs***
    - Designed to support ***terminological services***

Ontological formulas are (assumed to be) *necessarily true*

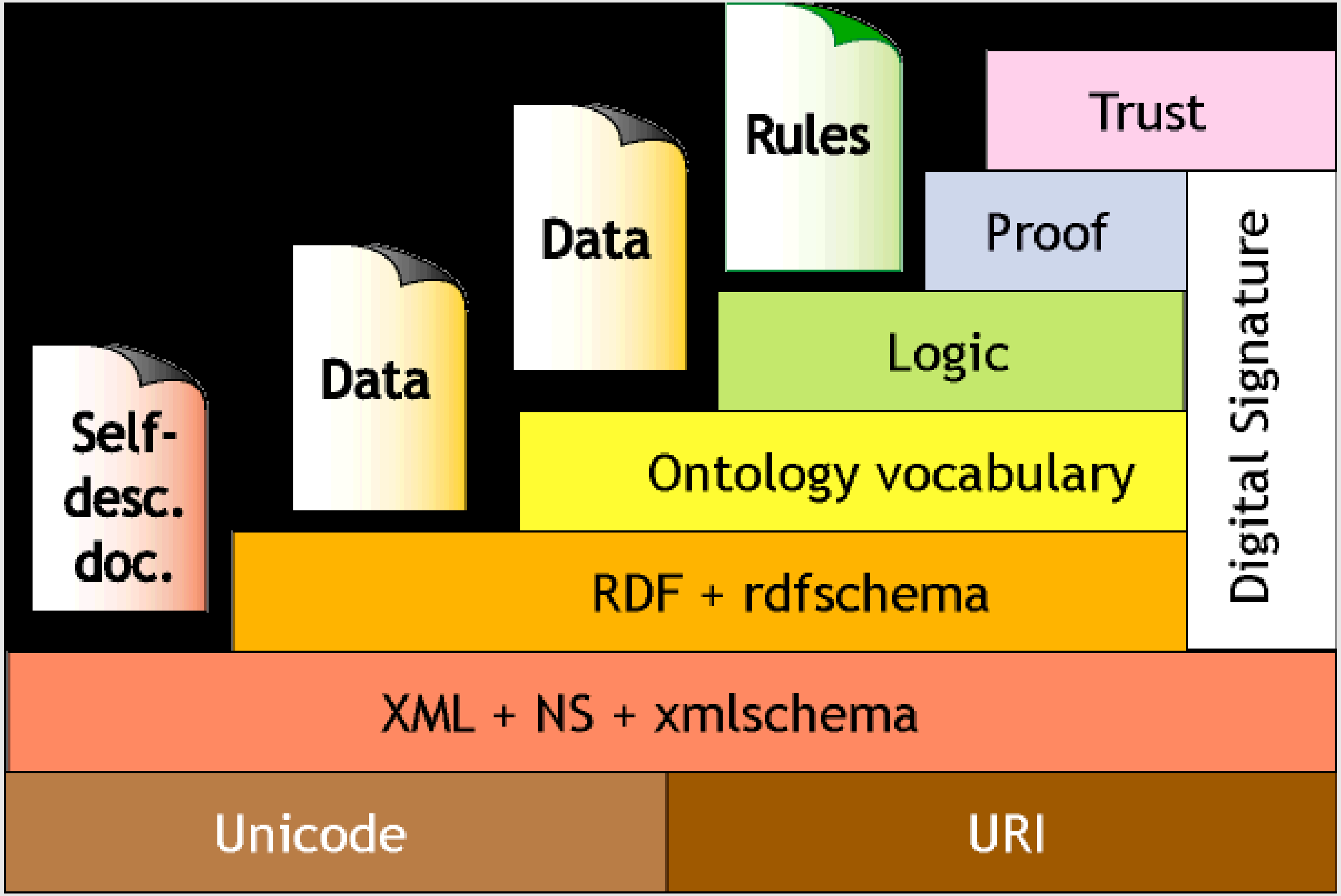
Ontologies vs. Conceptual Schemas

- **Conceptual schemas**
  - not accessible at run time
  - not always have a formal semantics
  - constraints focus on ***data integrity***
  - attribute values ***taken out of the UoD***

