# Restructuring WordNet's Top-Level

Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari

National Research Council, Institute for Cognitive Science and Technologies (ISTC-CNR), Italy
gangemi@ip.rm.cnr..it, {Nicola.Guarino, Alessandro.Oltramari, Claudio.Masolo}@ladseb.pd.cnr.it

## 1 Introduction

The number of applications where WordNet is being used as an ontology rather than as a mere lexical resource seems to be ever growing. However, WordNet is only really serviceable as an ontology if some of its lexical links are interpreted according to a formal semantics that tells us something about (our conceptualization of) "the world" and not (just) about the language. One of these more problematic 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 [Guarino 1998].

In the recent years, we developed a methodology for testing the ontological adequacy of taxonomic links called OntoClean [Guarino and Welty 2002a, Guarino and Welty 2002b], which was used as a tool for a first systematic analysis of WordNet's upper level taxonomy of nouns [Gangemi *et al*. 2001]. The early 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* linked to WordNet, which is presented here in some detail, although still in an informal way. This ontology aims at playing the role of a first reference module within a minimal library of *foundational ontologies* that we shall develop within the *WonderWeb*[1] project.

This paper is structured as follows. After a brief discussion of our experimental setting, we discuss in the next section some ontological inadequacies of WordNet's taxonomy of nouns. Then we introduce the most recent version of our ontology, and 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 WordNet's Preliminary Analysis

### 2.1 Experiment Setting

We applied our methodological principles and techniques to the noun synsets taxonomy of WordNet 1.6.To perform our investigation, we had to adopt some preliminary assumptions in order to convert WordNet's databases[2] into a workable knowledge base. At the beginning, we assumed that the hyponymy relation could be simply mapped onto the subsumption relation, and that the synset notion could be mapped into the notion of concept. Both subsumption and concept have the usual description logics semantics [Woods and Schmolze 1992]. In order to work with named concepts, we normalized the way synsets are referred to lexemes in WordNet, thus obtaining one distinct name for each synset: if a synset had a unique noun phrase, this was used as concept name; if that noun phrase was polysemous, the concept name was numbered (e.g. window_1). If a synset had more than one synonymous noun phrase, the concept name linked them together with a dummy character (e.g. Equine$Equid).

First, we created a Loom[3] knowledge base, containing, for each named concept, its direct super-concept(s), some annotations describing the quasi-synonyms, the gloss and the synset topic partition, and its original numeric identifier in WordNet; for example

```
(defconcept Horse$Equus_Caballus
    :is-primitive Equine$Equid
    :annotations ((topic animals)
    (WORD |horse|)
    (WORD |Equus caballus|)
    (DOCUMENTATION "solid-hoofed herbivorous quadruped domesticated since prehistoric times"))
    :identifier |101875414|)
```

| | |
|---|---|
| noun entries | 116364 |
| equivalence classes: synonyms, spelling variants, quasi-synonyms | 50337 |
| noun synsets (with a gloss and an identifier for each one) | 66027 |
| nouns | 95135 |
| monosemous nouns | 82568 |
| polysemous nouns | 12567 |
| one-word nouns | 70108 |
| noun phrases | 25027 |

Table1: Elements processed in the Loom WordNet kb

The elements processed in the Loom WordNet knowledge base are reported in Table 1. We report in Figure 1 an overview of WordNet's noun top-level as translated in our Loom knowledge base. The nine Unique Beginners are shown in boldface.

*In the process of moving to ISTC-CNR, Rome, Italy.

[1] http://wonderweb.semanticweb.org/

[2] We used the Prolog WordNet database, the Grind database, and some others from the official distribution.

[3] Loom is a knowledge representation system that implements a quite expressive description logic [MacGregor 1991].

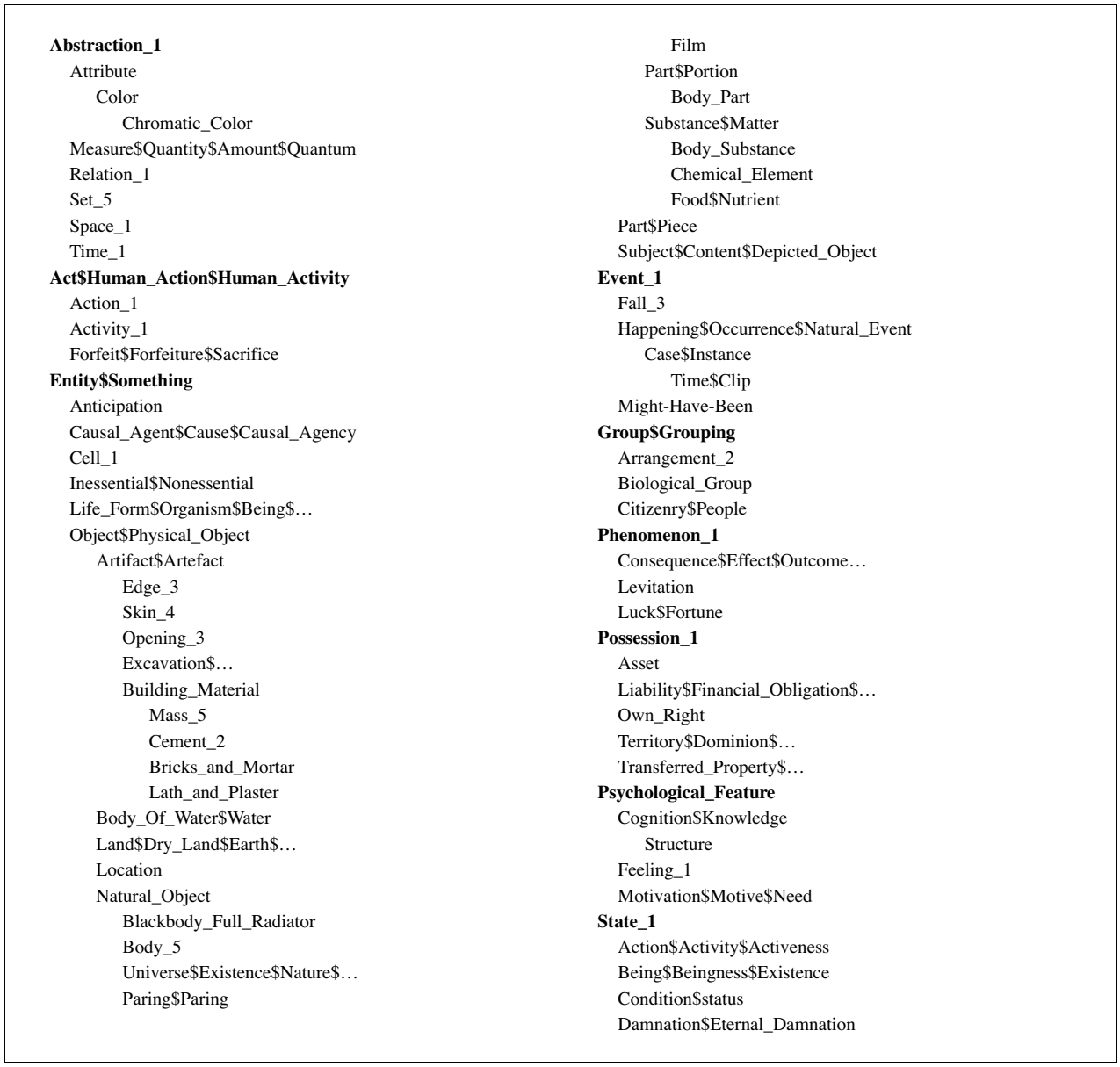


Figure 1: WordNet's top Level

## 2.2 Main problems found

Once the Loom WordNet was created, we systematically applied the OntoClean methodology to the upper taxonomy of noun senses. Let us discuss now the main ontological drawbacks we found after applying this cleaning process.

### 2.2.1 Confusion between concepts and individuals

The first critical point 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.[4] 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 (instan-

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

tiation).

#### 2.2.2 Confusion between object-level and meta-level

The synset Abstraction 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 the result of a psychological process of generalization, in accordance to Locke's position ([Lowe 1998], p.211). This meaning seems to fit the latter group of terms (Attribute, Relation, and possibly some hyponyms of Quantity), but not to 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.

#### 2.2.3 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 upper levels, 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 (like Student) and types (like Person). Roles are *anti-rigid:* every instance of a student can possibly be a non-student. Types, on the other hand, are rigid: every instance of a person must be a person. Therefore, roles cannot subsume types. 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 human beings. 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 is the case 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.

#### 2.2.4 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. It seems that this fact can be explained by the difference between types and roles. 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)[5].

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

Although problematic on the side of ontological distinctions among event-classes, the hyponyms of Phenomenon represent another relevant 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 a role (every event can the consequence of a previous event, but it seems that this is not an essential characteristic of the event itself).

## 3 A biased ontology of particulars

Before presenting our ontology of particulars, a couple of clarifications may be useful. First of all, we do *not* intend this as a candidate for a "universal" standard ontology. Rather, our ontology has a clear *cognitive bias*, in the sense that we aim at capturing the ontological categories lying behind natural language and human commonsense. Hence, we do not claim that our categories have "deep" metaphysical implications related to the intimate nature of the world: rather, they are thought of as "conceptual containers" useful to describe ontologies as cognitive artifacts ultimately depending on human perception, cultural imprints and social conventions. So, especially with respect to natural language, our attitude is more "descriptive" than "revisionary" [Strawson 1959, Loux 1998].

Moreover, even within this general cognitive choice, we are well aware that our particular ontological commitments (and more so the axioms we adopted to formalize them) may not reflect a universal consensus. Indeed, the vision we support is that of a *library of foundational ontologies*, reflecting different commitments and purposes.

The ontology we present here is intended to be a first *reference module* within such a library, intended to allow in the future easy and rigorous comparisons among different approaches. In this view, the most important challenge is not so much the agreement on a monolithic set of ontological categories, but rather the careful isolation of the fundamental ontological options and their formal relationships. If general ontologies reflecting different commitments and purposes are described in terms of these formal notions, making the rationales and alternatives underlying the different choices as explicit as possible, then we can hope they will form different but systematically related modules of the general foundational library.

As a further comment, let us clarify that we are dealing here with an ontology of *particulars*: this means that properties and relations are therefore not included in the

[5] We can draw similar observations for relation and set with respect to abstraction, etc.

domain. Some proposals for an ontology of properties have been made in [Guarino and Welty 2000]. We are not aware of any systematic work on the ontology of relations.

## 3.1 General notions

Before presenting our ontological categories, let us first introduce the general notions we shall use to characterize them. Some of these notions (like rigidity and unity) have already been defined in previous papers (respectively, [Guarino and Welty 2002] and [Gangemi *et al*. 2001]), and will not be discussed here. So we shall limit ourselves to the basic distinction between *enduring and perduring* entities, and the varieties of dependence relationships involving particulars.[6] We shall keep the discussion to an informal, introductory level; a rich axiomatization will be presented in a forthcoming paper.

### 3.1.1 Enduring and perduring entities

A fundamental distinction we assume is that between *enduring* and *perduring* entities. This is almost identical, as we shall see, to the distinction between so-called *continuants* and *occurrents* [Simons 1987], which is still being strongly debated both in the philosophical literature [Varzi 2000] and within ontology standardization initiatives[7]. Again, we must stress 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 [Simons 2000].

But let us see what this distinction is about. The difference between enduring and perduring entities (which we shall also call *endurants* and *perdurants*) is related to their behavior in time. Endurants are always *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 parts (e.g., their previous or future phases) may be not present. For instance, 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).

This different behavior affects the notion of change in time. Endurants can "genuinely" change in time, in the sense that the very same whole endurant 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.

We have already mentioned that endurants and perdurants can be taken as synonyms of the more common terms *continuants* and *occurrents*. We prefer to avoid this terminology because the continuants/occurrents distinction is sometimes considered only within so-called *concrete* entities, while, as we shall see, we take it as spanning the whole domain of particulars, including abstracts that we shall consider as endurants. Finally, we shall take *occurrence*, and not *occurrent*, as synonym of *perdurant*, since it seems natural to use *occurrent* to denote a type (a *universal*), whose instances are occurrences (*particulars*).

The endurants/perdurants distinction evidences the general necessity of temporally indexing the relationships within endurants. This means that, in general, it is necessary to know *when* a specific endurant bears a certain relation to other endurants. Consider for instance the classical example of Tibbles the cat [Simons 1987]: Tail is part of Tibbles before the cut but not after it, i.e. we have to "temporalize" the part relation: $P$(Tail, Tibbles, *before*(cut)) and $\neg P$(Tail, Tibbles, *after*(cut)).

With respect to a temporalized relation $R$, we can distinguish *R-constant* endurants from *R-variable* endurants. An endurant $e$ is called *R-constant* iff, when $R(x_1, \ldots, x_n, e, t)$ holds for a temporal interval $t$, then $R(x_1, \ldots, x_n, e, t')$ also holds whenever $e$ is present at $t'$.

We can also strengthen this definition introducing the modal notion of an *R-invariant* endurant. An endurant $e$ is called *R-invariant* iff, if it is possible that $R(x_1, \ldots, x_n, e, t)$ then necessarily $R(x_1, \ldots, x_n, e, t)$ holds whenever $e$ is present at $t'$.

For the purpose of characterizing the ontological categories here sketched, the property of being constant (or invariant) with respect to the parthood relation (*mereologically constance/invariance*) has a special relevance. For example, we usually take ordinary material objects as mereologically variable, because during their life they can lose or gain parts. On the other hand, amounts of matter are taken as mereologically invariant (all their parts are *essential part*), and so on.

### 3.1.2 Dependence

Let us now introduce informally some useful definitions based on the notion of dependence, adapted from [Thomasson 1999]. We focus here on *ontological dependence* (holding primarily between particulars, and only by extension between properties), to be distinguished from *notional dependence*, which only holds between properties.

A particular $x$ is *specifically constantly dependent* (SCD) on another particular $y$ iff, at any time $t$, $x$ can't be present at $t$ unless $y$ is also present at $t$. For example, a person might be specifically constantly dependent on her/his brain.

A particular $x$ is *generically constantly dependent* (GCD) on a property $\phi$ iff, at any time $t$, $x$ can't be present at $t$, unless a certain instance $y$ of $\phi$ is also present at $t$. For example, a person might be generically constantly dependent on having a heart.

[6] Only dependence between properties is used in the OntoClean methodology.

[7] See for instance the extensive debate about the "3D" vs. the "4D" approach at suo.ieee.org

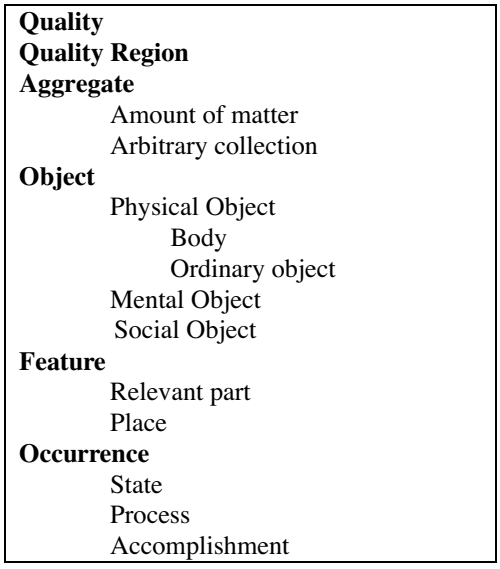


Figure 2: The top categories.

## 3.2 The Top Categories

The most general kinds of particulars assumed here are described in Figure 2. They are assumed to be mutually disjoint, and covering the whole domain of particulars. They are also considered as *rigid* properties, according to the OntoClean methodology that stresses the importance of focusing on these properties first.

### 3.2.1 Qualities and quality regions

'Quality' is often used as a synonymous of 'property', but this is not the case in the current ontology: qualities are particulars, properties are universals. According to our view, every entity comes with certain qualities, which exist exactly as long as the entity exists. These qualities belong to different *quality types* (like color, size, smell, etc.), and are characteristic for (*inhere to*) specific individuals: no two particulars can have the same quality, and each quality is *specifically constantly dependent* on the entity it inheres to. 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*) [Gärdenfors 2000], 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 qualities.

This distinction between qualities and qualia is inspired by [Goodman 1951] and the so-called *trope theory* [Campbell 1990] (with some differences that can't be discussed here[9]). Its intuitive rationale is mainly due to the fact that natural language – in certain constructs – seems often 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.[10]

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 cognive bias.

Under 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"): 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).

*Space and time locations as special qualities*

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).

Notice that quality regions can have qualities themselves (for instance, the spatial location of a certain object can have a shape). In particular, we assume that all quality regions are temporally located, and that their temporal qualia coincide with the temporal universe, i.e. quality regions are always present.

### 3.2.2 Aggregates

The common trait of aggregates is that they are endurants and 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[11] 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 ab-

[8] An important difference is that standard tropes 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".

[9] An important difference is that standard tropes 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".

[10] The possibility of talking of qualia as particulars rather than reified properties is another advantage of our approach.

[11] We assume here that a member is a special part of a collection, see [Gangemi *et al*. 2001]

stract entities, and *group* for something having an intrinsic unity.

#### 3.2.3 Objects

The main characteristic of objects is that all of them are endurants and essential wholes. They have no *common* unity criterion, however, as different subtypes of objects may have different unity criteria. Often objects (indeed, all endurants) are considered as ontologically independent from occurrences (discussed below). But, if we admit that every object has a life, it is hard to exclude a mutual specific constant dependence between the two. Nevertheless, we can still use the notion of dependence to (weakly) characterize objects as being not specifically constantly dependent *on other objects*.

Within objects, we distinguish those that have a spatial location (*physical objects*) from those that do not *(abstract objects)*. Among physical objects, we further distinguish between *physical bodies* and *ordinary objects*. Bodies are mereologically invariant, and then they are material objects in the sense of physics.[12]. Ordinary objects have a more cognitive nature, as they are admitted to change some of their parts while keeping their identity: they can have therefore *temporary parts*.

Among abstract objects, special relevance have those depending on mental "products" of a single intentional being, like (singular) intentions, believes, competences, feelings and so on, here collected by the very general category *mental object*. On the other hand, those depending on more than one intentional being are called *social objects*. Within them, special relevance have those depending on *conventions stipulated by a community of persons*, like projects, legal norms, moral values, aesthetic notions, ecc.

#### 3.2.4 Features

Typical examples of features are "parasitic entities" such as holes, bumps, surfaces, or stains, which are generically constantly dependent on physical objects[13] (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).

#### 3.2.5 Occurrences

Occurrences are synonymous of perdurants. They comprise what are variously called events, processes, happenings, and states. Occurrences 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* to them.

Within occurrences, we consider two main ontological distinctions: homeomery and relationality. The first dimension has been introduced by Parsons, Cresswell, and Mourelatos [Casati and Varzi 1996]: intuitively, we say that an occurrence is homeomeric iff all its temporal parts can be described *in the same way* used for the whole occurrence: for instance, every temporal part of "my sitting here" for an hour is still a "sitting here of mine". But if we consider "Messner's ascent to Everest" (intended in the complete sense), no parts of it are a "Messner's ascent to Everest". To formalize this notion, we need to refer to a certain property that holds for all the temporal parts of a certain occurrence *o*. We individuate this property by considering the most specific *occurrent* of *o*, i.e. the most specific occurrence type *o* is instance of. Then we can say that *o* is homeomeric iff all its temporal parts are instances of the same most specific occurrent. The second distinction takes inspiration mainly from [Smith 1982]. An occurrence is said *non-relational* when only one object participates to it, while it is *relational* when it has two or more objects as participants. Occurrences involving qualities varying in time (i.e., which can change their qualia in time) are prototypical examples of non-relational occurences: the change of color of a rose has only one object as a participant (there may be other participants, such as the rose's color, but this is a quality and not an object). In our proposal, homeomery seems to be enough to account for the distinctions proposed in the literature (especially [Mourelatos 1996]) among *states*, *processes*, and *accomplishments*.

It is easy to see that states are homeomeric occurrences (e.g., "the air smelling of jasmine"), while *accomplishments* are non-homeomeric (e.g. "the sunset"). Processes can be characterized as *weakly non-homeomeric*, in the sense that *some* temporal parts of them are instances of the same most specific *occurrent*, and some are not. For instance, in the case of "running", if you consider that instantaneous temporal part of your running through the park in which your right foot touches the ground while your left foot does not (think about photo-finish in a race), this sub-event is no more a "running". Together, processes and accomplishments are often described as *dynamic events*, just because of an (apparent) change of some of their properties across their different temporal parts.

## 4 Mapping WordNet synsets

Let us consider now the results of mapping the WordNet top synsets into the ontological categories we have presented. 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.

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

[12] Notice that, differently from amounts of matter, physical bodies are essential wholes.

[13] We may think that features are specifically constantly dependent on their host, but an example like "a whirlpool" is very critical in this sense. Notice that we are not considering as features entities that are dependent on mental-objects.

elsewhere, like Anticipation, Imaginary_Place, Inessential, etc. Such synsets have only a few children, and these have been already excluded in our analysis.
The results of our integration work are sketched in Table 2. Our categories are reported in the first column; the second column shows the WordNet synsets that are *covered* by such categories (i.e., they are either equivalent to or included by them); the third column shows some hyponyms of these synsets that were rejected according to our methodology. Finally, the last column shows possible further hyponyms that have been appended under our categories, coming from different places in WordNet. The problems encountered for each category are discussed below.

### 4.1 Aggregates, Physical Objects, and Features

Aggregate, Object and Feature acquire their children mainly from the hyponyms of the Unique Beginner Entity. Entity is a very confused synset and only few subordinate concepts are usable in our ontology (Physical_Object, Substance$matter, Imaginary_place,...[14]): a lot of hyponyms have to be "rejected", such as roles (Causal_Agent, Subject_4), unclear synsets (Location[15]) and so on.
The branch formed by our ontological category Feature has been populated with suitable hyponyms of Physical_Object.
By removing roles like Arrangement and Straggle, Group$grouping becomes a partition of the Ordinary Object category. In fact, hyponyms like Collection, Biological_Group, Kingdom, and so on, are nothing but plural objects, supporting a clear unity criterion (on the contrary, arbitrary collections are pluralities with no unity criterion, namely arbitrary sums of *things*:; interesting enough, these strange entities are not lexicalized in WordNet).

### 4.2 Mental objects, Social objects and Qualities

Abstraction is the most heterogeneous Unique Beginner: mental objects such as Chromatic_Color (an example of quality space[16]), qualities (mostly from the synset Attribute), a hybrid concept (Relation) that contains mental objects, concrete entities (as Substance[17]), and even meta-level categories (see §2.2.2). Each child synset has been mapped appropriately.
Psychological_feature contains both mental objects (Cognition[18], Feeling) and events (at the third level, under Feeling, you can find Emotional_state, Cognitive process, ecc.). We consider Motivation as a material role, so to be added to lower levels of the taxonomy of mental objects. Under Mental Object we add also Imaginary_place, a synset from Entity.
The Unique Beginner Possession has been mapped entirely to the category Social Object. This choice depends on the nature of the hyponyms of Possession, which refer globally to entities set up by financial or legal conventions.
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 desiderata in the WordNet community (see [Fellbaum 1998], p. 66). In Table 2 and Figure 3 we present some qualities found out among the hyponyms of Attribute. Since in WordNet there isn't a clear lexicalized concept for quality regions, and since they are in a sense "abstractions" from qualities, in some cases we decided to duplicate the synset: i.e., you will find chromatic_color-quality under Quality and chromatic_color-region under Quality Region.

### 4.3 Occurences

Since we focused on nouns, and the ontology of occurrences mainly influences the classification of verbs, our discussion will be necessarily limited here. Event, Phenomenon, State and Act are the Unique Beginners of those branches of WordNet denoting occurrences. Under Act we find in general events of two kinds: processes (see activity and its hyponyms) and accomplishments (see the homonymous synset under action). Event has a too much generic composition in order to be partitioned clearly 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

The final results of our mapping are sketched in Figure 3. As one can see, a substantial taxonomy rearrangement has been performed. The blind application of OntoClean's taxonomy evaluation methodology provided a first guideline, but stronger ontological commitments seemed to be unavoidable in order to get a "disciplined" taxonomy. In our opinion, strong (and explicit) ontological distinctions do also reduce the risk of classification mistakes in the ontology development process, and simplify the update and maintenance process.
WordNet is largely used mainly because of its coverage, and has proven to be a key resource in many strategic applications. What we are curious to see now is whether a "principled" restructuring like the one we have proposed will have some positive impact on the performance of these applications. Unfortunately we have no glues yet.

## 6 Acknowledgements

We would like to thank Stefano Borgo and Luc Schneider for the fruitful discussions and comments on the final 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).

[14] Compare figure 1 with table 2.
[15] Under Location we find roles (There, Here, Home, Base), individuals (Earth), geometric concepts like Line, Point, etc.
[16] By looking to the corresponding hyponyms, it becomes clear that this synset could also be viewed as denoting a quality (by means of this we decide to append it both under Quality and Quality Region top concepts).
[17] "The stuff of which an object consists".
[18] "The psychological result of perception, and learning and reasoning".

## References


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

Casati, R. and Varzi, A. (eds.) 1996. *Events*. Dartmouth, Aldershots, USA.

Fellbaum, C. (ed.) 1998. *WordNet - An Electronic Lexical Database*. MIT Press.

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., and Oltramari, A. 2001. 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: 285-296.

Gärdenfors, P. 2000. *Conceptual Spaces: the Geometry of Thought*. MIT Press, Cambridge, Massachussetts.

Goodman, N. 1951. *The Structure of Appearance*. Harvard University Press, Cambridge MA.

Guarino, N. 1998. 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: 527-534.

Guarino, N. and Welty, C. 2000. 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: 97-112.

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).

Loux, M. J. 1998. *Metaphysics, a Contemporary Intro duction*. Routledge.

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

MacGregor, R. M. 1991. Using a Description Classifier to Enhance Deductive Inference. In *Proceedings of Seventh IEEE Conference on AI Applications*: 141-147.

Mourelatos, A. 1996. Events, Processes, States. In R. Casati and A. Varzi (eds.), *Events*. Dartmouth Publishing Company, Aldershot: 457-476.

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

Simons, P. 2000. How to Exist at a Time When You Have No Temporal Parts. *The Monist*, **83**(3): 419-436.

Smith, B. (ed.) 1982. *Parts and Moments: Studies in Logic and Formal Ontology*. Philosophia Verlag, München.

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

Thomasson, A. L. 1999. *Fiction and Metaphysics*. Cambridge University Press, Cambridge.

Varzi 2000. Foreword to the special issue on temporal parts. *The Monist*, **83**(3).

Woods, W. A. and Schmolze, J. G. 1992. The KL-ONE family. In F. W. Lehmann (ed.) *Semantic Networks in Artificial Intelligence*. Pergamon Press, Oxford: 133-177.

| *OCT Top Categories* | **Covered Synsets** | **Rejected Hyponyms** | **Imported Hyponyms** |
|---|---|---|---|
| *Quality* | attribute* | trait, ethos, inheritance, … | |
| *Temporal Location* | time_interval$interval* | eternity, greenwich_mean_time, present, past, future,… | |
| *Spatial Location* | position$place | | |
| *Color* … | chromatic_color | | |
| *Quality Region* | attribute* | trait, ethos, inheritance, … | |
| *Time Region* | time_1, time_interval$interval* | eternity, greenwich_mean_time, present, past, future,… | |
| *Space Region* | space_1* | subspace, … | |
| *Color Region* … | chromatic_color | | |
| *Aggregate* | aggregate_2 (**!**) | | |
| *Amount of Matter* | =substance$matter* | bedding_material, ballast, atom, … | mass_5 , cement_2, substance_4, ball_7 air_3 ,… |
| *Arbitrary Collection* | | | |
| *Object* | | | |
| *Physical Object* | | | |
| *Body* | body_5*; blackbody$full_radiator, universe$existence$nature$creation | mass_5 | |
| *Ordinary Object* | physical object_1*; group* | finding_3, catch_4, vagabond_1, substance_1, location_1; circuit_5, law_1, political system_1, economic system_1, language system_1, social system_1 | belt_3, aerospace_1, atmosphere_3, d-region_1, e-region_1, f-region_1, heliosphere_1, ionosphere_1, black hole_1, organism_1, cell_2, whole_2, unit_7 |
| *Mental Object* | cognition_1*; feeling_1*; imaginary place_1 | food for thought_1, kernel_3, noumenon_1, equivalent_1, cognitive process_1, unconscious process_1, structure_3, public knowledge_1; emotional state_1 | |
| *Social Object* | possession_1; circuit_5, law, political system_1, economic system_1, language system_1, social system_1; person_1 | | |
| *Feature* | | | |
| *Relevant Part* | edge_3, bump_2; bottom_3, depth_2, top_1; paring_2, peak_5, surface_1 | | |
| *Place* | =region$part* | atmosphere_3, aerospace_1, air_3, belt_3, bottom_3, depth_2, distance_2, d-region_1, e-region_1, f-region_1, eden_1, heliosphere_1, ionosphere_1, extremity_4, hell_2, mare_2, zodiac_1, sign of the zodiac_1, top_1, black hole_1 | hole_1, hole_2,hole_4, hole_5, rear_3, front_10, right_2, left_1 orifice_1, opening_1, excavation_3, cave_1, blank space_1 |
| *Occurrence* | | | |
| *State* | state_5* | utopia, dystopia, nature | |
| *Process* | process*; activity_1 | consequence | cognitive process_1, unconscious process_1, emotional state_1 |
| *Accomplishment* | accomplishment$achievement | | |

Table 2: Synsets marked with '*****' are heterogeneous (some of their children are to be moved elsewhere, some are roles, or some are instances); those marked with '(**!**)' have no hyponyms; those in upper case are WordNet Unique Beginners.

**Quality**
- position$place
- time_interval$interval
- chromatic_color
- …

**Quality Region**
- space_1
- time_1
- time_interval$interval
- chromatic_color
- …
- **Aggregate**
- **Amount of matter**
  - body_substance
  - chemical_element
  - mixture
  - compound$chemical_compound
  - …
  - mass_5
  - cement_2
  - air_3
  - substance_4
  - …
- **Arbitrary collection**
  - …

**Object**
- **Physical Object**
  - **Body**
    - blackbody$full_radiator
    - body_5
    - universe$existence$nature$creation
    - …
  - **Ordinary Object**
    - natural_object
    - artifact$artefact
    - body_of_water$water
    - land$dry_land$earth$…
    - body$organic_structure
    - …
    - organism_1
    - whole_2
    - …
    - collection$aggregation
    - biological_group
    - kingdom

**Abstract object**
- **Mental Object**
  - imaginary_place_1
  - ability_2
  - structure_3
  - cognitive content_1
- **Social Object**
  - possession_1
  - circuit_2
  - law
  - social_system_1
  - economic_system_1
  - person_1
  - language system_1
- …

**Feature**
- **Relevant Part**
  - edge_3
  - bump_2
  - peak_5
  - surface_1
  - bottom_3
  - depth_2
  - top_1
  - skin_4
  - paring$parings
  - …
- **Place**
  - opening_1
  - excavation$hole_in_the_ground
  - cave_1
  - right_2
  - left_1
  - …
  - interplanetary_space
  - zone_3
  - outside_1
  - inside_1
  - …

**Occurrence**
- **State**
  - cognitive_state
    - situation_1
    - condition_1
- **Process**
  - process_1
  - activity_1
  - cognitive process_1
  - unconscious process_1
- **Accomplishment**
  - Accomplishment$achievement

Figure 3: WordNet cleaned up: mapping WordNet into our top categories.