WonderWeb Deliverable D15

Ontology RoadMap

Stefano Borgo, Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari
ISTC-CNR
c/o ISIB-CNR, C.so Stati Uniti, 4
35127 Padova
Italy

email: {Nicola.Guarino, Claudio.Masolo, Alessandro.Oltramari}@ladseb.pd.cnr.it,
gangemi@ip.rm.cnr.it, stborgo@indiana.edu

Identifier	D15
Class	Deliverable
Version	1.0
Date	31-12-2002
Status	Final
Distribution	Restricted
Lead Partner	ISTC-CNR[1]

[1] National Research Council, Institute of Cognitive Sciences and Technology. The ontology group formerly at LADSEB-CNR has now joined this new institute. The rest of LADSEB-CNR has joined another new institute, ISIB-CNR (Institute for Biomedical Engineering). Formal changes to the WonderWeb's contract are in progress.

WonderWeb Project

This document forms part of a research project funded by the IST Programme of the Commission of the European Communities as project number IST-2001-33052.

For further information about WonderWeb, please contact the project co-ordinator:

Ian Horrocks
The Victoria University of Manchester
Department of Computer Science
Kilburn Building
Oxford Road
Manchester M13 9PL
Tel: +44 161 275 6154
Fax: +44 161 275 6236
Email: wonderweb-info@lists.man.ac.uk

Table of Contents

1 Introduction

Ontologies are considered nowadays a key element for semantic interoperability, and more generally for successful information exchange among people and machines in complex environments as the Web. Ontologies act as shared vocabularies for describing the relevant notions of a certain application area, whose semantics is specified in a (reasonably) unambiguous and machine-processable form. Describing this semantics, i.e. what is sometimes called the *intended meaning* of vocabulary terms, is exactly the job well developed ontologies do.

But what *kinds* of ontologies do we need? This is still an open issue. Some people believe that very general ontologies involving rich axiomatic characterizations are important, others think they are a waste of time, and prefer to concentrate on *lightweight* ontologies, focusing on the minimal terminological structure (often just a taxonomy) which fits the needs of a specific community.

As one may expect, we need different kinds of ontologies depending on the way we use them. Let us consider for instance the Semantic Web scenario, which is probably one of the most ambitious application perspectives for ontologies. Within such scenario, we can use ontologies in different ways. On the one hand, we can use them for *semantic access* to a certain resource within a specific community; in this case, the intended meaning of the terms used by the community is more or less known in advance by all members, and the ontology can be limited to those structural relationships among terms that are considered as relevant for the query (in many cases, taxonomic relationships are enough).

On the other hand, ontologies can be used for *meaning negotiation and explanation*:

- to *negotiate meaning* between (human or artificial) agents belonging to different (possibly related) communities;
- to *establish consensus* in a community that needs to adopt a new term; or simply
- to *explain* the meaning of a term to somebody new to the community.

This is a completely different task, which requires the explicit representation of *ontological commitment* in order to exclude terminological and conceptual ambiguities bound to unintended interpretations. In this case, a rich axiomatization (in addition to an adequate informal documentation) seems to be unavoidable.

Of course, building ontologies for meaning negotiation and explanation is extremely hard, both conceptually and computationally. However, it only needs to be undertaken *once*, before a cooperation process starts. The quality of a meaning negotiation process may drastically affect the *trust* in a service offered by the Semantic Web, but not the computational performance of the service itself. For example, a product procurement process involving multiple agents with distributed lightweight ontologies may be carried out in an efficient way by using simple terminological services, but the risk of semantic mismatch can be minimized only if the agents share a (more or less minimal) common ontology.

We shall use the term “foundational ontologies” for ontologies of the second kind above, ultimately devoted to facilitate mutual understanding. We see their role and nature as complementary to that of lightweight ontologies: the latter can be built with a moderate effort by exploiting machine learning techniques for mining extensive resources on specific domains; the former require more painful human labour, which can greatly benefit from the results and methodologies of disciplines such as philosophy and linguistics.

The variety of formal ontologies that have been proposed can be seen as a *library* of ontological systems coming from different areas; computer science, linguistics, cognitive science, and so on. Different systems reflect different commitments and purposes, but – this is the challenge – they should not form isolated items in a repository, like for instance the glorious Stanford's Ontolingua library established almost ten years ago (see ontolingua.Stanford.edu): rather, they should be somehow organized according to carefully isolated fundamental ontological options, and linked together depending on formal relationships among such options. Rationales and alternatives underlying the different ontological choices should be made as explicit as possible. In this way, a network of different but systematically related ontologies become available and people working on applications could easily find those systems that better capture their ontological needs. Such an architecture, we believe, would contribute to make people (and computers) commit to common ontological agreements as well as *understand the reasons of disagreement*, which seem to us to be more effective than enforcing interoperability by means of a single overarching ontology.

1.1 Deliverable structure

The structure of this Deliverable is as follows. In the next section we shall describe the main characteristics of foundational ontologies. Then we introduce some of the most important and basic ontological choices. Using these, we will discuss the differences between the main foundational ontologies in the literature: OpenCYC, the public upper level of the CYC project (see www.opencyc.org); DOLCE, an upper level developed in the framework of the WonderWeb project (see wonderweb.semanticweb.org); SUMO, an upper level developed with the contribution of the IEEE working group on Standard Upper Ontology; BFO, an upper ontology developed at IFOMIS (see www.ifomis.uni-leipzig.de).

2 Foundational ontologies

As we have seen in the previous section, the ultimate role of foundational ontologies is to facilitate mutual understanding. Let us explore in this section the technical implications of such perspective, and the practical role that foundational ontologies can have.

The main purpose of an ontology is to specify the *intended meaning* of a vocabulary, i.e. its underlying conceptualization [Guarino 1998]. As we have seen, such intended meaning is largely implicit for lightweight ontologies, which focus mostly on relevant relationships existing within the terms used by a given community. The meaning of the terms themselves is taken for granted and known to the users. Things are different however in the case of ambiguous or general terms belonging to everyday natural language, or when computerized agents belonging to different communities need to communicate.

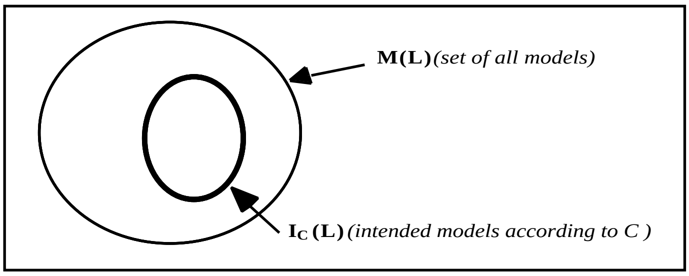

Figure 1. $\mathbf{I_C}$(**L**) is the set of *intended models* of **L** according to some conceptualization **C** of a domain of interest **D**. It is a subset of the set **M(L)** of all models of **L** with domain **D**. Adapted from [Guarino 1998][1].

