Sweetening Ontologies with DOLCE

Aldo Gangemi, Nicola Guarino, Claudio Masolo,
Alessandro Oltramari, Luc Schneider

[1] ISTC-CNR, Rome/Padua, Italy
gangemi@ip.rm.cnr.it, {Nicola.Guarino, Alessandro.Oltramari,
Claudio.Masolo, Luc.Schneider}@ladseb.pd.cnr.it

Abstract. In this paper we introduce the *DOLCE* upper level ontology, the first module of a Foundational Ontologies Library being developed within the WonderWeb project. *DOLCE* is presented here in an intuitive way; the reader should refer to the project deliverable for a detailed axiomatization. A comparison with *WordNet's* top-level taxonomy of nouns is also provided, which shows how *DOLCE*, used in addition to the *OntoClean* methodology, helps isolating and understanding some major WordNet's semantic limitations. We suggest that such analysis could hopefully lead to an "ontologically sweetened" WordNet, meant to be conceptually more rigorous, cognitively transparent, and efficiently exploitable in several applications.

1 Introduction

In the recent years, we developed a methodology for testing the ontological adequacy of taxonomic links called OntoClean [14, 13], which was used as a tool for a first systematic analysis of WordNet's upper level taxonomy of nouns [6]. The first version of OntoClean was based on an ontology of properties (unary *universals*), characterized by means of meta-properties. We are now complementing OntoClean with an ontology of *particulars* called DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering), which is presented here in some detail.

DOLCE is the first module of a *Library of Foundational Ontologies* being developed within the WonderWeb project[1]. In contrast with "lightweight" ontologies, which focus on a minimal terminological structure (often just a taxonomy) fitting the needs of a specific community, the main purpose of foundational ontologies is to *negotiate meaning*, either for enabling effective cooperation among multiple artificial agents, or for *establishing consensus* in a mixed society where artificial agents cooperate with human beings. The WonderWeb vision is to have *a library* of such ontologies, reflecting different ontological choices. The idea is to make the rationales and alternatives underlying such choices as explicit as possible, as a result of a careful isolation of the fundamental ontological options and their formal relationships. The library would form a network of different but systematically related modules which the various Semantic Web applications can commit to, according to their ontological assumptions.

[1] http://wonderweb.semanticweb.org/deliverables/D17.shtml

This paper is structured as follows. In the next section we introduce the basic assumptions and distinctions underlying DOLCE; then we discuss some ontological inadequacies of WordNet's taxonomy of nouns, revising and extending the analysis presented in [6]. Finally, we discuss the preliminary results of an alignment work aimed at improving WordNet's overall ontological (and cognitive) adequacy, and facilitate its effective deployment in practical applications.

2 The DOLCE Upper Ontology

According to the vision introduced above, we do *not* intend DOLCE as a candidate for a "universal" standard ontology. Rather, it is intended to act as starting point for comparing and elucidating the relationships with other future modules of the library, and also for clarifying the hidden assumptions underlying existing ontologies or linguistic resources such as WordNet.

As reflected by its acronym, DOLCE has a clear *cognitive bias*, in the sense that it aims at capturing the ontological categories underlying natural language and human commonsense. We believe that such bias is very important for the Semantic Web (especially if we recognize its intrinsic social nature [3]). We do not commit to a strictly referentialist metaphysics related to the intrinsic nature of the world: rather, the categories we introduce here are thought of as cognitive artifacts ultimately depending on human perception, cultural imprints and social conventions (a sort of "cognitive" metaphysics). We draw inspiration here from Searle's notion of "deep background" [18], which represents the set of skills, tendencies and habits shared by humans because of their peculiar biological make up, and their evolved ability to interact with their ecological niches [9]. The consequences of this approach are that our categories are at the so-called *mesoscopic* level, and they do not claim any special robustness against the state of the art in scientific knowledge: they are just *descriptive* notions [21] that assist in making *already formed* conceptualizations explicit. They do not provide therefore a *prescriptive* (or "revisionary" [21, 15]) framework to conceptualize entities. In other words, our categories describe entities in a post-hoc way, reflecting more ore less the surface structure of language and cognition.

DOLCE is an ontology of *particulars*, in the sense that its domain of discourse is restricted to them. The fundamental ontological distinction between *universals* and *particulars* can be informally understood by taking the relation of *instantiation* as a primitive: particulars are entities which have no instances[2]; universals are entities that do have instances. Properties and relations (corresponding to predicates in a logical language) are usually considered as universals. We take the ontology of universals as formally separated from that of particulars. Of course, universals *do* appear in an ontology of particulars, insofar they are used to organize and characterize them: simply, since they are not in the domain of discourse, they are not themselves subject to being organized and characterized (e.g., by means of *metaproperties*). An ontology of unary universals has been presented in [12]. In this paper, we shall occasionally use notions (e.g., rigidity) taken from such work in our meta-language.

---

[2] More exactly, we should say that they *can't* have instances. This coincides with saying that they have no instances, if we include *possibilia* (possible instances) among instances.