- **Ontologies**
  - **accessible at run time** (at least in principle)
  - **formal semantics**
  - constraints focus on ***intended meaning***
  - attribute values **first-class citizens**

The role of *Formal* Ontology

The semantic web architecture [Tim Berners Lee 2000]



Formal Ontology

- Theory of ***formal distinctions and connections*** within:
  - entities of the world, as we perceive it (***particulars***)
  - categories we use to talk about such entities (***universals***)
- Why ***formal***?
  - Two meanings: ***rigorous*** and ***general***
  - Formal logic: connections between truths - neutral wrt ***truth***
  - Formal ontology: connections between things - neutral wrt ***reality***

Formal Ontological Analysis

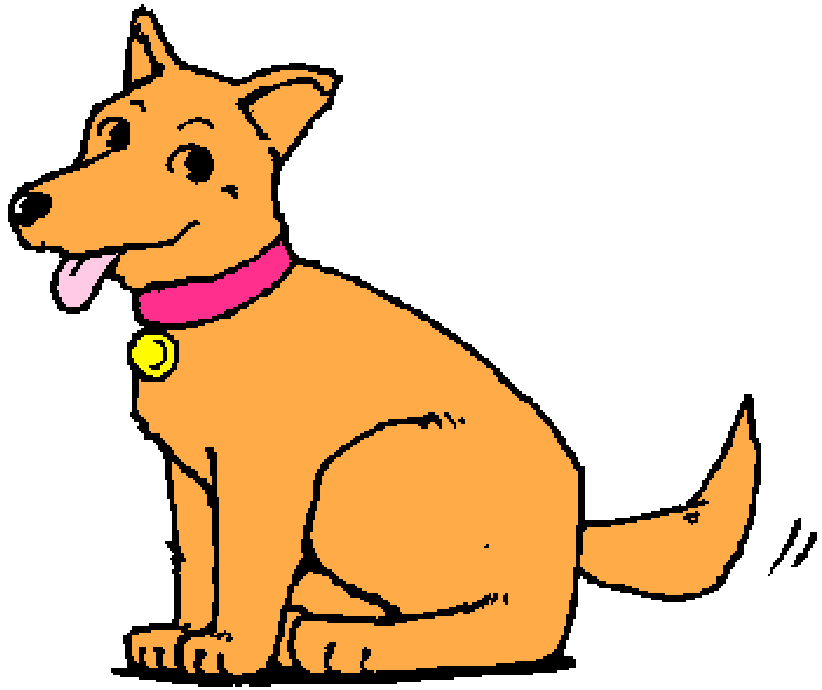
- Theory of Essence and Identity
- Theory of Parts (Mereology)
- Theory of Wholes
- Theory of Dependence
- Theory of Composition and Constitution
- Theory of Properties and Qualities

The basis for a common ontology vocabulary

Identity, Unity, and Essence

- **Identity**: is this my dog?
  - Essential properties of ***dogs***
  - Essential properties of ***my dog***

- **Unity**: is the collar part of my dog?
  - ***Being a whole*** (of a certain kind) is also an essential property



Kinds of Whole

- Depending on the **nature of the *unifying relation***, we can distinguish:
    - ***Topological wholes*** (a piece of coal, a heap of coal)
    - ***Morphological wholes*** (a constellation)
    - ***Functional wholes*** (a hammer, a bikini)
    - ***Social wholes*** (a population)

* a whole can have ***parts that are themselves wholes*** (with a different unifying relation)

OntoClean:
useful distinctions among properties

Essential properties and rigidity

- Certain entities must have some properties in order to exist;
  - John must have a brain
  - John must be a person.
- Certain properties are essential to ***all*** their instances (compare ***being a person*** with ***having a brain***).
- These properties are ***rigid*** - if an entity is ever an instance of a rigid property, it must necessarily be such.