[1] With respect to [Guarino 98], notice the further clarification that the set of all models is relative to a specific domain, namely the conceptualization's domain.

In this case, it is necessary to characterize the intended meaning of a vocabulary by means of a suitable *axiomatization*, which aims at excluding the unwanted models (**Figure 1**). Notice that, in general, it is not possible to *define* the intended meaning in this way, but just to *approximate* it by excluding at least some unwanted interpretations (**Figure 2**).

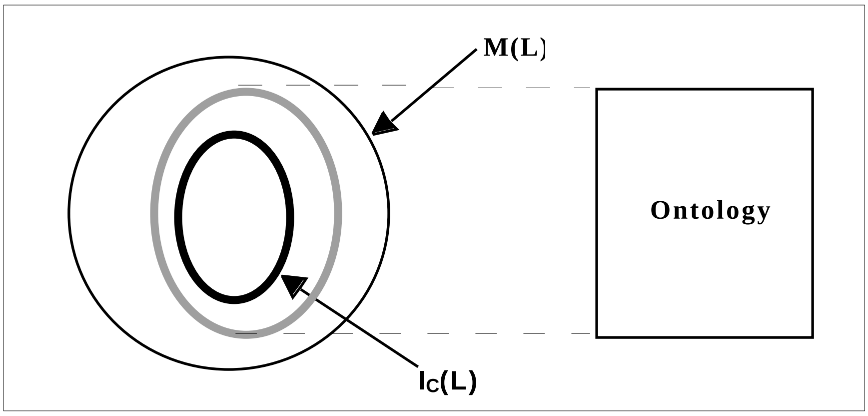

Figure 2. An ontology for a language L is a specification of a conceptualization C (relative to some domain D) if the set of its models relative to D is as close as possible to the set of intended models I_C (L). Adapted from [Guarino 98].

The scheme of **Figure 2** allows us to introduce two important dimensions for evaluating and comparing ontologies with respect to a certain conceptualization: *completeness* and *precision*.

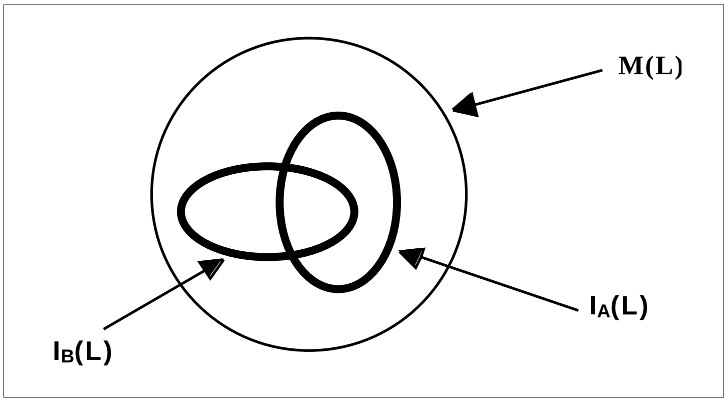

Figure 3. Two agents **A** and **B** using the same language **L** can communicate only if the set of intended models $\mathbf{I}_A(\mathbf{L})$ and $\mathbf{I}_B(\mathbf{L})$ associated to their conceptualizations overlap. Adapted from [Guarino 1997a].

Let us see now how these notions can help to understand the role foundational ontologies may have in meaning negotiation and explanation. Consider **Figure 3** above, which depicts a situation where two agents, **A** and **B**, use the same language **L** but two different conceptualizations of the same domain **D**. Effective communication is possible only if the intended models associated with the two conceptualizations do overlap (as shown in the figure), i.e. the two conceptualizations are mutually consistent. In the situation shown, a possible strategy could be to adopt an ontology that covers both conceptualizations[2]; in this case, the correctness is guaranteed, but the precision is low,

[2] We say that an ontology covers a conceptualization if the set of its models includes the intended models related to that conceptualization.

and the ontology cannot help discriminating or explaining differences in intended meaning. Another strategy is to design an ontology, which is precise enough to cover only the area of intersection: such an ontology would impose to the two agents a common, disciplined interpretation of the vocabulary in order to enable the cooperation. But what happens if the two sets of intended models do not overlap? Consider for instance Fig. 4, which depicts a worst-case scenario where each agent commits to a poor, largely imprecise ontology, whose models are delimited by the big dotted ovals, and these two sets intersect while the intended models do not. This means that a bottom-up approach to systems integration based on the integration of multiple local ontologies may not work, especially if the local ontologies are not precise enough, maybe because they just include those axioms that are relevant to a specific *context*, and therefore are only weak and *ad hoc* approximations of the intended models. The risk in this case is that of *false agreement*: the local ontologies are mutually consistent, but they agree on some models which are not among the intended ones. A way to reduce this risk is assuming that the two local ontologies both commit to a third, more general ontology whose intended models (the green oval in the picture) include as special cases those of the two agents. That is, we are assuming that the two agents, although disagreeing on the intended meaning of certain specific terms, still share some common understanding of some more general terms. If the general ontology is precise enough, the result will be that some of the false agreement models will be now excluded.

In reality, we do not find clear-cut examples as the one described here. Nonetheless, we believe such an example is useful to make explicit one of the roles of foundational ontologies: reducing the risk of false agreement, showing therefore the reasons of disagreement, as well as the conditions for mutual agreement.

The role of *foundational ontologies*

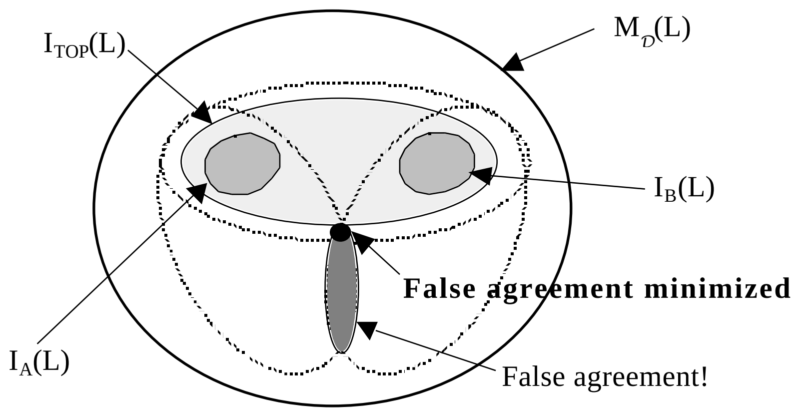

Figure 4. The sets of models of two different axiomatizations, corresponding to different ontologies, may intersect while the sets of intended models do not.

