Restructuring WordNet's Top-Level: The *OntoClean* approach

Alessandro Oltramari[(1)], Aldo Gangemi[(2)], Nicola Guarino[(1)], Claudio Masolo[(1)]
[(1)]LADSEB-CNR*, Padova, Italy: {Nicola.Guarino, Alessandro.Oltramari, Claudio.Masolo}@ladseb.pd.cnr.it
[(2)]ISTC-CNR, Rome, Italy: gangemi@ip.rm.cnr.it

Abstract

In this paper we propose an analysis and a rearrangement of *WordNet's* top-level taxonomy of nouns. We briefly review WordNet and identify its main semantic limitations, in the light of the ontology evaluation principles lying at the core of the *OntoClean* methodology. Then we briefly present a first version of the *OntoClean Top* (OCT) ontology, and show how WordNet can be aligned with it. The result is a "cleaned-up" WordNet, which is meant to be conceptually more rigorous, cognitively transparent, and efficiently exploitable in several applications.

1 Introduction

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 interpreted according to some formal semantics, which tells us something about "the world" and not (just) about the language. 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 (Guarino, N., 1998).

In the recent years, we developed a methodology for testing the ontological adequacy of taxonomic links called OntoClean (Guarino, N. & Welty, C., 2002; Guarino, N. & Welty, C., 2002), which was used as a tool for a first systematic analysis of WordNet's upper level taxonomy of nouns (Gangemi, A. *et al.*, 2001). The first version of OntoClean was based on an ontology of properties (unary *universals*), characterized by means of meta-properties. We are now extending OntoClean with an ontology of *particulars* called OCT (OntoClean Top ontology), which is presented here in some detail, although still in an informal way. The OCT will be the first module of a minimal library of *foundational ontology* that we shall develop within the *WonderWeb*[1] project.

This paper is structured as follows. In the next section we present an extension of our FOIS paper (Gangemi, A. *et al.*, 2001), concerning some ontological inadequacies of WordNet's taxonomy of nouns. Then we introduce the most recent version of our OntoClean Top 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, W. A. & Schmolze, J. G., 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).

Firstly, 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

*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, R. M., 1991).

base are reported in Table 1. We report in Figure 2 an overview of WordNet's noun top-level as translated in our Loom knowledge base. The nine Unique Beginners are shown in boldface.[4]

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.[5] 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, Tammany Society[6], 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).

2.2.2 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 ((Lowe, E. J., 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 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 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 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.

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. 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)[7].

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,

[4] Note that the sense numeration reported in our Loom kb is different from the WordNet's original one. Nevertheless, the reader will easily recognize the synsets we are referring to.

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

[6] “A political organization in New York city (late 1800’s early 1900’s) seeking political control by corruption and bossism”.

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

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[8]).
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.

3 The OntoClean Top Ontology

Before presenting our (still preliminary!) OCT ontology, a couple of clarifications may be useful. First of all, we do *not* intend this as a candidate for a "universal" standard ontology. Rather, we support the vision of a *library* of *foundational ontologies*, reflecting different commitments and purposes. In our opinion, the most important challenge today 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, then we can hope they will form a library of "foundational" ontologies accessible in a modular way, keeping the necessity of largely shared ontological commitments to the very minimum, and making the rationales and alternatives underlying the different ontological choices as explicit as possible. This is one of the goals of the *WonderWeb* project, where the OCT ontology will be linked to other foundational ontologies.
A second clarification concerns the general attitude underlying our ontological choices. The OCT 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, P. F., 1959; Loux, M. J., 1998).
Finally, we have to point out that the ontology presented here is an ontology of *particulars*. Properties and relations are therefore not part of its domain. Some proposals for a ontology of properties have been made in (Guarino, N. & Welty, C., 2000). We are not aware of any systematic work on the ontology of relations.

[8] For instance, the extinction of dinosaurs could have be the consequence of the impact of an asteroid on the Earth, or of a sudden glaciation, or of a mortal epidemic – scientists are not sure about this – but in terms of ontology of events, it is a conclusive event, at most an annihilation event, and there is no need (and here no possibility) to model it as a consequence.

3.1 General notions

Before introducing the OCT 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, N. & Welty, C., 2002) and (Gangemi, A. *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.[9] 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 in the OCT ontology 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, P., 1987), which is still being strongly debated both in the philosophical literature (Varzi, A., 2000) and within ontology standardization initiatives[10]. 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, P., 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 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 travel 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

[9] In the OntoClean taxonomy evaluation methodology only dependence between properties is used.
[10] See for instance the extensive debate about the "3D" vs. the "4D" approach at www.suo.org.

continuants and *occurrents*. We prefer however the adopted 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, P., 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 ¬*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 $R(x_1, \dots, x_n, e, t)$ holds for a temporal interval *t*, then $R(x_1, \dots, x_n, e, t')$ 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 $R(x_1, \dots, x_n, e, t)$ then necessarily $R(x_1, \dots, x_n, e, t)$ holds whenever *e* is present at *t'*.