Note: what does "exist" mean?

For concrete objects, being present at t...

Carrying essential properties

- A property P *carries* a (relevant) essential property Q (different from P) iff Q is essential to all instances of P, and still Q is not rigid:
  - Every person must have a brain.

- Compare with:
  - Every person must be a mammal.

Sortals and other properties

- **Sortals** (*horse, triangle, amount of matter, person, student...)*
  - Carry (non-trivial) identity conditions
  - Usually correspond to ***nouns***
  - High organizational utility
- **Non-sortals** *(red, big, old, decomposable, dependent...)*
  - No identity
  - Usually correspond to ***adjectives***
  - Span across different sortals
  - Limited organizational utility (but high semantic value)
- **Categories** *(universal, particular,object, event, substance...)*
  - No identity
  - Useful generalizations for sortals
  - Characterized by a set of (only necessary) formal properties
  - Good organizational utility

Carrying vs. Supplying Identity

- ***Supplying*** (global) identity (**+O**)
  - Carrying an IC (or essential property) that doesn't hold for *all* directly subsuming properties
- ***Carrying*** identity (**+I**)
  - Not supplying identity, while being subsumed by a property that does.
- ***Common sortal principle***: x=y -> there is a common sortal supplying their identity

- Theorem: only rigid properties supply identity

Heuristics for Identity

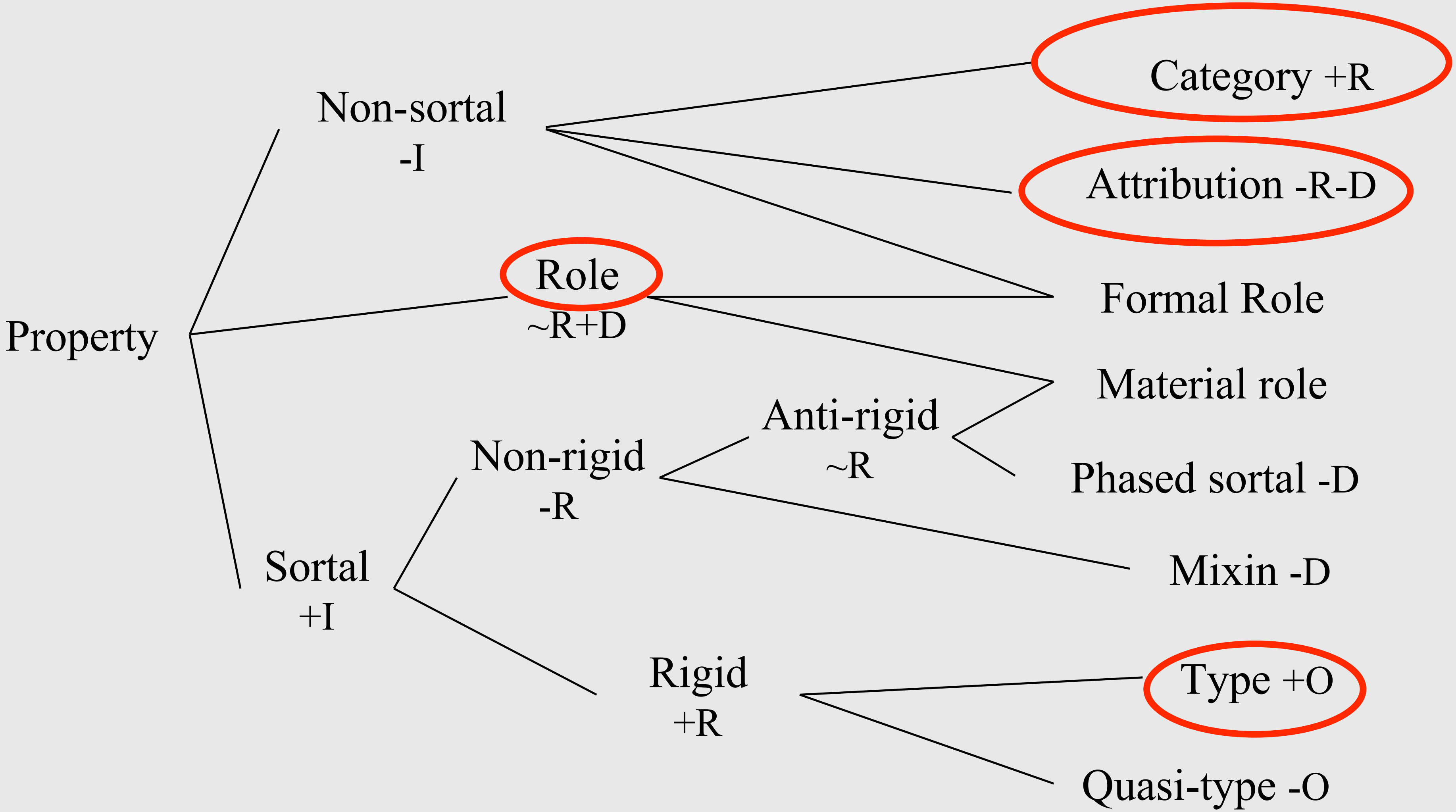
- Finding necessary *and* sufficient ICs for a given property may be ***very hard***.

