The role of ontologies for the Semantic Web (and beyond)

Nicola Guarino

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

Summary

- From classifications to ontologies
- Why ontologies
- What ontologies are (or should be…)
- Ontology quality
- Foundational ontologies

A familiar example: classifications

- A set of entities organized according to *access criteria*

- Examples
  - My holidays pictures according to country, sea/mountain/cities…
  - Yahoo directory, Google directory

Problems with multiple classifications

- Different domains
- Different terminology
- Different choices of relevant features
- Different meanings of features
- Different relevant relationships

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

The key problems

- *Semantic matching*
- *Semantic integration*

Simple queries: need more knowledge about what the user wants

- Search for "Washington" (the person)
  - Google: 26,000,000 hits
  - 45th entry is the first relevant
  - Noise: places
- Search for "George Washington"
  - Google: 2,200,00 hits
  - 3rd entry is relevant
  - Noise: institutions, other people, places

The vision: ontology+semantic markup

- *Ontology*
  - *Person*
    - George Washington
    - George Washington Carver
  - *Place*
    - Washington, D.C.
  - Artifact
    - George Washington Bridge
  - Organization
    - George Washington University
- Semantic disambiguation/markup
  - What Washington are you talking about?

The role of taxonomy and lexical knowledge

- Search for "Artificial Intelligence Research"
  - Misses subfields of the general field
  - Misses references to "AI" and "Machine Intelligence" (synonyms)
  - Noise: non-research pages, other fields…

Standard solutions

- Extra knowledge
  - *Taxonomy: specializations*
    - Knowledge Representation
    - Machine Vision etc.
    - Neural networks
  - *Lexicon: synonyms*
    - Artificial Intelligence
    - Machine Intelligence
- Techniques
  - Query Expansion
    - Add disjuncted sub-fields to search
    - Add disjuncted synonyms to search
  - Semantic Markup of question and data
    - Add "general terms" (categories)
    - Add "synonyms"

The vision: ontology-driven search engines

- Idealized view
  - Ontology-driven search engines act as ***virtual librarians*** (or, more realistically, ***librarian assistants***)
    - Determine what you "really mean"
    - Discover relevant sources
    - Find what you "really want"
- Requires common knowledge on all ends
  - Semantic linkage between questioning agent, answering agent and knowledge sources
- Hence the "Semantic Web"

But…

Is the Semantic Web just hype?

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***"