Since an ontology only indirectly accounts for a conceptualization, we often want to measure *how close it can get*. In fact, we can classify ontologies according to the *accuracy* they characterize the conceptualization they commit to. Perhaps surprisingly, completeness and precision are not enough to characterize the *quality* of a foundational ontology, i.e. its ability to specify the intended meaning of a vocabulary. The reason lies in the difference between the *logical* notion of (intended) *model* and the *ontological* notions of *state of affairs* and *situation.*

In a way, the notion of intended model is an artificial one, it is bound to two choices that have an intrinsic degree of arbitrariness: the choice of the *domain of discourse*, and the choice of the *logical primitives* on which the axiomatization is based. On the other hand, the notion of conceptualization, according to the definition provided in [Guarino 1998], relies on an intuitive notion of *situation*, intended as a maximal state of affairs involving the elements of the domain[3].

There are two possible ways to improve the approximation of a conceptualization: by developing a richer axiomatization, and by adopting a richer domain and/or a richer set of relevant conceptual relations. In the first case, the difference between the set of ontology models and the set of intended models is reduced. In the second case, it is possible – at least in principle – to include in the set of relevant conceptual relations (some of) those relations that characterize a world state, extending at the same time the domain in order to include the entities involved by such relations: for instance, in the toy example of the blocks world, we may consider the spatial location of a block as a relevant conceptual relation, including therefore locations in the domain, and considering a relation like $on(a,b)$ as completely derivable from the locations of a and b. Since every model now carries the information concerning the state of the world it refers to, the underlying conceptualization can be reconstructed from the set of its intended models. In this case, if an ontology is axiomatized in such a way to have exactly the same models, then it would be a “perfect” ontology.

Another way to increase the accuracy of an ontology consists of either adopting a modal logic, which allows one to express constraints across worlds, or reifying worlds as ordinary objects of the domain.

Of course, there is a tradeoff between a coarse and a fine-grained ontology committing to the same conceptualization: the latter gets closer to specifying the intended meaning of a vocabulary (and therefore may be used to *establish consensus* about sharing that vocabulary, or a knowledge base which uses that vocabulary), but it may be hard to develop and to reason on, both because of the number of axioms and the expressiveness of the language adopted. A coarse ontology, on the other hand, may consist of a minimal set of axioms written in a language of minimal expressivity. Such an ontology may support only a limited set of specific services, intended to be shared among users which *already agree* on the underlying conceptualization. We can distinguish therefore between detailed *reference ontologies* and coarse *shareable ontologies,* or maybe between *off-line* and *on-line ontologies:* the former are only accessed from time to time for reference purposes, while the latter support core system’s functionalities.

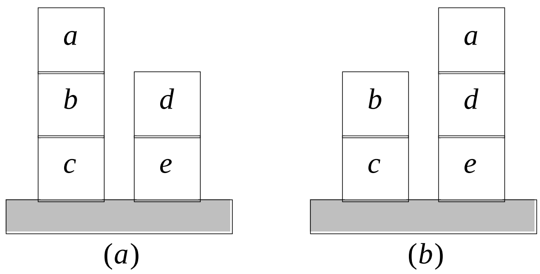

Figure 5 (a) A possible arrangement of blocks. (b) A different arrangement.

3 State of the art in foundational ontology design methodologies

We can envision two main areas of interest when talking about methodologies for foundational ontologies. On the one hand, there are many technical issues related to building, merging, integrating, and evaluating existing ontologies. On this subject, we address the reader to [Corcho et al. 2001], the deliverable 1.1 of the OntoWeb Project,

[3] Technically, a situation is usually intended as a non-maximal state of affairs, while a world is a maximal state of affairs. However, we shall use these two terms as synonymous.

where these aspects have been extensively analyzed.

In this section we concentrate only on the most important foundational ontologies presented in literature comparing their basic assumptions in ontological analysis and highlighting the general methodologies adopted, leaving aside the problems linked to coding, purposes, maintenance, and the like, that is, the engineering aspects of ontological research.

3.1 General attitudes toward ontological analysis

Before addressing specific issues about domain of discourse and basic categories and relations[4], it may be important to clarify the general attitude towards ontological analysis, or – in other words – the constraints and the motivations that influence the *conceptualization* of reality. In this respect, two main ontological choices highlighted in the philosophical literature are: a) *descriptive* vs. *revisionary*, and b) *multiplicative* vs. *reductionist*.

(a)A descriptive ontology aims at describing the ontological assumptions behind language and cognition by taking seriously the *surface structure* of natural language and commonsense. Under this approach, ontological categories are postulated in a rather unrestricted way, independently of evidence coming from other areas like physics or astronomy. A revisionary ontology, on the other hand, gives less importance to linguistic and cognitive aspects, and does not hesitate to suggest paraphrases of linguistic expressions or re-interpretations of cognitive phenomena in order to avoid ontological assumptions considered debatable on scientific grounds.

To give an example: commonsense distinguishes between *things* (*spatial objects*) like houses and computers, and *events* (*temporal objects*) like bank transfers and computer repairs. In the wake of relativity theory, however, time is only another dimension for objects and some philosophers and computer scientists have come to believe that the commonsense distinction between things that are and things that happen should be abandoned for a view according to which everything extends in space *and* time.

In a descriptive ontology, the categories refer to cognitive artifacts more or less depending on human perception, cultural imprints and social conventions. Thus, such an ontology maintains the distinction between things and events. A revisionary ontology, by contrast, is committed to capture the intrinsic nature of the world, its objective structures independently from how an agent conceptualizes it. As a consequence, an ontology of this type may impose that only entities extended in space *and* time exist.
Classic examples of descriptive ontologies are [Strawson 1959] and [Searle 1983]**)**.

(b)In designing ontologies, one wants to have a rich array of basic concepts from which to build the system. On the other hand, there are considerable advantages in using small taxonomies. A reductionist ontology adopts this latter approach; he aims at describing a great number of ontological differences with the smallest number of concepts. Instead, in a multiplicative ontology expressivity is more relevant: the aim is to give a reliable account of reality despite of the need of a larger number of basic concepts.