- Heuristic 1: ***at least a sufficient IC***.
- Heuristic 2: ***some essential parts or qualities***
- Heuristic 3: ***some essential (non-rigid) properties***

Dependence

- Between particulars
  - ***Existential dependence*** (specific/generic)
    - Hole/host, person/brain, person/heart
  - Historical dependence
    - Person/parent
  - Causal dependence
    - Heat/fire
- Between universals
  - ***Definitional dependence***
    - *P* depends on *Q* iff *Q* is involved in the **definition** of *P*.
    - Metaproperties: +D/-D

The OntoClean ontology of properties



DOLCE: a *Descriptive Ontology for Linguistic and Cognitive Engineering*

DOLCE
a Descriptive Ontology for Linguistic and Cognitive Engineering

- Strong cognitive/linguistic bias:
  - ***descriptive*** (as opposite to *prescriptive*) attitude
  - Categories mirror cognition, common sense, and the lexical structure of natural language.
- Emphasis on ***cognitive invariants***
- Categories as ***conceptual containers***: no “deep” metaphysical implications
- Focus on ***design rationale*** to allow easy comparison with different ontological options
- Rigorous, systematic, interdisciplinary approach
- ***Rich axiomatization***
  - *37 basic categories*
  - *7 basic relations*
  - *80 axioms, 100 definitions, 20 theorems*
- Rigorous quality criteria
- Documentation

DOLCE's basic taxonomy

- Endurant
  - Physical
    - Amount of matter
    - Physical object
    - Feature
  - Non-Physical
    - Mental object
    - Social object
  - …
- Perdurant
  - Static
    - State
    - Process
  - Dynamic
    - Achievement
    - Accomplishment
- Quality
  - Physical
    - Spatial location
    - …
  - Temporal
    - Temporal location
    - …
  - Abstract
- Abstract
  - Quality region
    - Time region
    - Space region
    - Color region
    - …
  - …

DOLCE's Basic Ontological Choices

- **Endurants** (aka *continuants* or *objects*) and **Perdurants** (aka *occurrences* or *events*)
  - distinct categories connected by the relation of *participation*.

- **Qualities**
  - Individual entities *inhering in* Endurants or Perdurants
  - can live/change with the objects they inhere in
  - Instance of *quality kinds,* each associated to a **Quality Space** representing the *"values" (qualia)* that qualities (of that kind) can assume. Quality Spaces are neither in time nor in space.

