Some Organizing Principles For A Unified Top-Level Ontology

Nicola Guarino

National Research Council, LADSEB-CNR
Corso Stati Uniti 4, I-35127 Padova, Italy
guarino@ladseb.pd.cnr.it
http://www.ladseb.pd.cnr.it/infor/Ontology/ontology.html

Abstract

This paper presents a set of principles and distinctions used to isolate a taxonomy of top-level ontological concepts aimed to conciliate clarity with semantic rigour, generality and commonsense. The general perspective adopted is that of Formal Ontology.

Introduction

Currently, a number of efforts in the ontological engineering community are aimed to the development of large "top-level" ontologies, to be used for a variety of tasks (PANGLOSS, Penman Upper Level, CYC, MikroKosmos...)

An important challenge would be the realization of a unified top level ontology, which should be the result of the integration of the above cited top-level ontologies. I shall discuss in this paper some general organizing principles which can be of help in this integration effort, based on the quest for semantic rigour, clear ontological foundation, and cognitive plausibility.

The general perspective I have in mind is that of Formal Ontology (Guarino 1995, Smith 1995), which can be intended as the theory of formal distinctions between the possible elements of a domain, independently of their actual reality. I shall briefly present what I call the basic *conceptual tools* of formal ontology, showing how they can be used to design the overall structure of a top-level ontology.

Such conceptual tools are theories aimed to address the classical ontological issues of philosophical logics: parthood, integrity, identity, dependence. With the help of these theories, we can define a number of formal ontological properties which – combined in various ways – contribute to characterize the concepts used in an ontology.

The use of formal ontological distinctions inspired by philosophical research has been recently advocated by John Sowa (Sowa 1995), who proposes to view an ontology as organized around the boolean lattice induced by such distinctions. Despite its mathematical appeal[1], the problem of this proposal is that it is difficult to isolate a basic backbone, a sort of natural skeleton or "conceptual coat rack" (Woods 1986). On the other hand, many current top-level proposals, especially MikroKosmos and Pangloss, are based on a graph structure which *tends* too be a tree (in the sense that the authors seem to have tried to minimize multiple inheritances), at least for what concerns the topmost levels[2]. As an extreme approach in this sense, we may mention the proposal of (Bouaud *et al.* 1995), who argue that an ontology *should* always be a tree.

I present in the following a set of principles and distinctions I have used to isolate a preliminary taxonomy of top-level ontological concepts aimed to conciliate clarity with semantic rigour, generality and commonsense. Such a preliminary top-level ontology is reported in Appendix 2. I do not claim (yet?) it *must* be a tree, but it *turns out to be* a tree up to the depth of 5 or 6. The main principles can be summed up as follows:

1. Individuals are kept separate from relations. Since the latter themselves can be seen as (second-order) individuals, I adopt the philosophical distinction between *particulars* and *universals*.[3]
2. Not all the unary relations appearing in the ontology of universals correspond to explicit concepts in the top-level ontology of particulars. Such concepts are called *taxons*, and must be chosen according to suitable organization criteria.
3. Taxons correspond to *material* ontological categories (like *objects* and *events*), related to the particular way our world is structured due to its laws of nature, and not to the *formal* ontological properties (like *decomposability*, for instance) which do not depend on a particular set of laws of nature.
4. Taxons are all primitive. Formal ontological properties (many of whom are defined) contribute to characterize their meaning, in terms of necessary conditions.
5. Taxons are mainly organized according to their *identity criteria* and the dependence relations between them. Disjoint sets of identity criteria correspond to disjoint taxons. Identity criteria can be grouped in classes, corre-

1 Which however presents serious limits concerning the definitions of the properties used

2 CYC is an exception in this sense, since it starts tangling up from the very beginning...

3 The definition of these terms are discussed in the following. For a general reference on philosophical terminology, see for instance (Burkhardt and Smith 1991).

sponding to a stack of *ontological strata* linked together by a dependence relationship (Borgo *et al.* 1996).

6. Universals are organized along three dimensions: their role as organizing criteria (taxons) in the ontology of particulars; the nature of their domain, and the distinction between formal and material universals.

In summary, we argue that a top-level ontology should not simply be represented as a tangled taxonomy of concepts, each one corresponding to (more or less arbitrary) unary predicates: within these predicates, important distinctions can be done in order to isolate what have been called the "factors" of an ontology. These "factors" can be of two kinds: formal factors and material factors. A clear understanding of such factors may contribute to avoid contradictions and to increase the reusability of an ontology.

The tools of formal ontology

Theory of parthood. A theory of parthood is, in our opinion, at the basis of any form of ontological analysis. Relevant questions that must be addressed are:

- What does count as a part of a given entity?
- What properties does the part relation have?
- Are there different kinds of parts?

An important example of a theory of parthood is given by *extensional mereology*. Much work must be addressed however in order to come up to a satisfactory theory of *intensional mereology*, where integrity and identity are taken into account. See (Simons 1987) for a thorough reference to the problems of mereology.

Theory of integrity. A given entity (like the sum of two chairs) can have parts without being considered as a single whole. The theory of integrity studies the different ways of connecting together the different parts to form a whole. Relevant questions that must be addressed are:

- What does count as a whole? Why?
- In which sense are its parts connected? What are the properties of such connection relation?
- How is the whole isolated from the background? What are its boundaries?

Together, the theory of parthood and the theory of integrity form what may be called mereotopology (Varzi 1996).

Theory of identity. The theory of identity studies the conditions under which an entity can be considered as identical to another. Relevant questions that must be addressed are:

- How can an entity change while keeping its identity?
- What are its essential properties?
- Under what conditions does an entity loose its identity?
- Does a change of parts affect identity?
- Does a change of "point of view" change the identity conditions?

The last question is especially relevant in our case to distinguish between ontological strata. For instance, physical objects (or animals) are different from the matter they are made of since the identity criteria underlying the two points of view are different.

Theory of dependence. The theory of dependence studies the various forms of existential dependence involving specific individuals and the class they belong to. We refer here to the notion of "existence" as "ontological existence", not as "logical existence". In this sense, existence can be represented by a specific predicates rather than by a logical quantifier. Relevant questions that must be addressed are:

- Can an individual exist alone, independently of any other individual?[1]
- Can an individual exist independently of another indvidual belonging to a specific class?[2]
- Can an individual belong to a particular class without implying the existence of a different individual belonging to a different class?[3]

In our opinion, the notion of dependence is at the core of Peirce's distinction between Firstness, Secondness and Thirdness, as reported in [Sowa 1995]. However, the examples reported by Sowa are definitely not clear, mainly because they don't take into account the further distinction between particulars and universals (see below).

Theory of universals. In the last paragraph we have seen how formal ontology also needs to address distinctions among universals. In particular, the following formal properties of classes (monadic universals) turn out to be extremely useful[4] :

- *Countability*: A class is countable if, for all of its instaces, their parts are not instances of the class.
- *Rigidity*: A class C is rigid if, when Cx is true in a possible world, then Cx is necessarily true.
- *Foundation:* Implies dependence on another class. See footnote 3 above.
- The property of *introducing a specific identity criterion* [I have no single word for this. See the discussion below].

[1] If this is not true, we say that the individual is *rigidly dependent* on another individual.

[2] If this is true, we say that the individual is *generically dependent* on another individual.

[3] If this is true for all the instances of the class, we say that the class is *independent*, or *not founded.* Otherwise it is *dependent,* or *founded.*

[4] The definition reported are simplified. See (Guarino *et al.* 1994) for an account of countability and rigidity, and (Simons 1987) for an account of foundation.

Some basic organizing principles

Being clear about the domain.

An ontology is often presented as a structured collection of classes, supposed to be defined on a certain domain of discourse. Sometimes the exact nature of this domain is however not clear, with the result that the formal semantic interpretation of the ontology turns to be difficult or even impossible. Consider for example the natural language descriptions associated to the leaves of an ontology: it is often not clear whether they refer to real world individuals, word senses or abstract relations. Sometimes it is even difficult to tell whether a leaf node is an instance or a class.

Particulars and Universals. The top-level ontology presented in Appendix 2 addresses to separate domains, corresponding to the classical distinction between *Particulars* and *Universals.* According to the Aristotelian perspective, particulars are entities that "cannot be said of anything" (except themselves); they correspond to individuals existing either in the actual or in a possible world, and are divided into *objects* and *events.* Universals are entities which "can be said of something", usually regarded as (reified) unary or n-ary relations. In this case however it is important to clarify that we give them a Montague-style *intensional* semantics, as discussed in (Guarino and Giaretta 1995). This means that they are not just regarded as sets of tuples, but rather as functions from the set of all possible worlds to the set 2^{D^n}, where n is the "arity" or the relation and D is the domain where they are defined. *First-order universals* (like *red* or *event*) are defined on the domain of particulars, while *Higher-order universals* (like *color*) are (also) defined on the domain of first-order universals.

Notice that the *possible worlds* mentioned above are not mysterious entities, but they are assumed to be special instances of events, corresponding to global (maximal) *states* of the whole set of objects. By including possible worlds in the domain, we can have formulas of a modal flavour while sticking to standard first-order logic.

This brief notes should give an idea of the formal semantics I have in mind. As a final observation, notice that no "Top" node appears: this means that it is not possible to quantify on the union of the two domains within the same formula, avoiding some of the most common paradoxes.

Taxons and ontological properties

In a recent paper, Peter Simons compares two possible methodologies for ontology building, which he labels "top-down" and 'bottom-up" (Simons 1995). The top-level approach directly addresses – in an ad-hoc and piecemeal way – the most relevant ontological distinctions emerging from natural language and commonsense, while the bottom-up approach focuses on the most general, "topic-neutral" ontological primitives. In other words, the top-down approach focuses on *material* distinctions, while the bottom-up approach on *formal* distinctions. Formal distinctions apply to very general domains, while material distinctions only apply to specific subdomains. For instance, the distinction "atomic/compound" applies to the whole domain of particulars, while the distinction "materia/immaterial" only applies to physical objects.

In a sense, the methodology I propose here takes advantage of the two approaches, keeping them at the same time clearly separated: the ontology of particulars is structured around a backbone of "basic" material distinctions, while the intrinsic structure of formal distinctions is described within the more general ontology of universals.

The ontologies of universals distinguishes among *properties, relations,* and *categories.*

Categories correspond to specific classes used to organize the ontology of particulars, which have the following properties: i) they are material; ii) they are associated to specific identity criteria. Notice that not all monadic universals are categories: for instance, the property *red* (which does not introduce a specific identity criterion for its instances) is not a category. Within categories, we distinguish between *taxons,* which have the further property of being rigid and have an ontological status akin to that of natural kinds, and *roles,* which are not rigid but countable. As mentioned in the introduction, only taxons are allowed to appear as structuring categories in the top-level ontology of particulars. Roles are allowed to be used as structuring categories only within more detailed ontologies built for specific purposes, like natural language translation or problem solving. Within taxons, a useful distinction can be made between *basic categories,* corresponding to uncountable, independent taxons, and *types,* corresponding to countable taxons.

This distinctions among unary relations are mainly taken from (Guarino *et al.* 1994). Besides some technical details, an important addition regards the distinction – within types – between *kinds* and *qualities.* corresponding to what have been called "determinables". In the vision endorsed here, these entities are just metaclasses (*quality spaces*) of properties: Red is therefore an instance of Color. Qualities can be characterized by being rigid, countable, and *dependent* (in order x to be a color, something else which has the property of being x must exist).

Properties (maybe we should call them *mere* properties) correspond to either formal unary universals or material unary universals which do not introduce a specific identity criterion. According to their domain of definition, they can be defined as *first-order* or *higher-order*.

Relations correspond to n-ary universals, organized around the two distinctions *first-order/higher-order* and *formal/material.*

Strata

Let us focus now on the material distinctions among taxons (Figure 1). The main criteria we adopt for this purpose are based on the theory of identity and the theory of dependence. *Substrates* have an extensional criterion of identity, in the sense that any two of them are identical if they have the same parts; objects, on the other hand, have an intensional criterion of identity, in the sense that they are more that a mere sum of parts. Moreover, *objects* depend on substrates, while *events* depend on objects (and substrates, too). Objects and events are "stratified" according to different "points of view", each with a specific kind of identity criterion. For instance, what we have called physical bodies are based on a simple mereo-topo-morphological identity criterion, while functional objects are characterized by a functional identity criterion. For a formal account of the relationships between physical objects and their substrates (not involving time), see (Borgo, Guarino, & Masolo 1996).

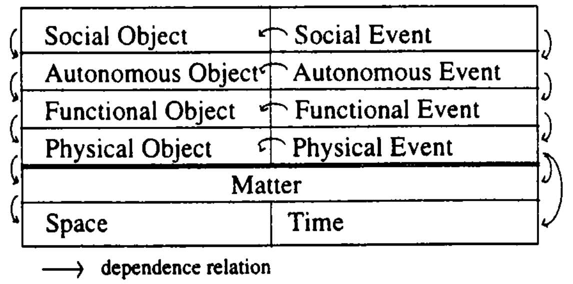

Figure 1. Strata and dependence relations

Further distinctions are made within the basic categories of objects and events according to the ways used to ascribe them identity criteria. For instance, an animal can be conceptualized as an autonomous object, as a biological organism or just as a piece of matter. We argue that different identity criteria correspond to disjoint sub-concepts within a basic category: in the case of an animal, three distinct individuals coexist in the same space and time, corresponding to the three conceptualizations above. Since the animal depends on the underlying biological organism, as well as the biological organism depends on the underlying amount of matter, we call these different sub-concepts *ontological strata.* See (Poli 1996) for a further account of this issue.

Abrupt conclusions

Due to the limits of time (and the vastity of the topic), I am forced to omit the discussion concerning the remaining distinctions adopted in the top-level reported in Appendix 2. I hope to be able to further discuss this issues in a future extended version of this paper.

Acknowledgements

This work has been done in the framework of the special CNR project "Strumenti Ontologico-Linguistici per la Modellazione Concettuale". It has also been partially funded by a contract between LADSEB-CNR and the Apple/IBM Consortium "CORINTO" regarding the development of tools for the classification and reuse of object-oriented software components.

I am grateful to Stefano Borgo, Massimiliano Carrara, Pierdaniele Giaretta, Claudio Masolo, Roberto Poli, Barry Smith, and Achille Varzi for the precious and lively discussions on the topics of this paper.

References

Borgo, S., Guarino, N., & Masolo, C. 1996. Stratified Ontologies: the Case of Physical Objects. In *Proceedings of ECAI-96 Workshop on Ontological Engineering.* Budapest.

Bouaud, J., Bachimont, B., Charlet, J., & Zweigenbaum, P. 1995. Methodological Principles for Structuring an "Ontology". In *Proceedings of IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing.* Montreal, Quebec, Canada.

Burkhardt, H. & Smith, B. (eds.) 1991. *Handbook of Metaphysics and Ontology.* Philosophia Verlag, Münich.

Guarino, N. 1992. Concepts, Attributes and Arbitrary Relations: Some Linguistic and Ontological Criteria for Structuring Knowledge Bases. *Data & Knowledge Engineering,* **8**: 249-261.

Guarino, N. 1995. Formal Ontology, Conceptual Analysis and Knowledge Representation. *International Journal of Human and Computer Studies,* **43**(5/6): 625-640.

Guarino, N., Carrara, M., & Giaretta, P. 1994. An Ontology of Meta-Level Categories. In D. J., E. Sandewall and P. Torasso (eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR94).* Morgan Kaufmann, San Mateo, CA: 270-280.

Guarino, N. & Giaretta, P. 1995. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N. Mars (ed.) *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995.* IOS Press, Amsterdam: 25-32.

Hirst, G. 1991. Existence Assumptions in Knowledge Representation. *Artificial Intelligence,* **49**: 199-242.

Poli, R. 1996. Ontology and Knowledge Organization. In *Proceedings of 4th Conference of the International Society of Knowledge Organization (ISKO 96).* Washington.

Simons, P. 1987. *Parts: a Study in Ontology.* Clarendon Press, Oxford.

Simons, P. 1995. New Categories for Formal Ontology. *Grazer Philosophische Studien,* **49**: 77-100.

Smith, B. 1995. Formal Ontology, Commonsense and Cognitive Science. In N. Guarino and R. Poli (eds.), *Formal Ontology in Conceptual Analysis and Knowledge Representation.* .

Sowa, J. 1995. Top-level ontological categories. *International Journal of Human and Computer Studies,* **43**: 669-685.

Varzi, A. 1996. Parts, Wholes, and Part-Whole Relations: The Prospects of Mereotopology. *Data and Knowledge Engineering*: (in press).

Woods, W. 1986. Important issues in knowledge representation. *Proceedings of the IEEE,* **74**(10): 1322-1334.

Appendix 1 - Mini-glossary

(clarifies the ad-hoc use of potentially ambiguous words used in the text and in the ontology)

Category: A unary universal used to organize an ontology of particulars. Can be either a *taxon* or a *role*. It is sometimes called *ontological category*, but the adjective 'ontological' seems superfluous in this context.

Concept: synonymous of *category*

Ontological property/relation: abbreviation of *formal* ontological property/relation.

Property: synonymous of *determinate*. Not synonymous of unary universal. Unary universals are partitioned into *properties* and *categories*.

Quality: synonymous of *determinable*. Not synonymous of property.

Relation: without arity qualification, synonymous of binary universal.

Role: it is a *unary* universal, not a binary universal as in KL-ONE based languages. See (Guarino 1992). Roles are all *categories*, that means that they may be used to organize an ontology, but their use is not advised in a *top-level* ontology.

Taxon (pl. *taxons*): a category used to organize the top-level ontology of particulars.

Appendix 2 - A preliminary top-level proposal

NOTE - This ontology is in no sense stable and complete. It is intended to be a preliminary contribution showing the possible advantages of philosophical distinctions on the practice of ontological engineering. Many of these distinctions still present serious philosophical problems, but I believe they can be however very useful for practical applications. Hopefully, their introduction in engineering systems can stimulate further fundamental research.

Formal axioms are intended to characterize the meaning of the concepts used, but they are not reported here due to the preliminary nature of this work.

- **Particular**
  - Substrate[1]
    - Space
    - Time
    - Matter
    - Knowledge[2]
  - Object
    - [Singular] object
      - Physical object
        - Physical body[3]
          - *this block of wood*
        - Physical region[4]
          - Spatial part
            - *the upper half of a block*
          - Physical feature
            - *a hole*
            - *a scratch in a block*
          - Boundary
            - *the boundary of a block*
      - Functional object
        - Functional body[5]
          - Artifact
            - *this bolt*
            - Symbolic object
          - Biological body
            - Animal body
              - *the body of this man*
            - Vegetal body
              - *this tree*
        - Functional region
          - Functional part
            - *the head of this bolt*
            - *the condyle of this femure*
            - *the hand of this man.*
          - Functional feature
      - Autonomous object[6]
        - Animal
          - Person
          - Non-human animal
        - Robot
      - Social object[7]
        - Social body
          - Organization
          - Geopolitical entity
          - Ecological system
        - Social region
          - Social part
            - *the people of the highlands*
            - *Southern Italy*
          - Social feature
            - *a sack of poverty*
      - Mathematical object[8]
        - Set
        - Number
        - ...
    - Plural object
  - Event
    - [Singular] Event

---

[1] Identity criterion is extensional. Dimensionality can be seen as a formal property of substrates.

[2] The status of this entry is not clear.

[3] Identity criteria based on mereo-topological (and also morphological?) properties. Maximally self-connected and independent. Pieces of matter, which are not necessarily self-connected, are not considered here.

[4] Dependent on physical objects.

[5] Identity criteria based on functional and pragmatic properties.

[6] Identity criteria based on *agentivity*

[7] Identity criteria based on social rules and conventions.

[8] Notice that we are considering here abstract objects which are not universals.

- Physical event[1]
    - Physical state
        - World state[2]
        - ...
    - Physical action[3]
        - Physical process
        - Physical performance
            - Physical accomplishment
            - Physical achievement
- Functional event
    - Functional state
    - Functional process
    - Functional performance
- Autonomous event
    - Mental event
    - ...
- Social event[4]
    - Communicative event
    - ...
- Plural event

Universal

- Property[5]
    - First order property
        - Formal —
            - mereology-related[6]
                - *atomicity/decomposability*
            - integrity-related
                - *(self) connectedness*
                - *singularity/plurality*[7]
            - identity-related
                - *extensionality*
            - dependence-related
                - *rigid dependence (on something)*
                - *generic dependence (on something)*
                - *independent*
            - *existence*[8]
            - *wholeness*[9]
            - First-order mathematical property
            - ...
        - Material —
            - Physical property
            - Functional property
                - Sign[10]
                    - Symbol
                        - Character
                        - String
                            - Word
                            - *Name*
                            - Text
                    - Icon
                - ...
            - Autonomous property
            - Social property
    - Higher-order property
        - Formal —
            - existence-related
                - *discriminating*
            - mereology-related —
            - integrity-related —
                - *countable*
            - identity-related —
                - *rigid*
            - dependence-related —
                - *foundation*
            - constituency-related —
                - *substance-constitution*
            - Higher-order mathematical property
                - *transitivity*
                - *simmetry*
                - *reflexivity*
                - *equivalence*
                - ...
            - ...
        - Material —
- Relation
    - First-order relation
        - Formal —[11]
            - mereology-related —
                - Part-whole relation
                    - *Part-Whole (generic)*
                    - *EssentialPart-Whole*
                - *Element-Collection*
            - integrity-related —
                - *abstract connection*
            - identity-related —
            - dependence-related —
                - *rigid dependence*
                - *generic dependence*
            - constituency-related
                - *constitution*[12]

[1] The underlying ontology of events is rather naive, and only indicative of important problems and distinctions.
[2] A maximal state of affairs of the whole set of individuals
[3] According to some authors, actions should be separated from events. I am myself sympathetic with this view, while however sticking to Mourelatos' classical distinction just for the sake of simplicity.
[4] Identity criterion is based on social conventions and agreements established among autonomous objects.
[5] Properties may be denoted either by adjectives by the corresponding nouns; i.e., no semantic difference is assumed between *transitive* and *transitivity*. Moreover, properties may be primitive or non-primitive.
[6] Other properties may be related to the *internal structure* of an object.
[7] A plurality is considered as a mereological sum. There are no pluralities of pluralities [Simons, Link, Franconi]
[8] *Existence* is not intended here as related to logical quantification, but rather to ontological status, like existence in a particular world; such a status may correspond to an ad-hoc logical predicate as discussed in (Hirst 1991).
[9] Independent and maximally self-connected
[10] In the sense of *conventional sign*, whose purpose (of representing something) has already been determined.
[11] For the sake of simplicity, higher arity relations are not considered.
[12] "An autonomous object is *constituted* by a physical object"

- formal comparison relation
  - *identity*
  - congruence relation
- Material —
  - Object relation
    - Physical relation[1]
      - Spatial extension
      - Spatial ordering
      - Spatial connection
      - Physical connection
      - Containment
      - *PhysicalPart-Whole*
    - Functional relation
      - *FunctionalPart-Whole (Component-Whole)*
    - Conscious —
      - Psychological possession
    - Social —
      - Legal possession
  - Event relation
    - Temporal location
      - Temporal extension
      - Temporal ordering
    - Causal relation
    - Participation[2]

Higher-order relation
- Formal —
  - *arity*
  - taxonomic relation
    - *Subclass-Superclass*
    - *Instance-Class*
- Material —
  - Measuring unit[3]

Category
- Taxon
  - Basic category[4]
    - Substrate[5]
    - *object, event...*
  - Type
    - Kind
      - *physical body*
    - Quality[6]
      - Formal quality
        - Mathematical quality
          - *cardinality*
        - Formal ontological quality
      - Material quality[7]
        - Physical quality
          - *color*
          - *size*
        - Functional quality
        - Conscious (?) quality
        - Social quality

Role
- Thematic role
- Structural role
- ...

[1] Physical relations "generate" events
[2] of an objects in an event.
[3] Intended as a relation between a quality and a reference particular (e.g., between *length* and a specific object used as a reference).
[4] Basic categories are all first-order
[5] Non-decomposable
[6] Qualities are all higher-order
[7] Qualities are "determinables", while properties are "determinates". Qualities are classes of properties. For instance, *color* includes *red* and *blue* among its instances.