A clear example is the attitude towards co-localized entities. A multiplicative ontology allows for different entities to be *co-localized* in the same space-time. These entities are *assumed* to be different because they have incompatible essential properties. The standard case is that of the vase and the amount of clay it is made of: the vase does not survive a radical change in shape or topology, while the amount of clay does. The multiplicativist concludes that these must be different entities, yet co-located: the vase is *constituted* by an amount of clay, but it is not an amount of clay[5]. When a vase-master

[4] See [Gangemi et al. 2001].
[5] One of the purposes of the OntoClean methodology [Guarino and Welty, 2002] is to help the user

shapes a particular amount of clay, new properties are instantiated thus justifying the *emergence* of a new entity: the vase. However, a reductionist ontology postulates that each space-time location contains at most one object: incompatible essential properties are regarded as being linked to different points of view from which one can look at the same spatio-temporal entity. The vase and the clay are surely different, the reductionist claims, although not as entities but as views of the same spatio-temporal object.

The problem of representing time and modality aiming at modal and temporal reasoning is an old and ever recurrent quandary in artificial intelligence. Basically, two approaches are possible: either one includes modal and temporal operators in the formal system from the very beginning, or reproduces modal reasoning into a first-order language adding time and world (or situation) parameters to the predicates. In the first case one can translate literally into the formalism the expression "It is possible that John is ill". In the latter one can rephrase the expression by "There is a world in which John is ill".

These options are well known to the practitioner. What is less known is that the decision for one or the other approach is often connected to the ontological choice between actualism and possibilism (or presentism and eternalism). Actualism claims that only what is real exists, while possibilism admits as existing also possibilia (situations or worlds). Similarly, presentism assumes that only what is present exists, while for an eternalist the past, the present and the future are all existing.

The decision to allow quantification over instants or worlds is a decision faced by the possibilism or eternalism approach, while actualism and presentism go hand in hand with the use of primitive modalities.

3.2 Basic Assumptions and Methodologies

We begin with a brief description of the basic assumptions and methodologies considered in the most important upper ontologies, namely DOLCE, OPENCYC, SUMO, and BFO, analyzing the different terminologies and structures adopted.

DOLCE

History

DOLCE has been developed by the research group at ISTC-CNR (previously at LADSEB-CNR) as a reference module for a library of ontologies (mainly in the context of the WonderWeb Project). The formal structure implemented in DOLCE is an enhanced application of the OntoClean methodology, whose task was to isolate formal constraints for ontology-building ([Oltramari et al. 2002]).

General attitude and basic assumptions

As reflected by its acronym, DOLCE (a Descriptive Ontology for Linguistic and Cognitive Engineering) has a clear *cognitive bias*, in the sense that it aims at capturing the ontological categories underlying natural language and human commonsense. DOLCE embraces the so-called *multiplicative approach*: starting from the observation that one tends to associate objects to incompatible essential properties, DOLCE provides a clear and detailed treatment of objects and properties assuming that different entities can be *co-located* in the same space-time. According to the general methodology introduced in [Gangemi *et al.* 2001], suitable primitive formal relations, such as parthood, dependence and constitution [Masolo *et al.* 2002], have been introduced to characterize DOLCE ontological commitments as neutrally as possible. The use of formal relations is very important in this field: they may be applied to multiple subdomains, depending only on their "relata", which means that there is no commitment to the entities linked by these relations. DOLCE provides also a deep axiomatisation, which is important for the exploitation of ontologies in the area of the Semantic Web. Furthermore, to employ

evaluating ontological choices of this type.

DOLCE in the field of linguistic resources, an alignment with WordNet has been provided [Gangemi *et al.* 2002], although limited to links between ontological categories and synsets for nouns.

Terminology and Presentation

DOLCE uses terms and expressions diffused in the philosophical debate and in linguistics: philosophy-oriented categories are `Perdurants`, `Endurants` and `Abstracts` (see i.e. [Loux 1998]), while the partition of Perdurants into, `Accomplishment`, `Achievement`, `State` and `Process,` uses terms from linguistics (i.e. [Frawley 1992]). Categories are always introduced as partitions of their relative upper node. As a consequence, DOLCE has a tree-structure revealing a "top-down" methodology

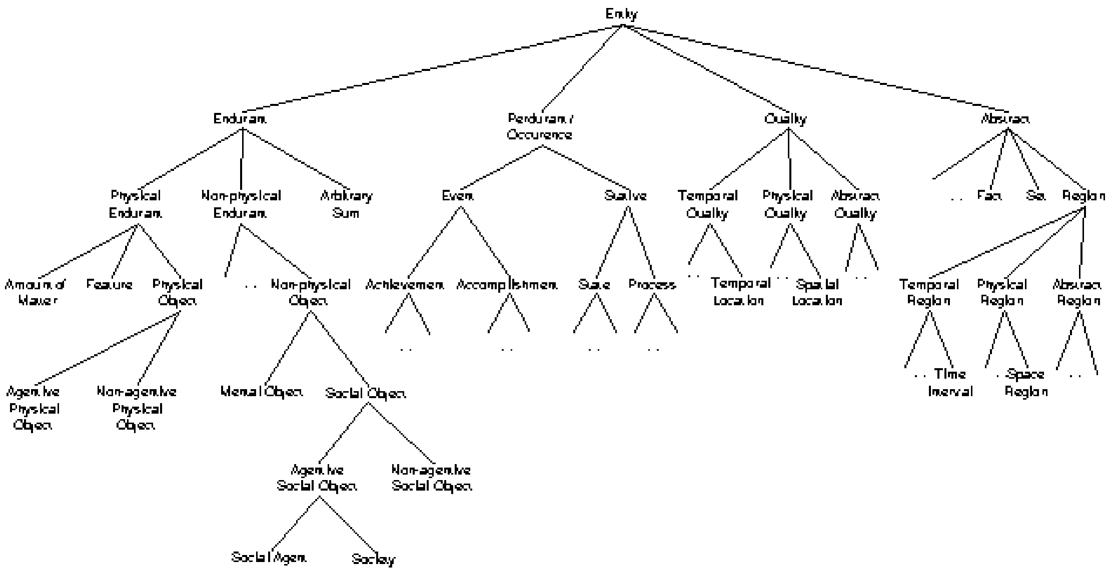

Figure 6. The Taxonomy of Dolce v2.0 Basic Categories

OpenCyc

History

The CYC project started in the mid eighties, when Douglas Lenat joined the MCC (Microelectronics and Computer Consortium), to develop a computer program "equipped" with a large amount of commonsense knowledge[6].

The name "CYC" derives from "encyclopedia", even though the intent of the project is not to build a sort of electronic encyclopedia, but rather a "complement" of it: «*one day* CYC *ought to contain enough commonsense knowledge to support natural language understanding capabilities that enable it to read through and assimilate any encyclopedia article, that is, to be able to answer the sorts of questions that you or I could after having just read any article, questions that neither you nor I nor* CYC *could be expected to answer beforehand*» [Guha and Lenat 1990]. In order to achieve this ambitious goal, Cyclists began to "prime the knowledge pump" of the program with millions of everyday terms, concepts, and rules constituting the human naïve knowledge of reality. Due to problems of consistency within a unique huge knowledge base, the information stored in CYC has been carved according to hundreds of contexts or "microtheories", namely bundles of assertions that share common assumptions about the world (a microtheory, in CYC terms, usually concern a specific domain of knowledge)[7].

[6] At the end of 1994, a new company called Cycorp took the place of MCC.
[7] Each microtheory is consistent, but there might be contradictions among different microtheories.

However it is not clear how CYC actually solves the problems of consistency when dealing with several microtheories at a time.

CYC is a huge system with many interesting elements and a complex architecture. In what follows we focus on the “semantic heart” of the project, namely the so-called OPENCYC ontology, which is the public version of the overall ontology.

General attitude and basic assumptions.

OPENCYC ontology appears to be deeply affected by cognitive assumptions, since its categories try to capture naïve conceptions of the real world, that is, as we said above, the human fund of common sense knowledge. For this reason we could consider OPENCYC ontology neither a reductionist nor a eliminativist account of what exists, but rather as the product of a multiplicative conceptualisation of reality, also known as “non-reductive pluralism” in the philosophical literature. OPENCYC makes available about 5000 concepts and 50000 axioms, subdivided in microtheories. It supplies minimal general categories (see fig. 2), and furnishes links to natural language, exploiting in particular WORDNET’s synset structure[8]. Unfortunately, it must be said that the characterization of the commitments on underlying ontological choices seems to be a secondary task in the current state of OPENCYC project: the documentation is still sketchy, and as a consequence there is lack of references to the established literature. Nevertheless, top categories and their mutual relations are explained in relevant sections of OPENCYC website (www.opencyc.com), allowing us to perform a close examination of the upper level of the ontology in comparison with DOLCE.

Terminology and Presentation

OPENCYC, on the contrary of DOLCE, does not stress very much the terminological aspect, since it uses common-language terms and locutions, like `SomethingExisting`, `Situation`, `Tangible/IntangibleThing`, and so on[9].

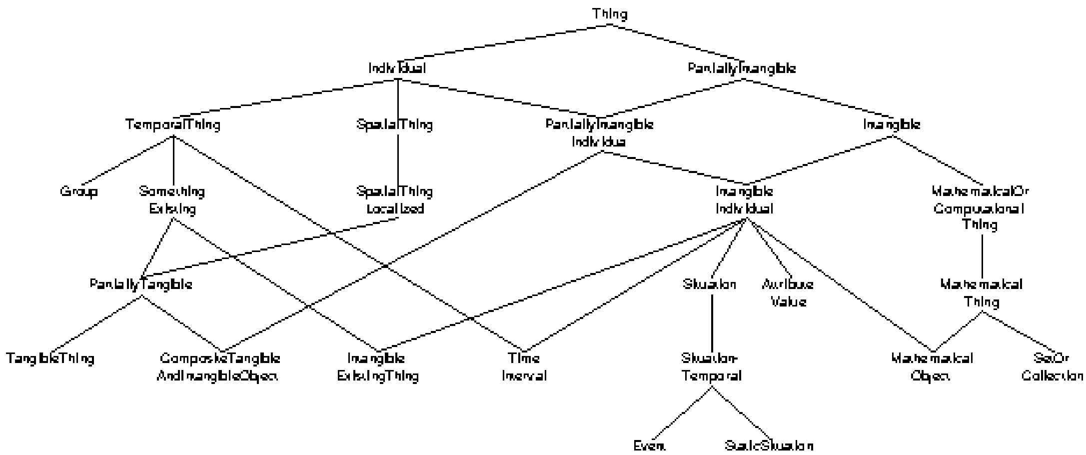

Figure 7. The Taxonomy of OpenCyc Basic Categories

In OPENCYC multiple inheritance has been broadly employed to supply an “attribute-like” characterization of lower categories[10] (i.e., a `MathematicalThing` is both a

[8] The most recent documentation (see http://www.opencyc.org/doc/#WordNet) reports that «*the Cyc-WordNet Linking Tool allows users to state links between WordNet synsets and Cyc constants* [categories]. *The interface also allows the user to browse WordNet, and to see existing Cyc-WordNet links*».

[9] At a first sight, some of OPENCYC’S categories sound a bit strange, like `Something_Existing`, that seems a kind of catch-all-entity, or `Attribute_Value`.

[10] A class subsumed by several super-concepts takes specific features from any of them, therefore

MathematicalObject and an IntangibleIndividual; a TimeInterval is both a TemporalThing and an IntangibleIndividual – see fig. 2) and then OPENCYC has a graph-like structure (compare fig.6 and fig.7) revealing a "bottom-up" approach applied to a given knowledge base.

Sumo

History

The Suggested Upper Merged Ontology (SUMO) has been created by the IEEE Standard Upper Ontology (SUO) working group as an attempt to link categories and relations coming from different top level ontologies (see [Pease 2002]) in order to improve interoperability, communication and search in the Semantic Web area. SUMO is divided in 11 "sections", kinds of complementary ontologies clustering categories by relevant topics: the Mereotopology ontology, for example, contains concepts that deal with the formalization of part/whole relation (see [Casati and Varzi 1994]), while the Unit of Measure ontology isolates definitions for unit systems. In the 4th paragraph we briefly analyse the so-called "Base ontology" section, namely the upper level layer containing the most basic distinctions.

General attitude and basic assumptions

Because of its characteristic merging of different upper level ontologies, SUMO is actually not influenced by a specific theoretical approach, rather it tends to take from various ontological proposals those general categories which seem to be largely shared by the computer science community. In this context, we should say that SUMO does not clearly adopt neither a multiplicative nor a reductionist approach. It provides a quite rich axiomatisation (see the documentation available on the web[11]) and it has been linked with WordNet 1.6 (probably the very standard lexical database used for populating ontologies).

Terminology and Presentation

From the terminological point of view, as you can understand considering the previous paragraphs, SUMO collects locutions and terms used in distinct top level ontologies. It has a tree-like structure, revealing a top-down methodology like DOLCE.

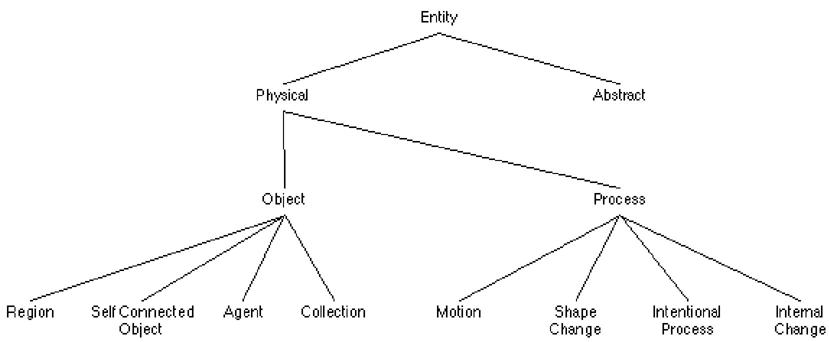

Figure 8. The Taxonomy of Sumo v1.36 Basic Categories

Bfo

BFO (Basic Formal Ontology) has been built by the research group of the Institute for Formal Ontology and Medical Information Science (IFOMIS) headed by Barry Smith at the University of Leipzig. Here we are not presenting a detailed overview of this upper

acquiring, in a sense, different "attributes".

[11] http://ontology.teknowledge.com/#docs

level ontology, mainly because it is still in a developmental phase (categories are fixed, but the axiomatisation and the alignment with lexical resources are still an on-going work).

BFO consists of sub-ontologies possibly linked together by formal relations. Every sub-ontology represents a particular view on reality, according to the level of granularity chosen by the modeler in order to "catch" determinate aspects of the world. For example, SNAP-BFO and SPAN-BFO are respectively the ontology for endurants and the ontology for perdurants, the two most important layers further sub-ontologies must be linked to. The former provides a list of all the entities existing *in* time ("a snapshot of reality"[12]), like cars, animals, mountains, while the latter is a "catalogue" of events (which occurs *through* time), like races, deaths, avalanches. Indeed, SNAP entities *participate* in SPAN entities (as endurants participates in perdurant according to DOLCE).

BFO reflects a realistic bias, and its terminological choices (Quality, Substance, Process, and so on) have to be read in this conceptual framework.

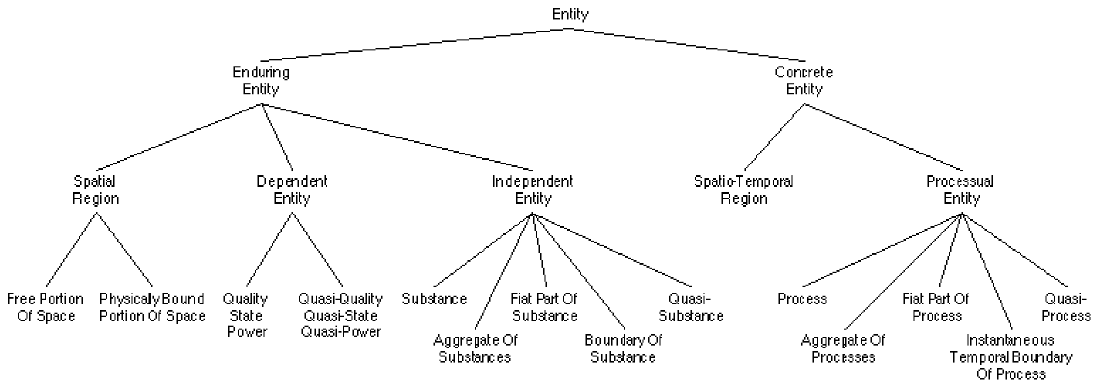

Figure 9. The Taxonomy of Bfo Basic Categories

4 Roadmap of major ontological choices

In this section we concentrate on the analysis of some important ontological choices adopted in DOLCE, OPENCYC, SUMO and BFO. We do not aim at an exhaustive and accurate examination of the four top level ontologies, rather we present a diffuse discussion of these systems with respect to some relevant ontological distinctions introducing the theoretical problems and the solutions adopted in these systems.

4.1 Universals, Particulars and Individual Properties

The ontological distinction between *universals* and *particulars* can be characterized by means of the primitive relation of *instantiation*: particulars are entities that *cannot* have instances; universals are entities that *can* have instances[13]. In linguistic, 'proper nouns' are normally considered to refer to particulars, while 'common nouns' to universals. For example, 'Varenne', the Italian racehorse, is an instance of 'horse', but it cannot be instantiated itself.

This characterization of the concept of universal is still vague since it does not clarify whether sets, predicates, and abstract should be considered as universals.

Sets are extensional entities, i.e. fully determined by their extension, and so is the *membership* relation: two sets are identical if and only if they have the same members.

[12] http://ontology.buffalo.edu/bfo/BFO.pdf
[13] Properties and relations are usually considered as universals.

The relation of instantiation is more generic and usually taken to be non-extensional. For example the two universals 'three-angled polygon' and 'three-sided polygon,' are considered different although they have exactly the same instances.

Predicates are sometimes closed with respect to the logical connectives, i.e. if *P* and *Q* are predicates, also '*P and Q*', '*P or Q*', and '*not P*' are predicates. This seems awkward for universals, for example one would probably count *table* and *pumpkin* as universals, but not predicates like '*table or pumpkin*' or '*not table*'.

If abstracts are entities non extended in space-time (see also the next paragraph), then they can differ from universals in many aspects: (*a*) not all abstracts are universals (like numbers, sets, etc.); (*b*) universals can be localized in space-time since they are sometimes associated to the spatio-temporal locations of their instances[14]. More radically, universals can be rejected as in the *trope theory* [Campbell 1990]. Tropes do not have instances, they are properties/qualities of specific material entities and depend ontologically on them. In trope theory, it is possible to speak of the 'whiteness' of this specific piece of paper, while the universal 'white' does not exist. In addition tropes are often considered as localized in the space-time of (the surface of) the material entities they depend on, and they cannot be considered as abstracts in the usual way.

A clear-cut ontological choice is whether universals are included in the domain[15]. DOLCE and SUMO are examples of a foundational ontologies of particulars that does not classify universals. Instead, OPENCYC and BFO admit both particulars and universals in the domain. The category `SetOrCollection` in OPENCYC furnishes a clear example.