For the purpose of characterizing the OCT categories, the property of being constant (or invariant) with respect to the parthood relation (*mereologically constant* (*invariant*)) 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, A. L., 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 its brain.

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

1.2 The OntoClean Top Categories

The most general kinds of particulars assumed in the OntoClean Top ontology are described in Figure 1. 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.

Quality
Quality Region
Aggregate
- Amount of matter
- Arbitrary collection

Object
- Physical Object
  - Body
  - Ordinary object
- Mental Object

Feature
- Relevant part
- Place

Occurrence
- State
- Process
- Accomplishment

Abstract

Figure 1: Onto Clean Top Categories.

1.2.1 Qualities and quality regions

'Quality' is often used as a synonymous of 'property', but this is not the case in the OCT 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 (*inhere to*) specific individuals: no two particulars can have the same quality. 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 "extension" (or "classification") of an individual quality with respect to a certain *conceptual space* (called here *quality space*) (Gärdenfors, P., 2000), So when we say that two roses have the same color their two colors are classified in the same way wrt the color space (they have the same *color quale*), but still they have two numerically distinct qualities.

The reason of this distinction between qualities and qualia, which is inspired to the theory of tropes (with some differences that can't be discussed here[11]), 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 an atom in the color space.[12]

[11] 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".

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

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.
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).
As a final remark, we note that qualities are assumed to be as specifically constantly dependent on the entities they *inhere to*.

1.2.1.1 Location

In the OCT 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* and it belongs to the associated quality space (respectively geometric space and temporal space). 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 occurence. 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.