2.1 Enduring and perduring entities

DOLCE is based on a fundamental distinction between *enduring* and *perduring* entities, i.e. between what philosophers usually call *continuants* and *occurrents* [19], a distinction still strongly debated both in the philosophical literature [22] and within ontology standardization initiatives[3]. Again, we must emphasise that this distinction is motivated by our cognitive bias: we do not commit to the fact that both these kinds of entity "really exist", and we are indeed sympathetic with the recent proposal made by Peter Simons, that enduring entities can be seen as equivalence classes of perduring entities, as the result of some kind of abstraction mechanism [20].

The difference between enduring and perduring entities (which we shall also call *endurants* and *perdurants*) is related to their behavior in time. Endurants are *wholly* present (i.e., all their proper parts are present) at any time they are present. Perdurants, on the other hand, just extend in time by accumulating different temporal parts, so that, at any time they are present, they are only *partially* present, in the sense that some of their proper temporal parts (e.g., their previous or future phases) may be not present. E.g., the piece of paper you are reading now is wholly present, while some temporal parts of your reading are not present any more. Philosophers say that endurants are entities that *are in time*, while lacking however temporal parts (so to speak, all their parts flow with them in time). Perdurants, on the other hand, are entities that *happen in time*, and can have temporal parts (all their parts are fixed in time)[4].

Hence endurants and perdurants can be characterised by whether or not they can exhibit change in time. Endurants can "genuinely" change in time, in the sense that the very same endurant as a whole can have incompatible properties at different times; perdurants cannot change in this sense, since none of their parts keeps its identity in time. To see this, suppose that an endurant has a property at a time t, and a different, incompatible property at time t': in both cases we refer to the whole object, without picking up any particular part. On the other hand, when we say that a perdurant has a property at t, and an incompatible property at t', there are always two different parts exhibiting the two properties.

The main relation between endurants and perdurants is that of *participation*: an endurant "lives" in time by *participating* in a perdurant. For example, a person, which is an endurant, may participate in a discussion, which is a perdurant. A person's life is also a perdurant, in which a person participates throughout its all duration.

In the following, we shall take the term *occurrence* as synonym of *perdurant*. We prefer this choice to the more common *occurrent*, which we reserve for denoting a type (a *universal*), whose instances are occurrences (*particulars*).

---

[3] See for instance the extensive debate about the "3D" vs. the "4D" approach at suo.ieee.org, or the SNAP/SPAN opposition sketched at ontology.buffalo.edu/bfo.

[4] Time-snapshots of perdurants (i.e., in our time structure, perdurants whose temporal location is atomic, and which lack therefore proper temporal parts) are a limit case in this distinction. We consider them as perdurants since we assume that their temporal location is fixed (a time-snapshot at a different time would be a different time-snapshot).

2.3 DOLCE's Top Categories

The taxonomy of the most basic categories of particulars assumed in DOLCE is depicted in Figure 1. They are considered as rigid properties, according to the OntoClean methodology that stresses the importance of focusing on these properties first. Some examples of "leaf" categories instances are illustrated in Table 1.