4.2 Abstract and Concrete Entities

We have mentioned that *abstracts entities* exist neither in space nor in time, i.e. they are not localized. On the other hand, *concrete entities* (or *concretes*) are defined as entities that exist at least in time. Mathematical objects (like numbers and sets) are examples of abstracts, while ordinary objects (like cars, books, etc.) or events (like the 2000 Olympiad) are examples of concrete entities. This characterization immediately raises a question: how is it possible that abstracts exist without existing at any time? Is it better to say that these are eternal and immutable, i.e., they exist at all times without changing? From an ontological point of view the answer is not trivial, and perhaps a weaker characterization is preferable. An alternative definition is based on the 'causal criterion': abstracts possess no causal power while concretes do. This second definition, although similar to the first, is quite different: if abstracts are 'timeless' entities, as in the first definition, then they cannot be involved in causal relations; vice versa it is possible to individuate entities localized in time and space (like 'the centre of mass of the solar system' see [Lowe 1998]) that lack any causal power. In what follows, we focus on the first characterization of abstracts[16].

In DOLCE, temporal and/or spatial localization is not defined on `Abstract Qualities` and on `Abstracts`. Regarding the localization in time, in space or in other kinds of `Regions`, DOLCE distinguishes between 'direct' and 'indirect' localization. Some entities do not have a direct localization but they *inherit* their localizations from entities on which they depend. For example, `Endurants` (like tables

[14] In this case, the location of a universal is the sum of the locations of its instances and, according to this philosophical stand, every universal is 'wholly present' in each instance. This thesis is controversial. The difficulty of understanding how there can be a class of entities extended spatio-temporally but not behaving like particulars, remains unsolved.
[15] Even when universals are excluded from the domain of quantification, they do occur indirectly to organize and structure the domain.
[16] The sense of abstractness introduced here is different from the one used in trope theory. Here concrete entities are 'material' (such as cars, tables, etc.), while tropes are properties or qualities of these entities (possibly with a spatio-temporal location).

or persons) inherit their temporal localizations from the temporal localizations of the `Perdurants` in which they participate. `Perdurants` (like a walking through the park or listening to music) have a direct temporal localization. Considering both 'direct' and 'indirect' localization, `Endurants`, `Perdurants`, `Temporal Qualities`, and `Physical Qualities` are all concrete entities.

We have seen in OPENCYC that instances of `SetOrCollection` do not have spatial or temporal locations, thus corresponding to abstracts. Among the subdivision of `Individuals`, `TemporalThings` are localized at least in time while `SpatialThings` are localized at least in space. It is not clear if there exist instances of `SpatialThing` that are not instances of `TemporalThing`, i.e. individuals that are localized in space but not in time. If not, then all `SpatialThings` (like all `TemporalThings`) are concretes.

In SUMO, the distinction between `Physicals` and `Abstracts` is very similar to the distinction between concretes and abstracts. `Physicals` are entities "that have a location in space-time", and `Abstracts` "cannot exist at any particular place and time without some physical encoding or embodiment". Note that the notion of `Physical` seems to be a specialization of the notion of 'concrete' since it seems not possible to have `Physicals` that are localized only in time. As seen in OPENCYC, in SUMO locations are themselves understood to have location in space-time.

From the available documentation, the BFO ontology considers only entities existing in space and time, i.e. only 'concrete' entities. This kind of entities includes also `Spatio-TemporalRegions`, i.e. regions occupied by a process.

4.3 Endurants and Perdurants

Classically, *endurants* (also called *continuants*) are characterized as entities that are 'in time', they are 'wholly' present (all their proper parts are present) at any time of their existence. On the other hand, *perdurants* (also called *occurrents*) are entities that 'happen in time', they extend in time by accumulating different 'temporal parts', so that, at any time *t* at which they exist, only their temporal parts at *t* are present[17]. For example, the book you are holding now can be considered as an endurant because (now) it is wholly present, while "your reading of this book" is a perdurant because, your "reading" of the previous section is not present now. Note that it is possible to distinguish between 'ordinary objects' (like the book) and 'events or process' (like 'the reading of the book') even when assuming in the domain only perdurants. In this case, one has to rely on properties that lie outside spatio-temporal aspects. This is particularly clear in the assumption of the existence of different four-dimensional entities, which have the same spatio-temporal location like a person and its life.

Another way of characterizing endurants and perdurants has been proposed recently by Katherine Hawley: something is an endurant if and only if (*i*) it exists at more than one moment and (*ii*) its parts can be determined only relatively to something else (for instance time) [Hawley 2001]. In other words, the distinction is based on the different nature of the parthood relation: endurants need a time-indexed parthood, while perdurants do not. Indeed, a statement like "this keyboard is part of my computer" is incomplete unless you specify a particular time, while "my youth is part of my life" does not require such a specification[18].

---

[17] Time-snapshots of perdurants (i.e., perdurants that are present only for an instant, and which lack proper temporal parts) are a limit case in this distinction.

[18] Accepting in the domain both 'objects' and 'events', without possibility of definition (reduction) of one kind of elements on the other kind, the *participation* relation represents a fundamental link between these two kinds of entities: objects can participates in events. For example, a person may participate in a discussion or a sword in a battle. Note that the participation relation is important also

DOLCE assumes a classical three-dimensionalist view, accepting both the concepts of `Endurant` and `Perdurant`. `Endurants` and `Perdurants` are concrete, ‘principally in space’ and ‘principally in time’, respectively. This means that, while the temporal location of a perdurant *p* is directly ‘linked’ to *p*, its spatial location is individuated by the spatial location of the endurants that *participate* in *p*. Similarly the temporal location of an endurant is individuated by the temporal locations of the perdurants it participates in. We have already observed that only `Physical Endurants`[19] (like cars, horse, hammers) have a ‘direct’ spatial-location, while `Non-physical Endurants` (the content of a written law) ‘inherits’ their spatial location from the `Physical Endurants` they depend on.

In OPENCYC, `SomethingExisting` (i.e. entities that remain relatively stable throughout their lifetimes) seems very close to the classical concept of endurant. `SituationTemporal` (that is the union of `Event` and `StaticSituation`) seems to correspond to the classical notion of perdurant. `Events` are things that we say are *happening*, or change in the state of the world and, thus, they seem close to `Events` in DOLCE, while `StaticSituations` are state of affairs between two or more things, persisting statically over some time interval, thus close to `Statives` in DOLCE. In any case, as in DOLCE, the instances of both `SomethingExisting` and `SituationTemporal` are necessarily extended in time but not necessarily in space.

In SUMO, `Objects` are opposed to `Processes`. `Processes` are classically characterized as “the class of things that happen and have temporal parts or stages”, while for `Objects` a less standard four-dimensional interpretation is accepted: “in a 4D ontology, an `Object` is something whose spatio-temporal extent is thought of as dividing into spatial parts roughly parallel to the time-axis”. In any case, note that in SUMO, `Objects` and `Processes` are considered as necessarily localized in the space-time.