- **Multiplicative approach**
  - Different Objects/Events can be spatio-temporally co-localized: the relation of *constitution* is considered.

Endurants and Perdurants

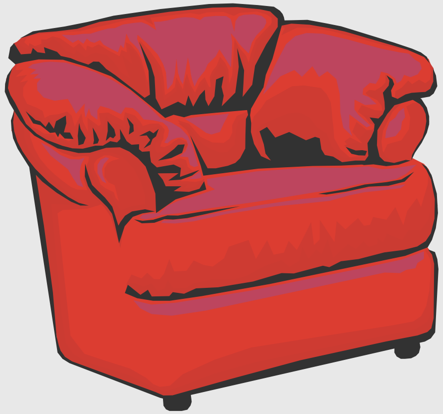
- Endurants (3D *continuants*)
  - Need a time-indexed parthood relation
  - Exist in time
  - Can genuinely change in time
  - May have non-essential parts
  - All proper parts are present whenever they are present (wholly presence, no temporal parts)

- Perdurants (4D *occurrences*[1]) [*Occurrents* are *occurrence-types*]
  - Do not need a time-indexed parthood relation
  - Happen in time
  - Do not change in time *(as a whole...)*
  - All parts are essential
  - Only some proper parts are present whenever they are present (partial presence,temporal parts)

- Endurants *participate to* Perdurants

(1)

Qualities



The rose and the chair have *the same color*:

- different color qualities inhere to the two objects
- they are located in the same quality region

Therefore, the same color attribute (red) is ascribed to the two objects

Qualities vs. Features

- **Features**: “parasitic” physical entities.
- **relevant parts** of their host…
  … or **places**
- Features have qualities, qualities have no features.

OntoWordNet

Using WordNet as an ontology

- Unclear semantic interpretation of hyperonimy
  - Instantiation vs. subsumption
  - Object-level vs. meta-level
  - Hyperonymy used to account for polysemy
    - (law both a document and a rule)
- Unclear taxonomic structure
  - Glosses not consistent with taxonomic structure
  - Heterogeneous leves of generality
  - Formal constraints violations (especially concerning roles)
- Polysemous use of antonymy (child/parent vs. daughter/son)
- Poor ontology of adjectives and qualities
- Shallow taxonomy of verbs

Mapping with lexicons: the OntoWordNet project

(Aldo Gangemi, Alessandro Oltramari, Massimiliano Ciaramita)

- 809 synsets from WordNet1.6 directly subsumed by a DOLCE+ class
  - Whole WordNet linked to DOLCE+
  - Lower WordNet levels still need revision
- Glosses being transformed into DOLCE+ axioms
  - Machine learning applied jointly with foundational ontology
- WordNet “domains” being used to create a modular, general purpose domain ontology
- Ongoing work on ontological analysis of specific WordNet domains (cognition, emotion, psychological feature)
- Ongoing cooperation with Princeton University.

The OntoWordNet methodology

1. **Populate** a general ontology (DOLCE) by adding single synsets (or whole taxonomy branches) from a c. lexicon (upon suitable classification)
2. **Restructure** a c. lexicon by checking ontological constraints (e.g. *OntoClean* meta-properties) throughout the branches
3. **Merge** an ontology and a c. lexicon (includes 1. and 2.)
4. **Enrich** the resulting structure by extracting relationships from the glosses.

A Selection of Most Relevant Projects (2003-2006)