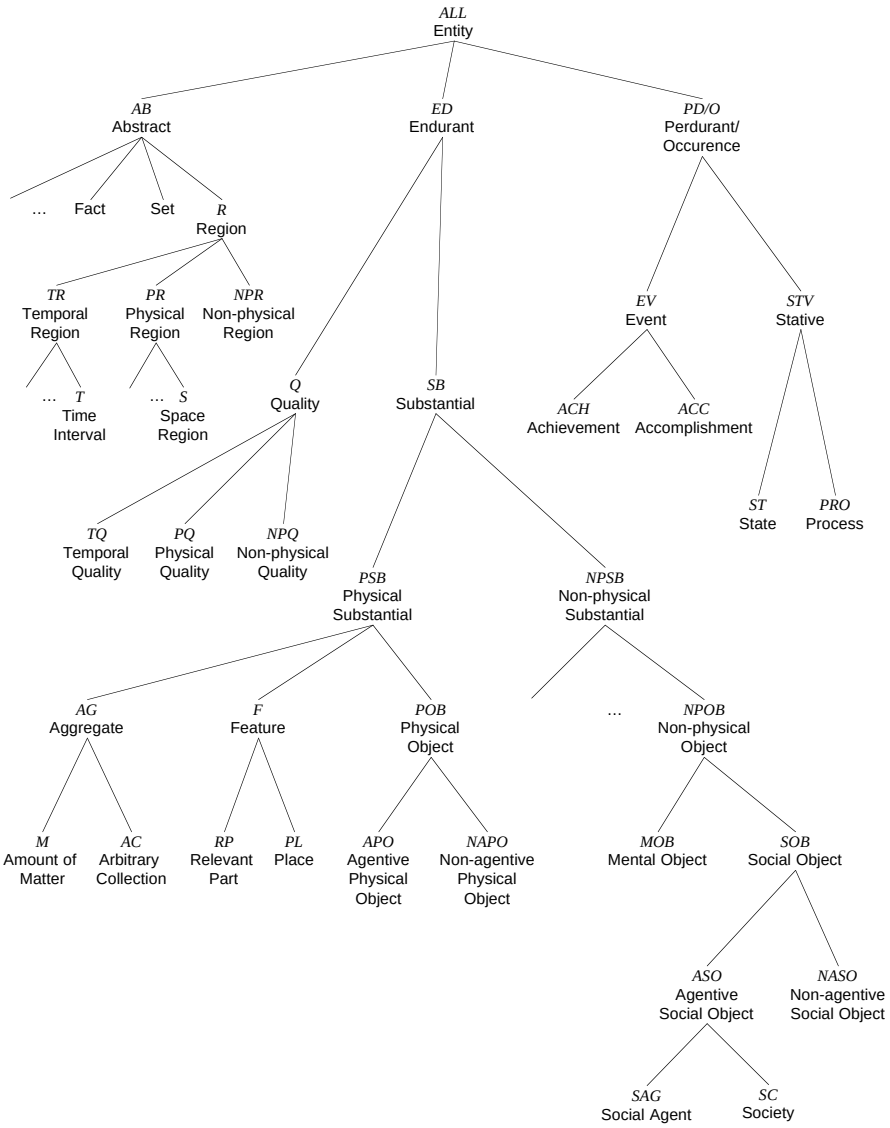

Fig. 1. Taxonomy of DOLCE basic categories.

Qualities and quality regions. Qualities can be seen as the basic entities we can perceive or measure: shapes, colors, sizes, sounds, smells, as well as masses, lengths, electrical charges… The term 'Quality' is often used as a synonymous of 'property', but this is not the case in DOLCE: qualities are particulars, properties are universals. Qualities *inhere* to entities: every entity (including qualities themselves) comes with certain qualities, which exist exactly as long as the entity exists. Within a certain ontology, we assume that these qualities belong to a finite set of *quality types* (like color, size, smell, etc.), and are characteristic for (*inhere in*) specific individuals: no two particulars can have the same quality, and each quality is *specifically constantly dependent* on the entity it inheres in: at any time, a quality can't be present unless the entity it inheres in is also present. So we distinguish between a quality (e.g., the color of a specific rose), and its "value" (e.g., a particular shade of red). The latter is called *quale*, and describes the position of an individual quality within a certain *conceptual space* (called here *quality space*) [8]. So when we say that two roses have (exactly) the same color their *two* colors have the same position in the color space (they have the same *color quale*), but still the two roses have numerically distinct color qualities.

Table 1. Examples of "leaf" basic categories.

"Leaf" Basic Category	**Examples**
Accomplishment	*a conference, an ascent, a performance*
Achievement	*reaching the summit of K2, a departure, a death*
Agentive Physical Object	*a natural person*
Amount of Matter	*some air, some gold, some cement*
Arbitrary Collection	*my left foot and my car*
Mental Object	*an idea*
Non-agentive Physical Obj.	*a hammer, a house, a computer, a human body*
Non-agentive Social Object	*a law, an economic system, a currency, an asset*
Non-physical Quality	*the value of a stock share*
Non-physical Region	*a 1Euro value*
Physical Quality	*the weight of a pen, the color of an apple*
Physical Region	*the Euclidean space, an area in the color spectrum*
Place	*a hole, a gulf, an opening*
Process	*running, writing*
Relevant Part	*a bump, an edge, a skin*
Social Agent	*a legal person, a contractant*
Society	*Fiat, Apple, the Bank of Italy*
State	*being sitting, being open, being happy, being red*
Temporal Quality	*the duration of a battle, the starting time of a race*
Temporal Region	*the time axis, 22 june 2002, one second*

This distinction between qualities and qualia is inspired by [10] and the so-called *trope theory* [1] (with some differences that can't be discussed here[5]). Its intuitive rationale is mainly due to the fact that natural language – in certain constructs – often seems to make a similar distinction. For instance, when we say "the color of the rose turned from red to brown in one week" or "the room's temperature is increasing" we are not speaking of a certain shade of red, or a specific thermodynamic status, but of something else that changes its properties in time while keeping its identity. This is why we assume that *qualities are endurants*.

On the other hand, when we say that "red is opposite to green" or "red is close to brown" we are not speaking of qualities, but rather of regions within quality spaces. The specific shade of red of our rose – its color quale – is therefore a point (or an atom, mereologically speaking) in the color space.

Each quality type has an associated quality space with a specific structure. For example, lengths are usually associated to a metric linear space, and colors to a topological 2D space. The structure of these spaces reflects our perceptual and cognitive bias.

In this approach, we can explain the relation existing between 'red' intended as an adjective (as in "this rose is red") and 'red' intended as a noun (as in "red is a color") (Figure 2): the rose is red because its color is located in the red region within the color space (more exactly, its color quale is a part of that region). Moreover, we can explain the difference between "this rose is red" and "the color of this rose is red" by interpreting "red" as synonymous of *red object* in the first case, and of *red color* in the latter case.

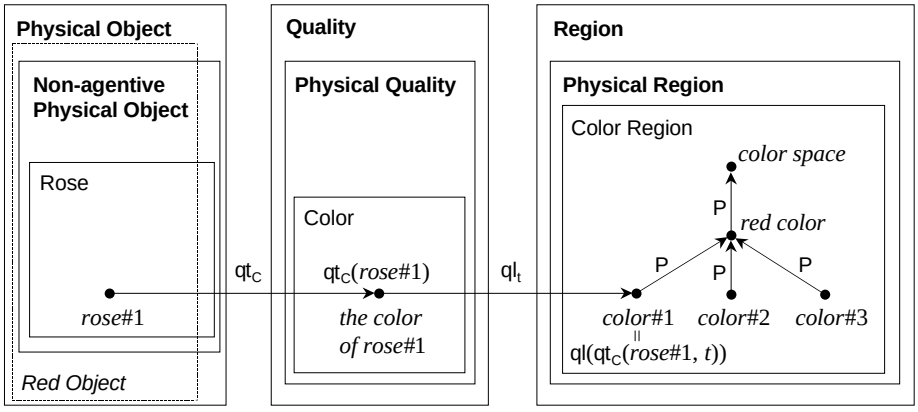

Fig. 2. Qualities and quality regions.