In BFO this distinction is used in order to individuate two kinds of sub-ontologies. Snap-BFO contains only endurants (`Enduring Entities`) and Span-BFO contains only perdurants (`Concrete Entities`). Among the perdurants, the `ProcessualEntities` are dependent on endurants (in particular on the subclass of `Substances`).

4.4 Co-localized entities

No matter what one decides about the ontological status of space and time, one has the option to include spatially and/or temporally co-located objects. It is quite natural to accept objects temporally co-localized, but it seems problematic to embody spatially or spatio-temporally co-localized distinct objects. For example, is a hole different from the region of space it occupies? Is a statue different from the stuff it is constituted by? Is a person different from its body? Are there holes, or only *holed objects*? Are there statues or only *statue-shaped stuffs*?

This subject is extremely complex and involves rather difficult issues like identity through time, material constitution, essentiality, modality, etc. Without entering into details we try to make explicit the position of the upper ontologies with respect to the possibility of accepting spatial and/or temporal co-localization. We distinguish entities that are spatially co-localized with ‘material entities’ – for examples statues, persons, etc. – and

---

if one considers only perdurants (i.e. also in a four dimensionalist position) but maintains a distinction between objects and events.

[19] `Physical Endurants` do not coincide with ‘material object’ or ‘material endurant’, i.e. endurants composed by matter, because for examples a hole is not material but it is ‘physical’ in the sense of DOLCE.

entities that are dependent on 'material entities' although *not* spatially co-localized with them – for example holes, places, spots, shadows, etc. (see [Casati and Varzi 1994]) for a detailed treatment of these entities).

DOLCE uses spatial co-localization and dependence/constitution relations in order to 'stratify' its domain. For example, `Non-agentive Social Objects` (a national financial manoeuvre) depend on `Societies` (a parliament), which are constituted by `Social` Agents (congressmen). These depend on `Agentive Physical Objects` (human persons), which are constituted by `Non-agentive Physical Objects` (bodies). In turn, the latter are constituted by `Amounts of Matter` (biological tissues, blood, bones, etc.). All these different entities are spatially co-localized. Analogously, `Mental Objects` depend on `Agentive Physical Objects`, and so on. `Features` like holes, shadows are clearly distinct from `Space Regions` and depend essentially on entities that are not spatially co-localized with 'material things'. Other kinds of `Features`, like edges, relevant parts, etc., are instead co-localized with (parts of) 'material entities'.

OPENCYC has a weaker position with respect to DOLCE with respect to spatially co-localized entities. Sometimes there is a genuine multiplication. For example a statue and the stuff that constitutes it, are distinct `TangibleThings` and spatially co-localized. On the other hand, in this ontology books are made of stuff but they are classified as `CompositeTangibleAndIntangibleThings`, i.e. the particular book in my office desk consists of a 'material support' and a 'content'. Vice versa in DOLCE the material support and the content of a book are two distinct entities. The entities of the second kind are considered `IntangibleExistingThings`[20].

SUMO and BFO do not pay much attention to this distinction although for different reasons. SUMO was developed as a mixture of different approaches which look at the issue of co-localization from different perspectives. Thus, this ontology suffers from the heterogeneity of the basic theories it started with. It seems that BFO does not present cases which address this issue directly while the documentation available does not mention the problem.

5 Bibliography

Campbell, K. 1990. *Abstract Particulars.* Basil Blackwell, Oxford.

Casati, R. and Varzi, A. C. 1994. *Holes and Other Superficialities.* MIT Press/Bradford Books, Cambridge (MA) and London (UK). Revised paperback edition, 1995.

Corcho, O., Fernander-López, M., Gómez Pérez, A., OntoWeb Deliverable 1.1: Technical RoadMap v1.0, 30.11.2001.

Frawley, W. 1992. *Linguistic Semantics.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Gangemi, A., Guarino, N., Masolo, C., and Oltramari, A. 2001. Understanding top-level ontological distinctions. In *Proceedings of IJCAI-01 Workshop on Ontologies and Information Sharing*. Seattle, USA, AAAI Press: 26-33. http://SunSITE.Informatik.RWTH-Aachen.DE/Publications/CEUR-WS/Vol-47/.

Gangemi, A., Guarino, N., Masolo, C., and Oltramari, A. 2002. Restructuring Wordnet's top-level. *AI Magazine, fall 2002 (in press).*

[20] Note that some strange entities like 'waves' or 'lights' are considered in OPENCYC as `Events` (and then perdurants), while in DOLCE as `Features` (and then endurant).

Guarino, N. 1998. Formal Ontology in Information Systems. In N. Guarino (ed.) *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998*. IOS Press, Amsterdam: 3-15. See amended version at www.ladseb.pd.cnr.it/infor/ontology/OntologyPapers.html.

Guarino, N. and Welty, C. 2002. Evaluating Ontological Decisions with OntoClean. *Communications of the ACM*, **45**(2): 61-65.

Guarino, N. and Welty, C. 2002. Identity and subsumption. In R. Green, C. Bean and S. Myaeng (eds.), *The Semantics of Relationships: an Interdisciplinary Perspective*. Kluwer (in press).

Guha, R. V. and Lenat, D. B. 1990. CYC: a Mid-Term Report. *AI Magazine*, **11**(3): 32-59.

Hawley, K. 2001. *How Things Persist.* Clarendon Press, Oxford, UK.

Loux, M. J. 1998. *Metaphysics, a Contemporary Introduction.* Routledge.

Lowe, E. J. 1998. *The possibility of metaphysics.* Clarendon Press, Oxford.

Masolo, C., Gangemi, A., Guarino, N., Oltramari, A., and Schneider, L. 2002. WonderWeb Deliverable D17: The WonderWeb Library of Foundational Ontologies. .

Oltramari, A., Gangemi, A., Guarino, N., and Masolo, C. 2002. Restructuring WordNet's Top-Level: the OntoClean approach. In *Proceedings of OntoLex2002: Workshop on Ontologies and Lexical Resources*. Las Palmas, Canary Islands. *(Accepted for publication on AI Magazine)*.

Pease A., 2002. The Suggested Upper Merged Ontology. A Large Ontology for the Semantic Web and its Applications. Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web.

Searle, J. 1983. *Intentionality.* Cambridge University Press, Cambridge.

Strawson, P. F. 1959. *Individuals. An Essay in Descriptive Metaphysics.* Routledge, London and New York.