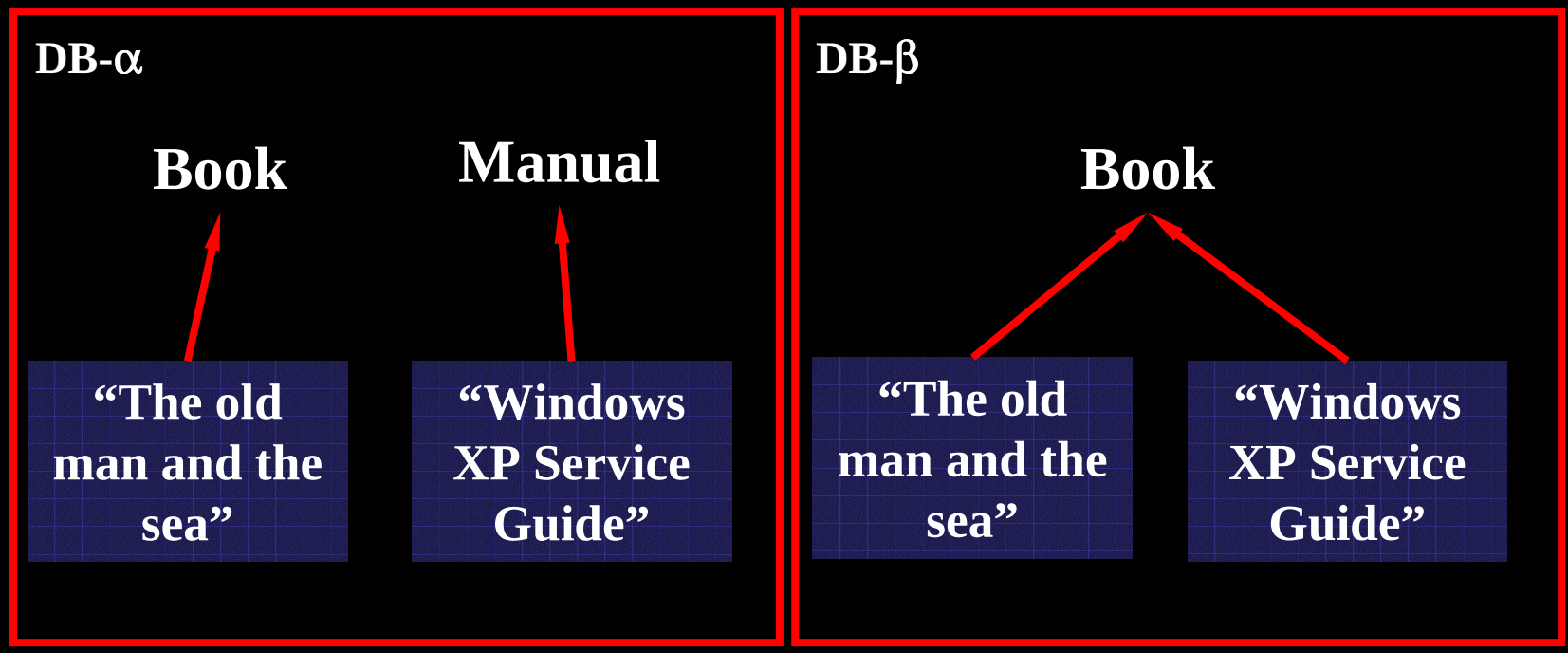
Harvard Business Review, October 2001

Where subtle distinctions in meaning are important

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

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

Same term, different concept



Unintended models must be taken into account!

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

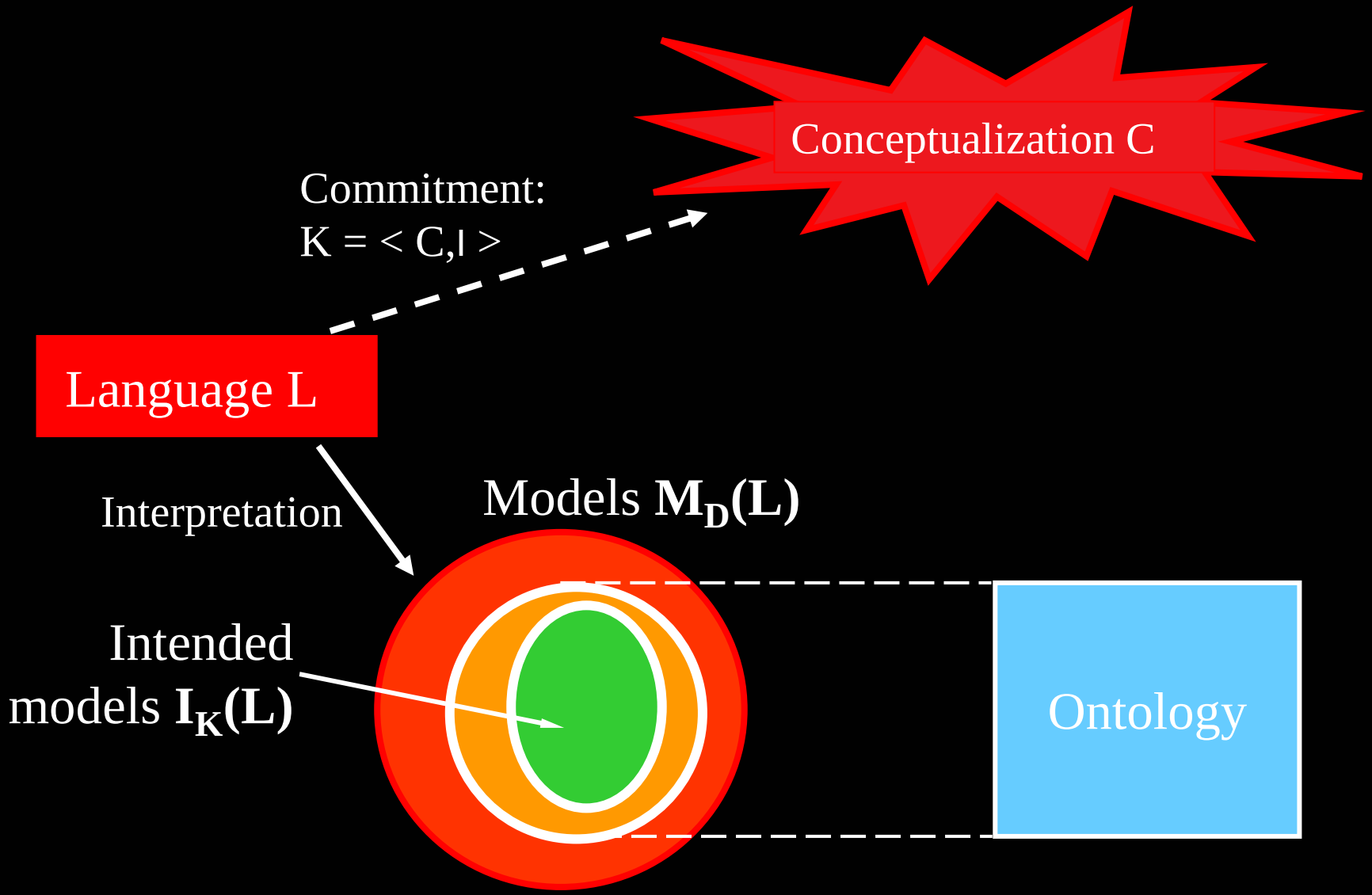
Standard vocabularies are not the solution

