Formal Ontology and Natural Language Semantics

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

Thanks to Achille Varzi and to the LOA people

Summary

- Ontology and semantics
- The ontological commitment of natural language
- The role of *formal* ontology
- The basic tools of formal ontological analysis
- Some language-motivated ontological choices
- The DOLCE ontology

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 focuses for semanticists:
  - Complex sentences
  - Primitive sentences
  - Quantifiers and modifiers
  - Prepositions
  - Nouns and verbs

Increasing ontological assumptions

The ontological commitment of natural language

- Every natural language [or maybe every contextualized sentence] *commits* to some ontology, in two ways [Talmy 2000]:
  - 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.

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 different ontological categories?

Descriptive vs. Revisionary Approach

- *Descriptive:* surface semantic structure of sentences is preserved (as best as possible)
- *Revisionary:* ontological eliminativism based on ***paraphrasability***:
  - There is a hole in this wall
  - This wall is holed
  - John gives a kiss to Mary
  - John kisses Mary
  - John gives a flower to Mary
  - *John flowers Mary
  - This statue has a long nose
  - This statue is long-nosed

Two forms of revisionism
[Varzi]

- Hermeneutic (re-interpretational) revisionism
  - What these words *really mean* is...
  - Serious troubles...
    - Holes exist -> something is holed
    - Holes do not exist -> nothing is holed ?!
- Revolutionary revisionism
  - What *we intend to say* with these words is... (Humpty-Dumpty style)

The traps of (revolutionary) 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?

An obvious (descriptive) distinction: Objects vs. Events

- Jones buttered the toast
  - Butter(Jones, toast)
- Jones buttered the toast with the knife
  - Butter(Jones, toast, knife)
- Jones buttered the toast with the knife in the bathroom
  - Butter(Jones, toast, knife, bathroom)
- Problem: *multiple meaning postulates*
  - Butter2(x, y, z) → Butter1(x, y)
  - Butter3(x, y, z, t) → Butter2(x, y, z) . . .
- Solution: *Reifying events* [Davidson]
  - $\exists e\ Butter(e,$ Jones, toast$)$
  - $\exists e\ (Butter(e,$ Jones, toast$) \wedge With(e,$ knife$))$
- Advantages:
  - Event anaphora: *It happened at midnight*
  - event nominalization: *The buttering was slow*
  - Predication over events: *I enjoyed reading the book*

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

The role of *formal* ontology

Once the basic choices are made, how to characterize slight differences in ontological assumptions?

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

Where subtle distinctions in meaning are important

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

…only *ontological analysis* solves these problems!!

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

What is an ontology?

- *An artifact* designed with the purpose of expressing the *intended meaning* of a (shared) *vocabulary*
- A shared vocabulary plus a specification (*characterization*) of its intended meaning

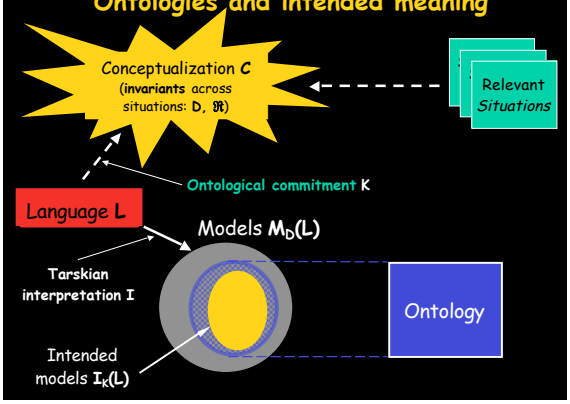
"An ontology is a specification of a conceptualization"
[Gruber 95]

i.e., an ontology accounts for the *commitment* of a language to a certain *conceptualization*

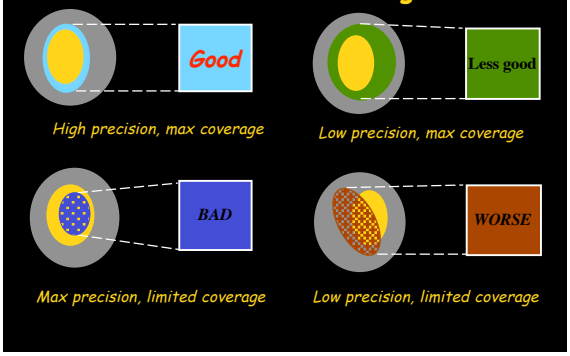
What is a conceptualization

- 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.

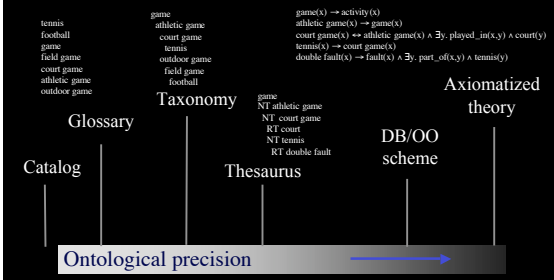
Ontologies and intended meaning



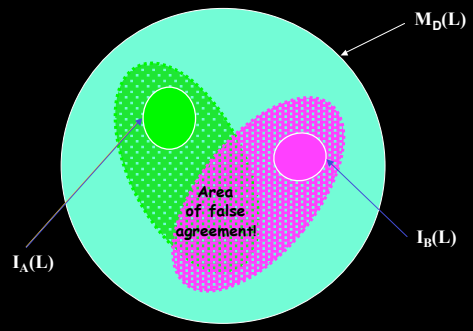
Ontology Quality: Precision and Coverage



Levels of Ontological Precision



Why precision is important



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

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*

The need for *general* ontological primitives

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 Qualities
- Theory of Participation
- Theory of Representation

The basis for a common ontology vocabulary

No axioms, no semantics…

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

An example: mereology

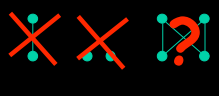
- Primitive: ***proper part-of*** relation (PP)
  - asymmetric
  - transitive
  - $Pxy =_{def} PPxy \vee x=y$
- Further axioms:

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

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

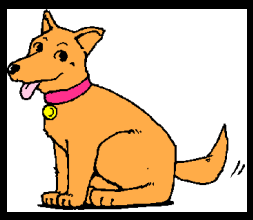
extensionality: $x = y \leftrightarrow (Pwx \leftrightarrow Pwy)$

Excluded models:



Unity, Identity, and Essence

- **Unity**: is the collar part of my dog?
  - *Being a topological whole* is an essential property of my dog
- **Identity**: is this my dog?
  - Essential properties allow us to keep track of my dog across time
    - Individual essential properties of *my dog*
    - **Generic essential properties of** *dogs*



Abstracting domain elements from situations

- Humans isolate *relevant invariances* on the basis of:
  - Perception (as resulting from evolution)
  - Cognition and cultural experience
  - Language
- A set of ***percepts*** is associated to each situation
- Synchronic level: **spatial invariants**
  - Unity properties are ascribed to **percepts patterns**: topological and morphological *wholes* emerge
  - Salient features are recognized
- Diachronic level: **temporal invariants**
  - *Continuants*: identity assumptions among percepts patterns belonging to different situations
  - *Occurrents:* unity properties ascribed to percepts patterns belonging to different situations
- Experience level
  - Categories of continuants are costructed, marked by their essential properties

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)

The role of essential properties

- The emergence of invariant properties is a fact of life
- Essential properties are often result of *conventions* reflecting these invariances
- Ascription of essential properties reflects the need to
  - *explore alternatives* to categorize things
  - rxtrapolate from experience
  - Create expectations

Incompatible essential properties are ***disjoint***

Ontological dependence and systematic polysemy

- Book: text/physical object
- Window: opening/artifact
- Apple: fruit/substance
- ....

- Separation lexicon/ontology [cf. Niremburg, Wilks]

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***

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**
    - **…**
  - **…**

Endurance vs. Perdurance

- **Endurants**:
  - All proper parts are present whenever they are present (***wholly presence***, no temporal parts)
  - Exist in time
  - Can genuinely change in time
  - 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
  - Do not need a time-indexed parthood relation

Qualities and qualia

- Linguistic evidence
  - *This rose is red*
  - *Red is a color*
  - *This rose has a color*
  - *The color of this rose turned to brown in one week*
  - *The room's temperature is increasing*
  - *Red is opposite to green and close to brown*
- Every entity comes with certain qualities that permanently ***inhere*** to it and are ***unique*** of it
- Qualities are perceptually mapped into ***qualia***, which are regions of ***quality spaces***.
- Properties hold because qualities have certain locations in their quality spaces.
- Each ***quality type*** has its own quality space

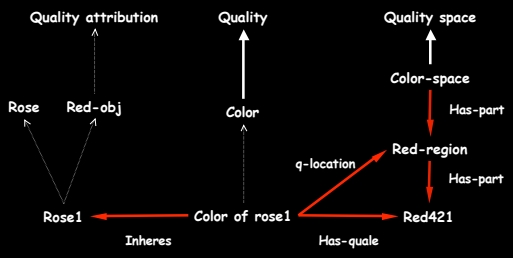
Qualities

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



Qualities vs. Features

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

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...

Physical vs. Non-physical Object

FIAT SpA

- Physical objects:
  - *inherent spatial localization*
  - *not dependent* on other objects (**physical objects, like cars**) or no inherent localization and be dependent on *agents* (**non-physical objects, like laws and institutions**).

•Non-physical objects can also be divided into *mental* (depending on *singular* agents) and *social* (depending on *communities* of agents).

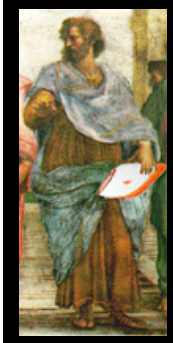
Different roles of ontologies

- **Community-based access *(lightweight ontologies)***
  - 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 *(foundational ontologies)***
  - ***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)

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
  - Cognitive science

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

Application of DOLCE (1) WordNet alignment and OntoWordNet

- 809 synsets from WordNet1.6 directly subsumed by a DOLCE+D&S class
  - Whole WordNet linked to DOLCE+D&S
  - Lower taxonomy levels in WordNet 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)

Applications of DOLCE (2)
Core Ontologies

based on DOLCE, D&S, and OntoWordNet

- Core ontology of plans and guidelines
- Core ontology of (Web) services
- Core ontology of service-level agreements
- Core ontology of (bank) transactions (anti-money-laundering)
- Core ontology for the Italian legal lexicon
- Core ontology of regulatory compliance
- Core ontology of fishery (FAO's Agriculture Ontology Service)
- Core ontology of biomedical terminologies (UMLS)

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**

The multiplicative approach

- The same physical process can participate to multiple equivalence relationships
- Multiple continuants can be located in the same field region
- ***Multiplication of continuants*** is motivated by our cognitive needs

Physical processes vs. events

- A person moves, its body moves. Two (three?) co-located events, one physical process.
- ***Once*** we have constructed continuants, ***then*** we focus on the events they participate…

When precision is not enough

Only one binary predicate in the language: on

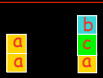
Only blocks in the domain: a, b, c, …

Axioms (for all x,y,z):

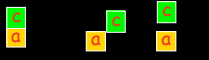
on(x,y) -> ¬on(y,x)

on(x,y) -> ¬∃z (on(x,z) ∧ 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