[5] An important difference is that standard trope theories explain a qualitative change in terms of a substitution of tropes (an old trope disappears and a new one is created). We assume instead that qualities are a sort of "enduring tropes".

In our ontology, space and time are considered as quality types like color, weight, etc. The spatial (temporal) individual quality of an entity is called *spatial* (*temporal*) *location*, while its quale is called *spatial* (*temporal*) *region*. For example, the spatial location of a physical object is just one of its individual qualities: it belongs to the quality type *space*, and its quale is a region in the geometric space. Similarly for the temporal location of an occurrence, whose quale is a region in the temporal space. This allows an homogeneous approach that remains neutral about the properties of the geometric/temporal space adopted (for instance, one may assume a circular time).

Concerning the inherence relation, we distinguish between *direct* and *indirect* quality inherence. So *temporal qualities* are those that directly inhere to occurrences, and *physical qualities* are those that directly inhere physical entities (physical entities, in turn, are those having a direct spatial location). Then, for example, occurrences have physical qualities only indirectly, insofar these qualities directly inhere to their participants.

Substantials. Roughly, we see substantials as stable aggregates of qualities: they are endurants that can have qualities, but are not themselves qualities. Most of such aggregations are cognitive artifacts, resulting from the tendency humans have to partition their environment around "islands of stability" that have enough permanence and features to be used as pervading frameworks of reference [21]. The term "substantial" is inspired to the Aristotelian notion of *substance*, but is indeed more general than the latter, which is closer to our notion of *object* (see below).

Substantials form the main branch of our taxonomy. We distinguish between *physical* and *non-physical substantials*, according to whether they have direct spatial qualities. At the moment, the ontology of non-physical substantials is still in progress. Within physical substantials, we distinguish between *aggregates*, *objects*, and *features*. This distinction is mainly based on the notion of unity we have discussed and formalized in [5]. In principle, the general structure of such distinction is supposed to also hold for non-physical substantials: nevertheless, we fully exploit it only on for physical substantials, since the characteristics of non-physical aggregates and features have not been considered yet.

Aggregates. The common trait of aggregates is that they are endurants with no unity (according to **[5]**, none of them is an essential whole). We consider two kinds of aggregates: amounts of matter and arbitrary collections. The former are *mereologically invariant*, in the sense that they change their identity when they change some parts. The latter are defined as mere mereological sums of essential wholes (e.g. objects, see below) which are not themselves essential wholes (like the sum of a person's nose and a computer keyboard). They are essentially mereologically pseudo-invariant, in the sense that they change their identity when a member[6] is changed, while a change in the non essential parts of a member is allowed. We may have called these arbitrary collections 'groups', or perhaps 'sets'; but we prefer to use 'set' for abstract entities, and 'group' for something having an intrinsic unity.

Objects. The main characteristic of objects is that they are endurants with unity. They have no *common* unity criterion, however, as different subtypes of objects may

[6] We assume here that a member is a special part of a collection, see [5].

have different unity criteria. Differently from aggregates, (most) objects are admitted to change some of their parts while keeping their identity: they can have therefore *temporary parts*. Often objects (indeed, all endurants) are considered as ontologically independent from occurrences (discussed below). However, if we admit that every object has a life, it is hard to exclude a mutual specific constant dependence between the two. Nevertheless, we may still use the notion of dependence to (weakly) characterize objects as being not specifically constantly dependent *on other objects*.

Features. Typical examples of features are "parasitic entities" such as holes, bumps, surfaces, or stains, which are (in most cases) specifically constantly dependent on physical objects[7] (their hosts). All features are essential wholes, but no common unity criterion may exist for all of them. However, typical features have a topological unity, as they are singular entities. Features may be *relevant parts* of their host, like a bump or an edge, or *places* like a hole in a piece of cheese, the underneath of a table, the front of a house, which are not parts of their host. We include within features also boundaries, which may be conceptualized in various ways, and are not discussed here.

It may be interesting to note that we do not consider body parts like heads or hands as features: the reason is that we assume that a hand can be detached from its host (differently from a hole or a bump), and we assume that in this case it retains its identity. Should we reject this assumption, then body parts would be features (relevant parts).