- **WonderWeb** (FP5): Ontology Infrastructure for the Semantic Web (LOA: foundational ontologies for the Semantic Web)
- **OntoWeb** (FP5 - NoE): Ontology-based information exchange for knowledge management and electronic commerce (LOA: *SIG on Content Standards*)
- **METOKIS** (FP6): Methodologies and tools infrastructure for the development of multimedia knowledge units
- **SEMANTIC MINING** (FP6 - NoE): Semantic Interoperability and Data Mining in Biomedicine
- **TICCA** (PAT&CNR): Tecnologie cognitive per l'interazione e la cooperazione con agenti artificiali (LOA: ontology of social interaction)
- **MOSTRO** (PAT); Modelling Security and Trust Relationships in Organizations
- **IKF** : Intelligent Knowledge Fusion (Eureka Project)
  - *Ontology of banking transactions* (*with* ELSAG Banklab_)
  - *Ontology of Service-Level Agreement and IS monitoring* (*with* SELESTA_)
  - *Ontology of Insurance Services (with Nomos SpA)*
- **FOS** (UN/FAO): Alignment of legacy fishery ontologies
- **NEON** (FP6) - Networked Ontologies
- **ONTOGEO** (FP6) - Geo-spatial Semantic Web

DOLCE Extensions
(mainly by Aldo Gangemi @LOA-RM)

- Allen-based ontology of time for events
- Ontology of common-sense locations
- Descriptions and Situations (D&S) ontology (reified relations and relationships)
- Ontology of Functional Participation (cf. *thematic roles*)
- Ontology of Plans and Tasks (DDPO) (Metokis project)
- Ontology of Information Objects (DDIO (Metokis project)
- Ontology of Knowledge Content Objects (KCO), from Metokis, for multimedia description and negotiation
- Ontology of Services, based on DDPO (with UKA, VUA)
- Ontology of Semantic Middleware (by Daniel Oberle at UKA)
- Core Legal Ontology (CLO, with ITTIG-CNR)
- Metaontology of ontology as semiotic object (O2)
- Ontology of ontology evaluation and quality (oQual)
- Ontology of design patterns
- Ontology of social entities and organizations (MOSTRO project @LOA-TN)

When is a foundational ontology useful?

1. When *subtle distinctions* are important
2. When *recognizing disagreement* is important
3. When *rigorous referential semantics* is important
4. When *general abstractions* are important
5. When *careful explanation and justification* of ontological commitment is important
6. When ***mutual understanding*** is more important than interoperability.

Community-based Access vs. Global Knowledge Access

different roles of ontologies

- **Community-based access**
  - Intended meaning of terms more or less ***known in advance***
  - ***Taxonomic reasoning*** (e.g. for classification/retrieval purposes) is the main ontology service
  - ***Limited expressivity***
  - ***On-line reasoning*** (stringent computational requirements)

- **Global knowledge access**
  - ***Negotiate meaning*** across different communities
  - ***Establish consensus*** about meaning of a new term within a community
  - ***Explain meaning*** of a term to somebody new to community
  - ***Higher expressivity*** required to express intended meaning
  - ***Off-line reasoning*** (only needed ***once***, before cooperation process starts)

A new journal: ***Applied Ontology***

Editors in chief:

Nicola Guarino
ISTC-CNR

Mark Musen
Stanford University

IOS Press

Amsterdam, Berlin,
Washington, Tokyo, Beijing

www.applied-ontology-org

FOIS-2006

International Conference on Formal Ontology in Information Systems

http://www.formalontology.org/

November 9-11, 2006
Baltimore, Maryland (USA)

Conclusion

- Subtle meaning distinctions do matter
- Formal ontological analysis provides for a rigorous methodology to characterize intended meaning
- A humble, truly interdisciplinary approach is essential

…Is this hard?

Of course yes!

(Why should it be easy??)

Extra slides

How a formal ontological theory looks like: mereology

- Primitive: ***proper part-of*** relation (PP)
  - asymmetric
  - transitive
  - **$Pxy =_{def} PPxy \vee x=y$**
  - **$Oxy =_{def} \exists z(Pzx \wedge Pzy)$**
- Axioms:

supplementation: **$PPxy \rightarrow \exists z (PPzy \wedge \neg Ozx)$**

principle of sum: **$\exists z \forall w (Owz \leftrightarrow (Owx \vee Owy))$**

extensionality: **$x = y \leftrightarrow \forall w(Pwx \leftrightarrow Pwy)$**

Excluded models: