Helping people (and machines) understand each other: the role of formal ontology

Nicola Guarino

Laboratory for Applied Ontology (LOA)
Institute for Cognitive Sciences and Technology (ISTC-CNR)
Trento-Roma, Italy
www.loa-cnr.it

Summary

- Why ontologies
- What ontologies are (or should be…)
- Ontology quality
- Formal ontology and the quest for general primitives

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

The need for dynamic meaning mediation (and negotiation)

"Lack of technologies and products to ***dynamically mediate discrepancies in business semantics*** will limit the adoption of advanced Web services for large public communities whose participants have disparate business processes"

Gartner Research, February 28, 2002

Where subtle distinctions in meaning are important

- US elections: how many ***holes***?
- Twin towers catastrophe: how many ***events***?

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

Ultimately, communication is among PEOPLE

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 vocabularies are not the solution

- 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***

The key problems

- *Semantic matching*
- *Semantic integration*

Ontologies: *a **magic** solution?*

Community-based Access vs. Global Knowledge Access

different roles of ontologies

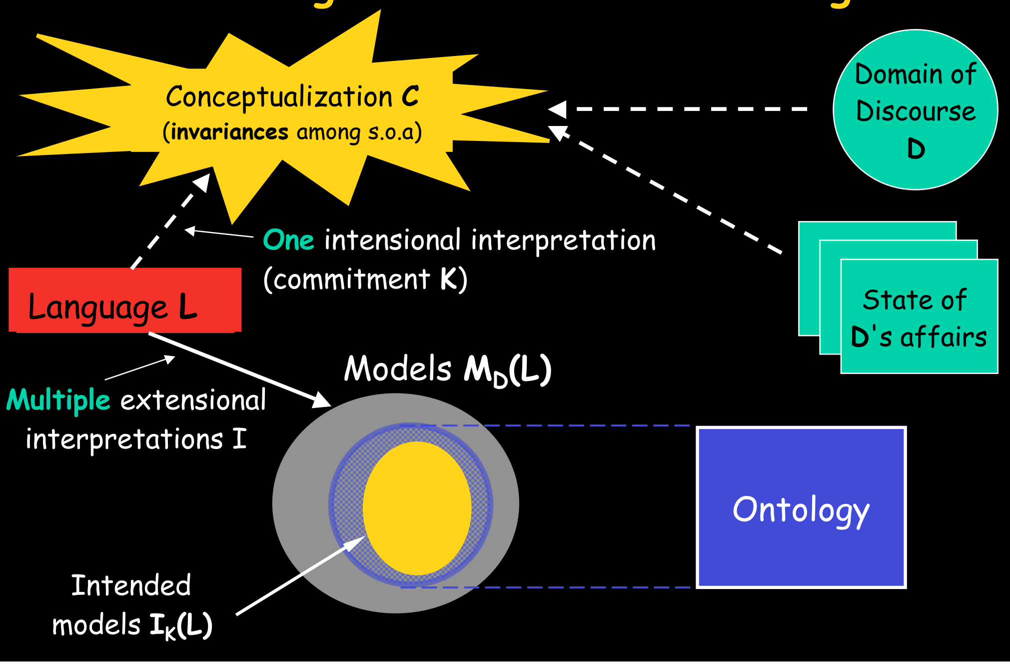
- **Community-based access**
  - Intended meaning of terms ***known in advance***
  - ***Taxonomic reasoning*** is the main ontology service
  - ***Limited expressivity***
  - ***On-line reasoning*** (stringent computational requirements)

- **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)

Ontology and ontologies

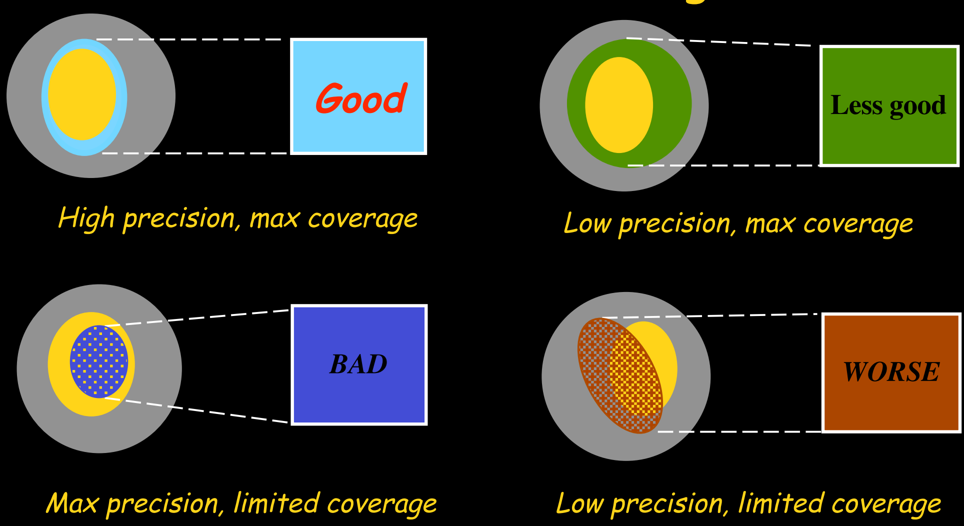
- Ontology (capital "o"):
  - a ***philosophical discipline***:
    - The study of the ***nature and structure*** of ***possible*** entities
- An ontology (lowercase "o"):
  - a ***specific artifact*** designed with the purpose of ***expressing the intended meaning of a vocabulary*** in terms of the nature and structure of the entities it ***refers*** to

Ontologies and intended meaning

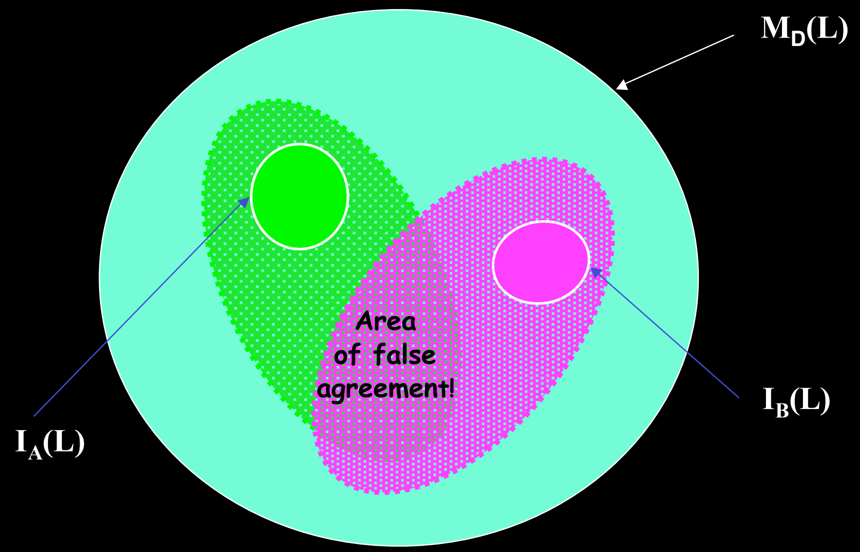


Ontology quality

Ontology Quality: Precision and Coverage



Why precision is important



When precision is not enough

Only one binary predicate in the language: on

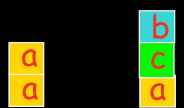
Only blocks in the domain: a, b, c, …

Axioms (for all x,y,z):

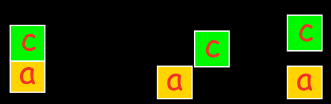
$$on(x,y) \rightarrow \neg on(y,x)$$

$$on(x,y) \rightarrow \neg \exists z\, (on(x,z) \wedge on(z,y))$$

Non-intended *models* are excluded, but the rules for the competent usage of on in different *situations* are not captured.



Excluded situations



Indistinguishable situations

Precision vs. Accuracy

- In general, a single intended *model* may not discriminate among relevant alternative *situations* *because of*
  - Lack of *primitives*
  - Lack of *entities*
- Capturing all intended models is not sufficient for a "perfect" ontology

  Precision: non-intended *models* are excluded

  Accuracy: non-intended *situations* are excluded

The problem of primitives

- Representation primitives vs. ***ontological primitives*** (against ***arbitrary*** interpretations)
- Let's aim at ***general*** primitives, similarly to what happens in mathematics: ***set, relation, transitive, symmetric…***

The Ontological Level

(Guarino 94)

Level	*Primitives*	*Interpretation*	*Main feature*
Logical	Predicates, functions	Arbitrary	Formalization
Epistemological	Structuring relations	Arbitrary	Structure
Ontological	**Ontological relations**	**Constrained**	**Meaning**
Conceptual	Conceptual relations	Subjective	Conceptualization
Linguistic	Linguistic terms	Subjective	Language dependence

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

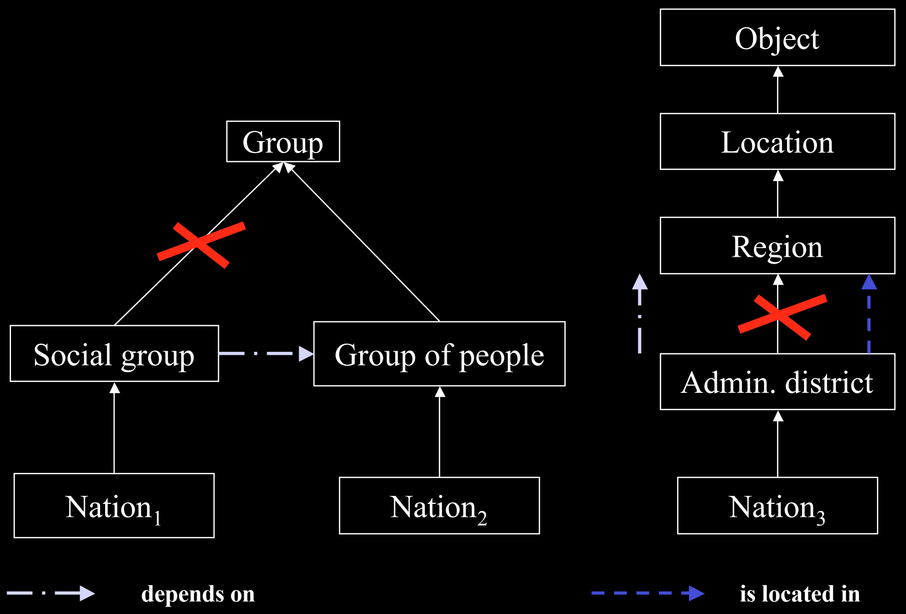
- Theory of Essence and Identity
- Theory of Parts
- Theory of Wholes
- Theory of Dependence
- Theory of Composition and Constitution
- Theory of Qualities
- Theory of Participation
- Theory of Representation

The basis for a common ontology vocabulary

Essence and identity

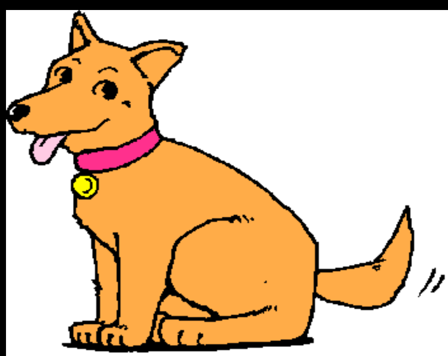
- Identity can be characterized in terms of essential properties
- Concepts with incompatible essential properties are ***disjoint***
- Some properties are essential to their instances (***rigidity***)
- Rigidity is at the basis of the distinction between ***types*** (*person*) and ***roles*** (*student*)