Non-physical substantials and the agentive/non-agentive distinction. Physical objects that have *intentionality* (the capability of heading for/dealing with objects or states of the world, see [18]) are called *Agentive*, those which do not are called *Non-agentive*. In general, the former are *constituted* by the latter: human persons are constituted by organisms, robots are constituted by machinaries, and so on (constitution is taken here as a primitive relation, which is axiomatized in [17]). Among non-agentive physical objects we have *ordinary objects* like houses, organs, pieces of wood, etc. *Non-physical Objects* are divided into *Mental* and *Social* according to whether they are "produced" by a single agent or recognized by a community of agents. In the first case we say that *mental objects* (like an idea) are specifically dependent on *agentive physical objects*, while in the second case we need to further distinguish between *Agentive* and *Non-agentive social objects*. Examples of the former category are *social agents* like the president of United States or a top manager of Microsoft, conceived as "reified roles" depending on agentive physical objects (certain persons) only in a *generic* way, as the role may survive a replacement of the person. Social agents are *not* constituted by agentive physical objects (although they depend on them), while they can constitute *societies*, like the CNR, Microsoft, etc. *Non-Agentive Social Objects* like laws, shares, peace treaties ecc. are generically dependent on societies, which are therefore the "conditio sine qua non" of their ontological status.

---

[7] In some cases, features are just *generically* dependent on their host, in the sense that *some* (suitable) object must exist whenever the feature exists: think for instance of a whirlpool: if it is a feature, what is its host?

Occurrences. Occurrences comprise what are variously called events, processes, phenomena, activities and states. They can have temporal parts or spatial parts. For instance, the first movement of (the execution of) a symphony is a temporal part of it. On the other side, the play performed by the left side of the orchestra is a spatial part. In both cases, these parts are occurrences themselves. We assume that objects can't be parts of occurrences, but rather they *participate* in them.

An ontology of occurrences has to take into account two basic aspects: *change* and *homeomericity*.

The first one concerns a naive view of our everyday experience of the world: for instance, if we see a ship standing still on the sea for an hour, we'll say that «the ship hasn't changed its position for an hour»; on the other side, if during the same interval we see the ship navigating from the harbor to an oil platform, we'll say «the ship has been moving for an hour». In the latter example, the detection of a movement implies that we are talking about a *dynamic occurrence*, while in the former we are speaking about a *stationary* occurrence.

The second aspect has been extensively discussed in [2]: intuitively, we can say that an occurrence is homeomeric if and only if all its temporal parts can be described *in the same way* used for the whole occurrence. Every temporal part of "John sitting here" for an hour is still a "sitting here of John". But if we consider "the complete ascent of Everest by Messner", there are no parts of such event which constitute a complete ascent of Everest by Messner. In linguistic as well as in philosophical terminology, the notion of the "*homeomericity*" of an occurrence is often introduced with respect to a property characteristic of (or *exemplified by*) the occurrent itself. If such property holds for all the temporal parts of the occurrence, then the occurrence is homeomeric. In our axiomatization, this presupposes a finite list of occurrence-types which have to be "declared" in advance. An occurrence-type is *stative* or *eventive* according to whether or not it holds of the mereological sum of two of its instances. For instance, a *sitting* occurrence is stative since the sum of two sittings is still a sitting occurrence. Within stative occurrences, we distinguish between *states* and *processes* according to whether the corresponding types hold of every part of their instances: so *sitting* is a state, while *running* is a process, since there may be (very short) temporal parts of a running that are not themselves runnings.

Finally, eventive occurrences (*events*) are called *achievements* if they are atomic, otherwise they are *accomplishments*.

3 Ontological problems in WordNet

Let us see now how the ontology we introduced, together with the general principles of the OntoClean methodology, can be of help in analyzing the ontological structure of WordNet[8]. We believe that such analysis is important, as the number of applications where WordNet is being used more as an ontology than just as a lexical resource seems to be growing more and more. To be used as an ontology, however, some of WordNet's lexical links need to be re-interpreted as semantic links, connecting together intended meaning of words, according to our own conceptualizations. One of such links is the hyponym/hypernym relation, which corresponds in many cases to the usual subsumption (or IS_A) relation between concepts. An early attempt at exploring the semantic and ontological problems lying behind this correspondence is described in [11]. Let us extend now such discussion in the light of the DOLCE ontology.

Confusion between concepts and individuals. The first critical problem we found in WordNet was the confusion between concepts and individuals. For instance, if we look at the hyponyms of the "unique beginner" Event, we'll find the synset Fall - an individual - whose gloss is "the lapse of mankind into sinfulness because of the sin of Adam and Eve", together with conceptual hyponyms such as Social_Event, and Miracle.[9] Under Territorial_Dominion we find Macao and Palestine together with Trust_Territory. The latter synset, defined as "a dependent country, administered by a country under the supervision of United Nations", denotes a general kind of country, rather than a specific country as those preceding it. If we go deeper in the taxonomy, we find many other examples of this sort. For instance, the hyponyms of Composer are a mixture of concepts and instances: there are classes corresponding to different special fields, such as Contrapuntist, or Songwriter, and examples of famous musicians of the past, such as Bach, and Beethoven.

Under Martial_Art, whose top hypernym is Act, we find Karate, and Kung Fu, but these synsets do not stand for concepts, they represent individuals, namely particular examples of martial arts.

If we look through Organization, under the branch whose root is Group, we find conceptual hyponyms such as Company, Alliance, Federation, Committee, together with instances like Irish_Republican_Army, Red Cross, and so on.

We face here a general problem: the concept/individual confusion is nothing but the product of an "expressivity lack". In fact, if there was an INSTANCE-OF relation, we could distinguish between a concept-to-concept relation (subsumption) and an individual-to-concept one (instantiation).

---

[8] We refer here to WordNet 1.6 (see [6] for a partial overview on the top-level structure)

[9] In the text body, we usually do not report all the synonyms of a synset (or their numeration), but only the most meaningful ones.

Confusion between object-level and meta-level: the case of Abstraction. The synset Abstraction_1 seems to include both object-level concepts, such as Set, Time, and Space, and meta-level concepts such as Attribute and Relation. From the corresponding gloss, an abstraction "is a general concept formed by extracting common features from specific examples". An abstraction seems therefore intended as a psychological process of generalization, in accordance to Locke's position ([16], p.211). This meaning seems to fit the latter group of terms (Attribute, Relation, and possibly some hyponyms of Quantity), but not the former. Moreover, it is quite natural to consider attributes and relations as meta-level concepts, while set, time, and space, seem to belong to the object domain.

OntoClean constraints violations. A core aspect of OntoClean is the analysis of subsumption constraints induced by the identity, rigidity, and unity meta-properties. In our analysis, we only found rigidity violations. We suspect that there are two reasons why we didn't observe other kinds of violation: on one hand, we limited our analysis to the consistency of lower levels against the upper level, where the criteria of identity and unity are very general; on the other hand, WordNet tends, notoriously, to multiply senses, so the chances of conflict are relatively limited.

The most common violation we have registered is bound to the distinction between roles and types. A role cannot subsume a type. Let's see an important clarifying example.

In its first sense, Person (which we consider as a type) is subsumed by two different concepts, Organism and Causal_Agent. Organism can be conceived as a type, while Causal_Agent as a formal role. The first subsumption relationship is correct, while the second one shows a rigidity violation. We propose therefore to drop it.

Someone could argue that every person is necessarily a causal agent, since 'agentivity' (capability of performing actions) is an essential property of persons. Causal_Agent should therefore be intended as a synonym of 'intentional agent', and considered as rigid. But, in this case, it would have only hyponyms denoting things that are (essentially) causal agents, including animals, spiritual beings, the personified Fate, and so on.

Unfortunately, this is not what happens in WordNet: Agent, one of Causal_Agent hyponyms, is defined as: "an active and efficient cause; capable of producing a certain effect; (the research uncovered new disease agents)". Causal_Agent subsumes roles such as Germicide, Vasoconstrictor, Antifungal. Instances of these concepts are not causal agents essentially. This means that considering Causal_Agent as rigid would introduce further inconsistencies.

These considerations allow us to add a pragmatic guideline to our methodology: when deciding about the formal meta-property to attach to a certain concept, it is useful to look at all its children.

Missing polysemy detection. WordNet is said to recognize most of the conventional senses of a word (obviously not all the possible contextual senses[10]). Nonetheless, there are cases where relevant polysemy has not been detected. Such a

[10] By the way, contextual polysemy does not usually affect the category of the sense of a word, but its so-called *connotation*.

case emerges in two modalities: the first is multiple hyperonymy, the second is sense gap. Here we show an example of the first.

Multiple hyperonymy is not widespread in WordNet nouns (about 900 synsets) and it is often used appropriately, as in Surgical_Knife, which has two hyperonyms: Surgical_Instrument and Knife_1. In this case, Surgical_Instrument is a role, then there is no conflict in specializing from Knife_1 (which is a type) to Surgical_Knife (which is a role). But there are cases of multiple and incompatible identity criteria, as in Law, which has the two hyperonyms Legal_Document and Rule. According to DOLCE, we consider Legal_Document as subsumed by Non-Agentive Physical Object, and Rule as subsumed by Non-Agentive Social Object. So the two categories are disjoint. Consequently, this multiple hyperonymy generates a logical incoherence, which could not be detected without an explicitly axiomatized upper-level.

Moreover, this is a case of *systematic polysemy*, since a legal document is the physical support for a law. The relation axioms in DOLCE help detecting systematic polysemy, which is a major source for building domain *core* ontologies [7].

Heterogeneous levels of generality. Going down the lower layers of WordNet's top level, we register a certain 'heterogeneity' in their intuitive level of generality. For example, among the hyponyms of Entity there are types such as Physical_Object, and roles such as Subject. The latter is defined as "something (a person or object or scene) selected by an artist or photographer for graphic representation", and has no hyponyms (indeed, almost any entity can be an instance of Subject, but none is necessarily a subject)[11].

For Animal (subsumed by Life_Form) this heterogeneity becomes clearer. Together with classes such as Chordate, Larva, Fictional_Animal, etc., we find out more specific concepts, such as Work_Animal, Domestic_Animal, Mate_3, Captive, Prey, etc. We are induced to consider the formers as types, while the latters as roles.

Although problematic on the side of ontological distinctions among event-classes, the hyponyms of Phenomenon_1 represent another meaningful example of heterogeneity. At the same taxonomic level there are "reasonably" general synsets like Natural_Phenomenon and Process together with a specific concept like Consequence, which could be modeled as anti-rigid (every event can be a consequence of the occurring of a previous event, but we could assume that this is not the essential characteristic of the event itself).

In short, intuitively some synsets sound too specific when compared to their siblings. Look at them from the formal point of view we are developing, we can pinpoint their "different generality" by means of the distinction between types and roles.