- 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

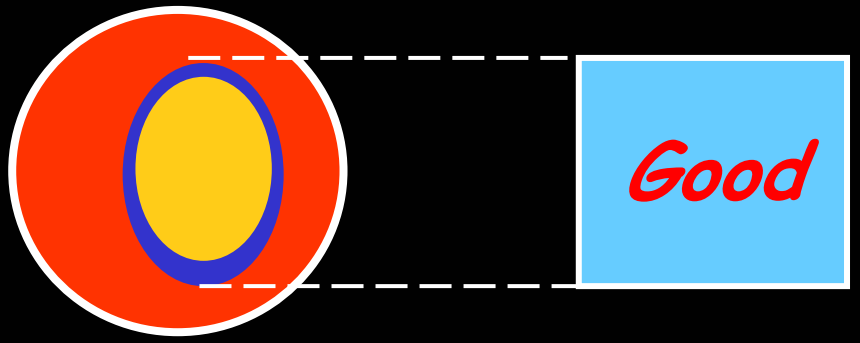
Definitions

- Ontology (capital "o"):
  - a ***philosophical discipline***:
    - The study of "being qua being"
    - The study of what is possible
    - The study of the nature of possible: distinctions among *possibilia*
- An ontology (lowercase "o"):
  - *a* ***specific artifact*** designed with the purpose of ***expressing the intended meaning of a vocabulary***

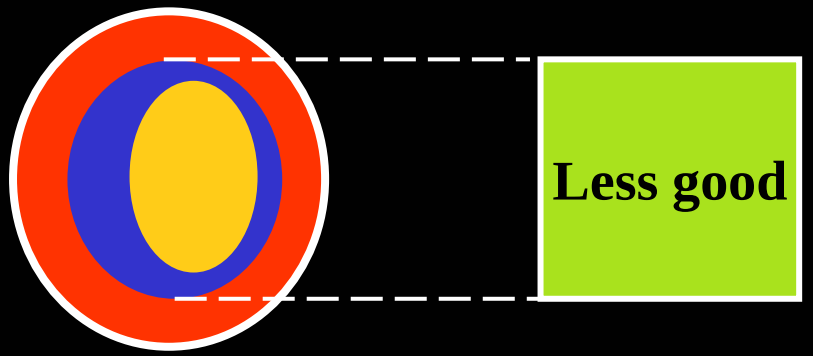
Ontologies and intended meaning



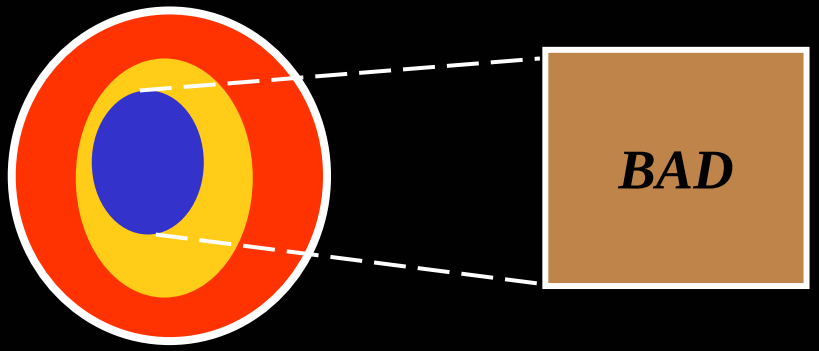
Ontology Quality: Precision and Coverage



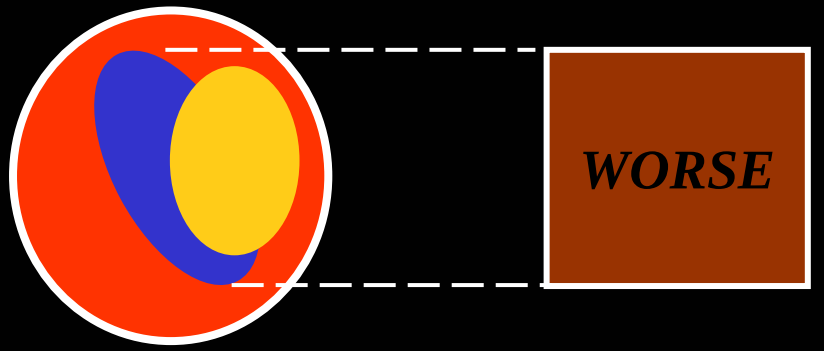
High precision, max coverage



Low precision, max coverage

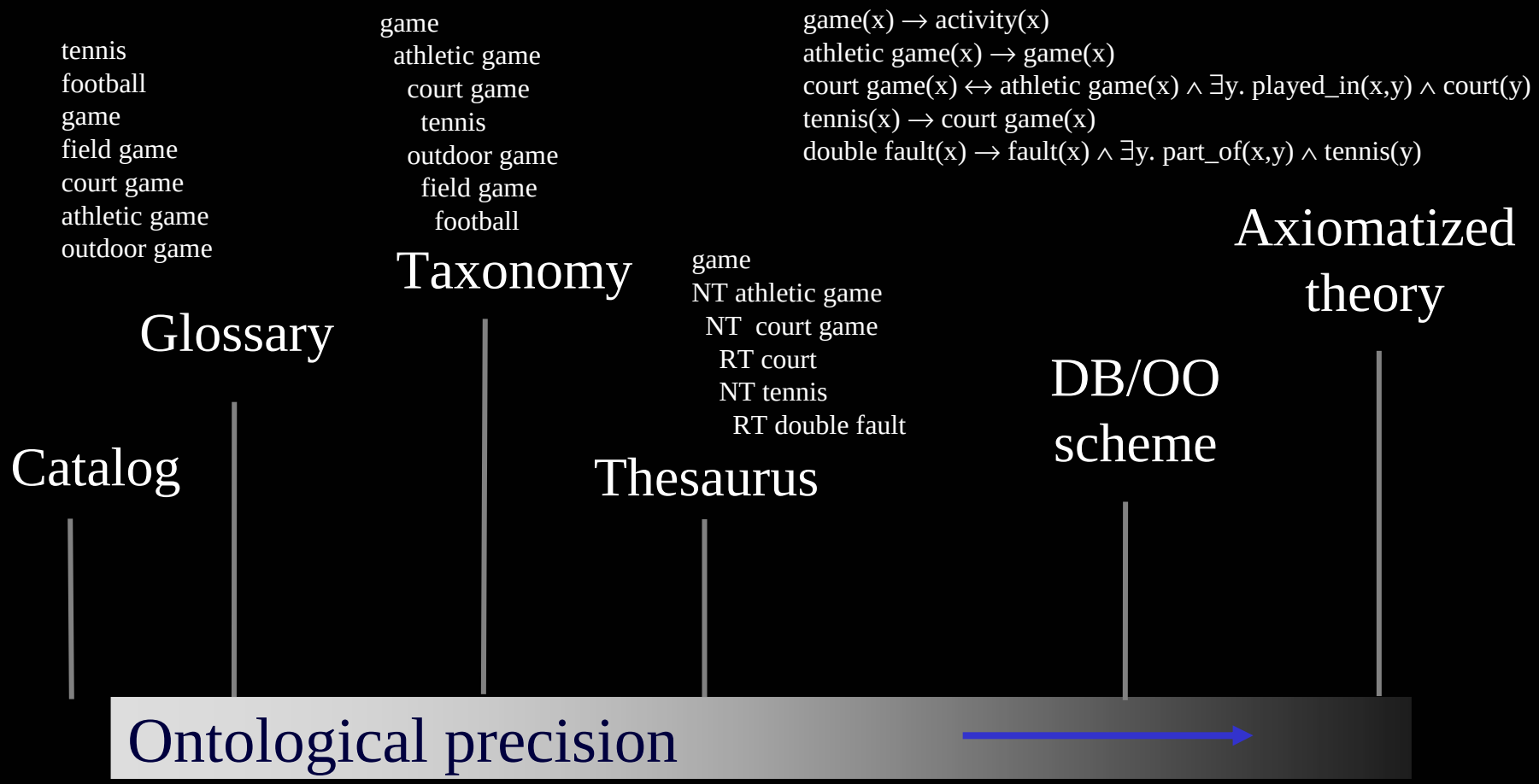


Max precision, low coverage

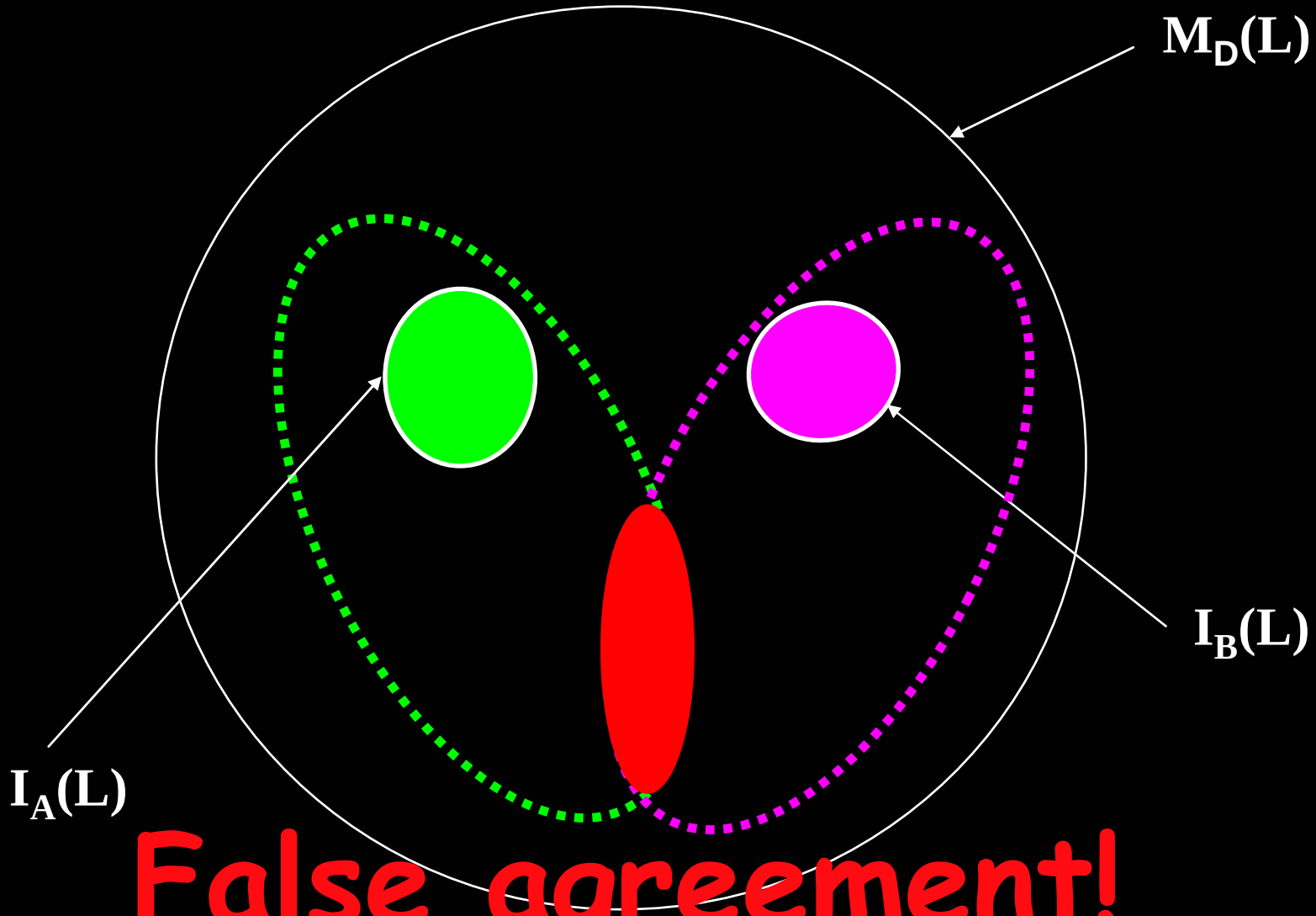


Low precision and coverage

Levels of Ontological Precision



Why precision is important



False agreement!

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

Different uses of ontologies

- Simple semantic access *(Processing time)*
  - Intended meaning of terms ***known in advance*** within a community
  - Lightweight ontologies support only services relevant for the query
  - *Limited expressivity* (stringent computational requirements)

- Meaning negotiation and explanation *(Pre-processing time)*
  - ***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*** and rich axiomatization needed to exclude ambiguities
  - Only needs to be undertaken ***once***, before cooperation process starts