The case of "Nation"



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 lump 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)

*Agreeing on conditions for identity and unity is at the basis of **meaning negotiation***

The need for *general* ontological primitives

No axioms, no semantics…

- No axioms, "free" interpretations
- Free interpretations = NO semantics
- Encoding primitive "formal" relations with no axioms does not solve anything

- Too much emphasis on encoding and representation, no ***shared principles for axiomatization…***

Mereology

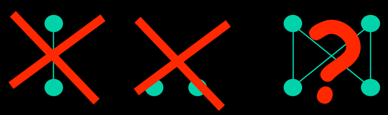
- Primitive: ***proper part-of*** relation (PP)
  - asymmetric
  - transitive
  - $Pxy =_{def} PPxy \vee x=y$
- Axioms:

supplementation: $PPxy \rightarrow \exists z\,(PPzy \wedge \neg z=x)$

principle of sum: $\exists z\,(PPxz \wedge PPyz \wedge \neg \exists w(PPwz \wedge \neg(Pwx \vee Pwy)))$

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

Excluded models:



Foundational ontologies

- Provide a ***carefully crafted taxonomic backbone*** to be used for domain ontologies
- Help recognizing and understanding ***disagreements*** as well as agreements
- Improve ontology development ***methodology***
- Provide a principled mechanism for the semantic integration and ***harmonisation*** of existing ontologies and metadata standards
- Improve the ***trust*** on web services

Mutual understanding **vs. mass interoperability**

An Interdisciplinary Approach

- Towards a unified ***Ontology-driven Modelling Methodology*** for databases, knowledge bases and OO-systems
  - Grounded in reality
  - Transparent to people
  - Rigorous
  - General
- Based on
  - Logic
  - Philosophy
  - Linguistics

DOLCE

a Descriptive Ontology for Linguistic and Cognitive Engineering

- **Strong cognitive bias: *descriptive* (as opposite to *prescriptive*) attitude**
- **Emphasis on *cognitive invariants***
- **Categories as *conceptual containers*: no "deep" metaphysical implications wrt "true" reality**
- **Clear *branching points* to allow easy comparison with different ontological options**
- ***Rich axiomatization***
  - ***37 basic categories***
  - ***7 basic relations***
  - ***80 axioms, 100 definitions, 20 theorems***

Conclusions

- *Subtle meaning distinctions* do matter
- *General ontological primitives* help making intended meaning explicit
- Realizing *reasons of disagreement* may be more important than forcing agreement
- A humble *interdisciplinary approach* is essential

...Is this hard?!

Of course yes! (Why should it be easy??)

Let's do SCIENCE!

Call for Papers

FOIS-2004

International Conference on Formal Ontology in Information Systems

http://www.fois.org

Picture courtesy of Fototeca Web del Comune di Torino

November 4-6, 2004, Torino (Italy)

Extra slides

Research priorities at the ISTC-CNR Laboratory for Applied Ontology

- Foundational ontologies and ontological analysis
- Domain ontologies
  - Physical objects
  - Information and information processing
  - Social interaction
  - Ontology of legal and financial entities
- Ontology, language, cognition
- Ontology-driven information systems
  - Ontology-driven conceptual modeling
  - Ontology-driven information access
  - Ontology-driven information integration

www.loa-cnr.it

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

Ontologies vs. Knowledge Bases

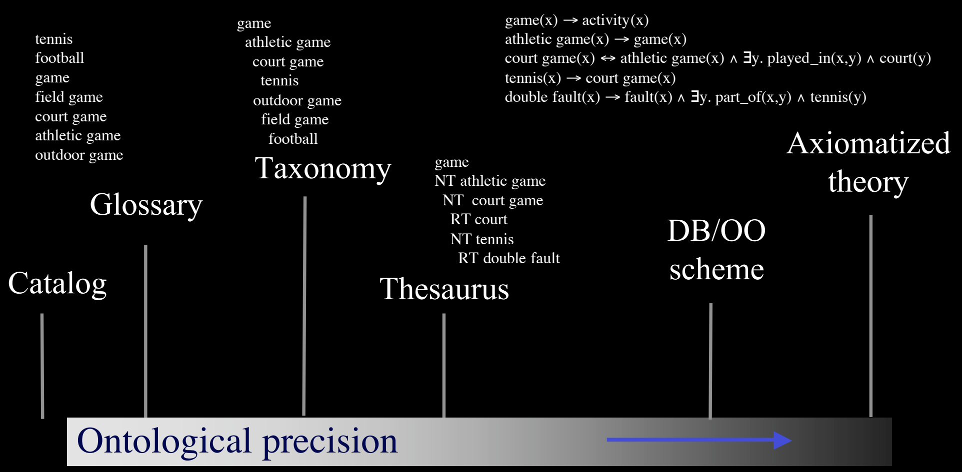
- Knowledge base
  - *Assertional component*
    - **reflects** ***specific (epistemic) states of affairs***
    - designed for ***problem-solving***
  - Terminological component (*ontology*)
    - ***independent*** of particular ***states of affairs***
    - Designed to support ***terminological services***

Ontological formulas are (assumed to be)
necessarily true

Ontologies vs. 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

Levels of Ontological Precision



Towards a practical procedure for rigorous ontology evaluation

- Determine a list of ***situations*** which the ontology is supposed to cover
- Document these situations by means of ***illustrations*** (annotated multimedia documents) showing the ***agreed intended use*** of ontology terms
- Make sure that for each term multiple examples and ***counter-examples*** are given
- Establish the *correspondence* between ***situations*** and ***ontology models***

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**
    - **…**
  - **…**

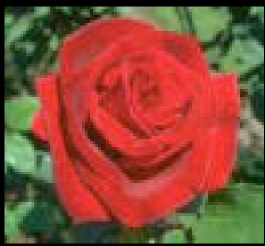
Abstract vs. Concrete Entities

- Concrete: located in space-time (regions of space-time are located in themselves)
- Abstract - two meanings:
  - Result of an abstraction process (something common to multiple exemplifications)
  - *Not located in space-time*
- Mereological sums (of concrete entities) are concrete, the corresponding sets are abstract...

Endurants vs. Perdurants

- **Endurants**:
  - All proper parts are present whenever they are present (***wholly presence***, no temporal parts)
  - Exist in time
  - Can genuinely change in time
  - May have non-essential parts
  - Need a time-indexed parthood relation
- **Perdurants**:
  - Only some proper parts are present whenever they are present (***partial presence***, temporal parts)
  - Happen in time
  - Do not change in time
  - All parts are essential
  - Do not need a time-indexed parthood relation

Qualities vs. Features



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

INTEROP panel

The key problems

- *Semantic matching*
- *Semantic integration*

Ontologies: *a* ***magic*** *solution?*

Same hard problems, new slogans

- 90's: enterprise integration, cuncurrent engineering, agile manufacturing…
  - STEP…
- More than 10 years later, we are in the same situation

Solved (?) problems at the theoretical level

Most of the representation/reasoning theoretical issues (especially concerning a single data/knowledge base)

- Management of global vs. local views
- The very idea of using ontologies and mediators for integration
- Some semi-automatic ontology learning techniques
- Some semi-automatic "ontology" alignment techniques
- Foundational work concerning the nature of ontologies
- Preliminary methodological work on ontology design
- Good deal of work on mereology, mereotopology, mereogeometry
- Some good, techniques for mining natural language

- The CIDOC experience

Open problems

- A unified conceptual modeling methodology
- Comparisons among ontologies
- Link between ontology and lexicon
- Core ontologies
- Recognize the problem of modeling **content** as a specific one…

How can ontologies help

- Think with interoperability in mind!