4 Mapping WordNet into DOLCE

Let us consider now the results of integrating the WordNet top concepts into our upper level. According to the OntoClean methodology, we have concentrated first on the so-called *backbone taxonomy*, which only includes the rigid properties. Formal and material roles have been therefore excluded from this preliminary work.

[11] We can draw similar observations for relation_1 and set_5 with respect to abstraction_1, etc.

Table 2. Mapping WordNet into DOLCE (some examples).

- **Aggregate**
  - **Amount of matter**
    - body_substance
    - chemical_element
    - mixture
    - compound$chemical_compound
    - mass_5
    - fluid_1
  - **Arbitrary collection**
    - …
- **Physical Object**
  - **Non-agentive**
    - body_of_water$water
    - landdry_landearth$…
    - body$organic_structure
    - artifact$artefact*
    - biological_group
    - kingdom
    - collection
    - **Body**
      - blackbody$full_radiator
      - body_5
      - universe$existence$nature$creation
      - …
  - **Agentive**
    - life_form$organism$being$…
    - citizenry
    - sainthood
    - ethnic group
- **Social Object**
  - **Non-agentive**
    - rule$prescript
      - law
      - …
    - circuit_5
  - **Agentive**
    - social_group
    - …
- **Feature**
  - **Relevant Part**
    - edge_3
    - skin_4
    - paring$parings
    - …
  - **Place**
    - opening_3
    - excavation$hole_in_the_ground
- **Quality**
  - position$place
  - time_interval$interval
  - chromatic_color
  - …
- **Occurrence**
  - **State**
    - condition$status
    - cognitive_state
    - existence
    - death_4
    - degree
    - medium_4
    - relationship_1
    - relationship_2
    - conflict
    - …
  - **Process**
    - decrement_2
    - increment
    - shaping
    - activity_1
    - chelation
    - execution
    - activity_1
    - …
  - **Accomplishment**
    - accomplishment$achievement
    - …
- **Abstract**
  - **Region**
    - space_1
    - time_1
    - time_interval$interval
    - chromatic_color
    - …
  - statement_1
    - proposition
    - …
  - symbol
  - set_5
  - …

Comparing WordNet's unique beginners with our ontological categories, it becomes evident that some notions are very heterogeneous: for example, Entity looks like a "catch-all" class containing concepts hardly classifiable elsewhere, like Anticipation, Imaginary_Place, Inessential, etc. Such synsets have only a few children and these have been already excluded in our analysis.

Some examples of our merging work are sketched in Table 2. Some problems encountered for each category are discussed below.

4.1 Aggregates, Objects, and Features

Entity is a very confused synset. A lot of its hyponyms have to be "rejected": in fact there are roles (Causal_Agent, Subject_4), unclear synsets (Location[12]) and so on. This Unique Beginner maps partly to our Aggregate and partly to our Object category. Some hyponyms of Physical_Object are mapped to our top concept Feature.

By removing roles like Arrangement and Straggle, Group$grouping appears to include Agentive Social Object (social group, ethnic group), Non-agentive Social Object (circuit), Agentive Physical Object (citizenry) and Non-agentive Physical Object (biological group, kingdom; collection).

Possession_1 is a role, and it includes both roles and types. In our opinion, the synsets marked as types (Asset, Liability, etc.) should be moved towards lower levels of the ontology, since their meanings seem to deal more with a specific domain - the economic one - than with a set of general concepts (except some concepts that can be mapped to Mental Object, such as Own_Right). This means that the remainder branch has also to be eliminated from the top level, because of its overall anti-rigidity (the peculiarity of roles).

4.2 Abstracts and Qualities

ABSTRACTION_1 is the most heterogeneous unique beginner: it contains abstracts such as Set_5, quality regions such as Chromatic_Color, qualities (mostly from the synset Attribute) and a hybrid concept (Relation_1) that contains social objects, concrete entities (as Substance_4[13]), and even meta-level categories. Each child synset has been mapped appropriately.

Psychological_feature contains both mental objects (Cognition[14]) and events (Feeling_1). We consider Motivation as a material role, so to be added to lower levels of the taxonomy of mental objects.

The classification of qualities deals mainly with adjectives. This paper focuses on the WordNet database of nouns; nevertheless our treatment of qualities foreshadows a semantic organization of the database of adjectives too, which is a current desideratum in the WordNet community (see [4], p. 66).

---

[12] Referring to Location, we find roles (There, Here, Home, Base, Whereabouts), instances (Earth), and geometric concepts like Line, Point, etc.).

[13] "The stuff of which an object consists".

[14] "The psychological result of perception, and learning and reasoning".

4.3 Occurences

Event_1, Phenomenon_1, State_1 and Act_1 are the Unique Beginners of those branches of WordNet denoting occurrences. In particular, the hyponyms of State_1 seem to fit well with our state category, as the children of Process (a subordinate of Phenomenon). For the time being, we restrict the mapping of our accomplishment category to the homonymous synset of WordNet. Event_1 is too heterogeneous to be clearly partitioned in terms of our approach: to a great extent, however, its hyponyms could be added to lower levels of the taxonomy of occurrences.

5 Conclusions

We are confident that foundational ontologies will eventually improve communication among agents in most cases of information exchange: information retrieval and extraction, semantic web services, software requirement analysis and unified modeling process, control knowledge, etc. In fact, foundational ontologies can act as a *reference* for agents to commit to certain theories, as a set of formal *guidelines* for domain modeling, and as a *tool* for making heterogeneous ontologies interoperate or merge. According to the needs, an upper level ontology can be used either in a light version, for computationally intensive applications, or as an off-the-shelf fully axiomatized theory, to be consulted as a reference source for more sporadic meaning negotiation purposes. That is why we intended DOLCE to be as detailed and rigorous as possible, and yet we plan to release a light-weight version.

In the light of this vision, we have started using DOLCE (or one of its preliminary versions) in several projects, either as a tool or a set of guidelines, with substantial results in the creation of well-founded and useful ontologies (by the way, there is still no benchmark or testbed for ontology quality, since there is small agreement on the criteria to adopt, and we are suggesting some of them …).

The WordNet experiment is one of the research applications of DOLCE, already presented in several contexts, which seems promising in bridging one of the multidisciplinary gaps in ontological engineering, between the domain of lexical technologies, and that of conceptual modelling.

6 Acknowledgements

We would like to thank Stefano Borgo for the fruitful discussions and comments on the earlier version of this paper. This work was jointly supported by the Eureka Project IKF (E!2235, Information and Knowledge Fusion), the IST Project 2001-33052 *WonderWeb* (Ontology Infrastructure for the Semantic Web) and the National project TICCA (Cognitive Technologies for Communication and Cooperation with Artificial Agents).

References

1. Campbell, K.: Abstract Particulars. Basil Blackwell, Oxford. (1990)
2. Casati, R. and Varzi, A. (eds.): Events. Dartmouth, Aldershots, USA (1996)

3. Castelfranchi, C.: Information Agents: The Social Nature of Information and the Role for Trust. In: M. Klusch and F. Zambonelli (eds.), Cooperative Information Agents V, 5th International Workshop, CIA 2001. Springer, Modena, Italy (2001) 208-210
4. Fellbaum, C. (ed.) WordNet - An Electronic Lexical Database. MIT Press (1998)
5. Gangemi, A., Guarino, N., Masolo, C., and Oltramari, A.: Understanding top-level ontological distinctions. In Proceedings of IJCAI-01 Workshop on Ontologies and Information Sharing. Seattle, USA, AAAI Press (2001) 26-33
6. Gangemi, A., Guarino, N., and Oltramari, A.: Conceptual Analysis of Lexical Taxonomies: The Case of WordNet Top-Level. In: C. Welty and S. Barry (eds.), Formal Ontology in Information Systems. Proceedings of FOIS2001. ACM Press (2001) 285-296
7. Gangemi, A., Pisanelli, D. M., and Steve, G.: Understanding Systematic Conceptual Structures in Polysemous Medical Terms. In Proceedings of AMIA Annual Symposium (2000)
8. Gärdenfors, P.: Conceptual Spaces: the Geometry of Thought. MIT Press, Cambridge, Massachussetts (2000)
9. Gibson, J. J.: The Theory of Affordances. In: R. E. Shaw and J. Bransford (eds.), Perceiving, Acting and Knowing. LEA, Hillsdale (1977)
10. Goodman, N.: The Structure of Appearance. Harvard University Press, Cambridge MA. (1951)
11. Guarino, N.: Some Ontological Principles for Designing Upper Level Lexical Resources. In: A. Rubio, N. Gallardo, R. Castro and A. Tejada (eds.), Proceedings of First International Conference on Language Resources and Evaluation. ELRA - European Language Resources Association, Granada, Spain (1998) 527-534
12. Guarino, N. and Welty, C.: A Formal Ontology of Properties. In: R. Dieng and O. Corby (eds.), Knowledge Engineering and Knowledge Management: Methods, Models and Tools. 12th International Conference, EKAW2000. Springer Verlag, France (2000) 97-112
13. Guarino, N. and Welty, C.: Evaluating Ontological Decisions with OntoClean. Communications of the ACM, **45**(2) (2002) 61-65
14. Guarino, N. and Welty, C.: Identity and subsumption. In: R. Green, C. Bean and S. Myaeng (eds.), The Semantics of Relationships: an Interdisciplinary Perspective. Kluwer (in press) (2002)
15. Loux, M. J.: Metaphysics, a Contemporary Introduction. Routledge (1998)
16. Lowe, E. J.: The possibility of metaphysics. Clarendon Press, Oxford (1998)
17. Masolo, C., Gangemi, A., Guarino, N., Oltramari, A., and Schneider, L.: WonderWeb Deliverable D17: The WonderWeb Library of Foundational Ontologies. (2002)
18. Searle, J.: Intentionality. Cambridge University Press, Cambridge (1983)
19. Simons, P.: Parts: a Study in Ontology. Clarendon Press, Oxford (1987)
20. Simons, P.: How to Exist at a Time When You Have No Temporal Parts. The Monist, **83**(3) (2000) 419-436
21. Strawson, P. F.: Individuals. An Essay in Descriptive Metaphysics. Routledge, London and New York (1959)
22. Varzi: Foreword to the special issue on temporal parts. The Monist, **83**(3) (2000)