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**

Formal Ontological Analysis

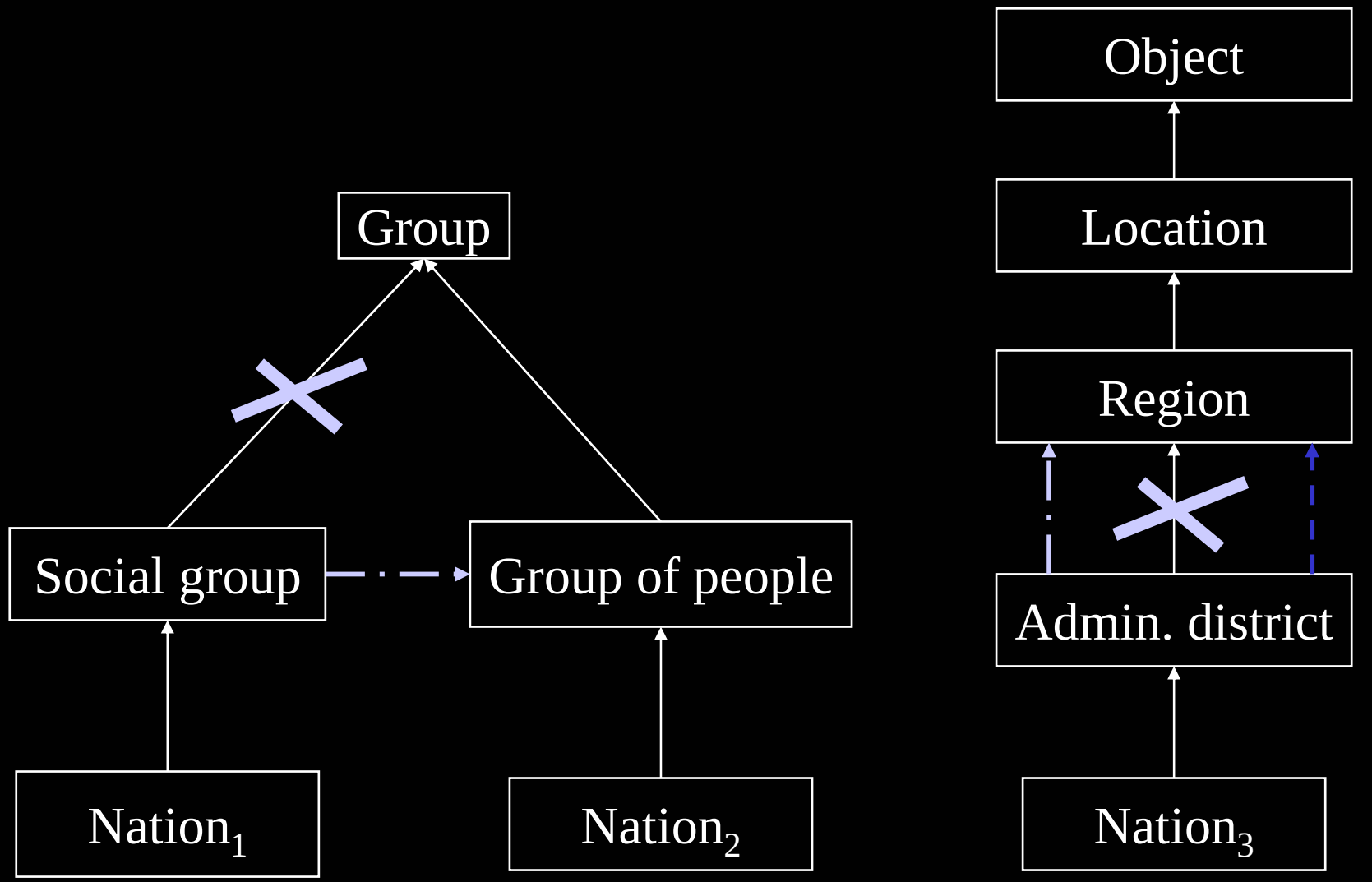
- Theory of Parts
- Theory of Wholes
- Theory of Essence and Identity
- Theory of Dependence
- Theory of Qualities
- Theory of Composition and Constitution
- Theory of Participation
- Theory of Representation

A common ontology vocabulary should be based on these theories!!

IS-A overloading

- *Overgeneralization:*
  1. A *physical object* is an *amount of matter* (Pangloss)
  3. An *amount of matter* is a *physical object* (WordNet)
  2. An *association* is a *group* (WordNet)
  4. A *place* is a *physical object* (μKosmos, WordNet)
  5. A passenger is a person
- *Clash of senses:*
  6. A *window* is both an *artifact* and a *place* (μKosmos)
  7. A *person* is both a *physical object* and a *living thing* (Pangloss)
  8. A *communicative event* is a *physical*, a *mental*, and a *social event* (μKosmos, Pangloss)

The case of "Nation"



The WonderWeb Library of Foundational Ontologies

- No single upper level
- Rather, a (small) set of *foundational ontologies* carefully *justified* and *positioned* with respect to the space of possible choices
- *Basic options* clearly documented
- Clear *branching points* to allow for easy comparison of ontological options)

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
    - …
  - …

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.

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

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)

Is the Semantic Web just Hype?

- Maybe yes.
- An "ontology vocabulary" is not enough
- Languages based on "semantic" primitives (OWL) are not enough (need for ***ontological* primitives**)

- …Unless the deep problems underlying ontology and semantics are attacked under an ***interdisciplinary*** approach

Europe is well ahead USA here…

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

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)