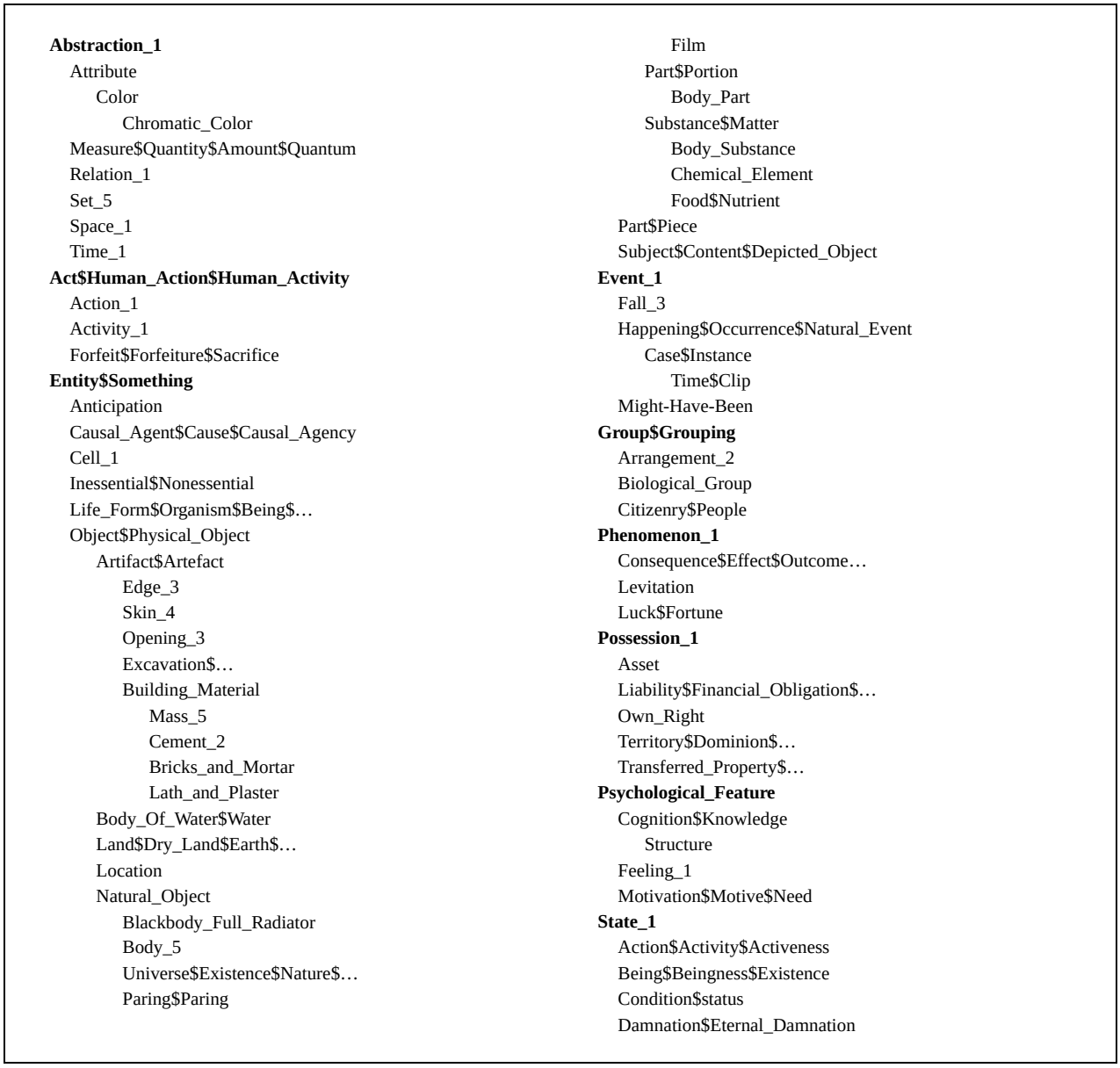

Figure 2: WordNet's top Level

1.2.2 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 which are not themselves essential wholes (like the sum of a person's nose and a computer keyboard). They are essentially mereologically *pseudo-constant*, in the sense that they change their identity when a member (i.e. a special part of a collection, see (Gangemi, A. *et al.*, 2001)) is changed, while a change in the non essential parts of a member is allowed. We may have called arbitrary collections *groups*, or perhaps *sets*; but we prefer to use *set* for abstract entities, and *group* for something having an intrinsic unity.

1.2.3 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 ontologically independent from occurrences (discussed below). But, if we admit that every object has a life, it is hard to exclude a mutual ontological dependence between the two. Nevertheless, we can use the notion of dependence to distinguish between objects that are not specifically constantly dependent on other objects and have a spatial location (*physical objects*) and objects that are generically constantly dependent on *persons* (that are also objects) and do not have a spatial location (*mental objects*). Among physical objects, we further distinguish between *bodies* and *ordinary objects*. Bodies are mereologically invariant, and then they are material objects in the sense of physics.[13]. Ordinary objects (and mental objects even more) 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 mental objects, we could distinguish between purely *subjective mental objects*, i.e. objects depending on a singular person (like an intention, or a competence), and *intersubjective mental objects*, i.e. objects depending on a community of persons (like a project, a legal norm, a moral value, an aesthetic notion).

1.2.4 Features

Typical examples of features are "parasitic entities" such as holes, bumps, surfaces, or stains, which are generically constantly dependent on physical objects[14] (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.

[13] Notice that differently from the amounts of matter they are essential whole.

[14] 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.

1.2.5 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. Clearly objects can't be parts of occurrences, rather they *participate* to them.

Within occurrences, we consider two main ontological dimensions of distinction: homeomery and relationality. The first dimension has been introduced by Parsons, Cresswell, and Mourelatos (see (Casati, R. & Varzi, A., 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 dimension takes inspiration mainly from (Smith, B., 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, A., 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.

In any case, we can further divide each of these categories into relational and non-relational occurrences.

1.2.6 Abstracts

Like mental-object and their qualities, abstracts are enduring entities that do not have a spatial location (indeed they do not have any "physical quality"). Differently from mental-object and their qualities, abstracts are independent from objects (and in particular from persons). Exam-

ples of abstracts are *sets, symbols, propositions, structures*, and *physical laws.*

4 Mapping WordNet into the OCT ontology

Let us consider now the results of integrating the WordNet top concepts into our top-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.

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.

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 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, Objects, and Features

Entity is a very confused synset. As sketched in the table, a lot of its hyponyms have to be "rejected": in fact there are roles (Causal_Agent, Subject_4), unclear synsets (Location[15]) 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 new top concept Feature.

By removing roles like Arrangement and Straggle, Group$grouping becomes a partition of the Ordinary Object category. In fact, hyponyms like Collection, Social_Group, Biological_Group, and so on, are nothing but plural objects, supporting a clear unity criterion.

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 is 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, mental objects such as Chromatic_Color (an example of quality space[16]), qualities (mostly from the synset Attribute) and a hybrid concept (Relation_1) that contains mental objects, concrete entities (as Substance_4[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]) 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 desiderata in the WordNet community (see (Fellbaum, C., 1998), p. 66).

4.3 Occurences

Event_1, Phenomenon_1, State_1 and Act_1 are the Unique Beginners of those branches of WordNet denoting events. WordNet does not support the distinction between relational and non-relational occurrences, so first of all, in order to restructure this partition of the top level, we need to separate the hyponyms of the above-mentioned four synsets by means of our defined first dimension. We see, for example, that State_1 maps in part to non-relational state (condition$status, cognitive_state, existence, death_4, degree, skillfulness…), in part to relational state (medium_4, relationship_1 and relationship_2, disorder, order, hostility, conflict…). We register a similar behavior for the children of Process (a subclass of Phenomenon_1): decrement_2, increment and shaping could be seen as kinds of process involving a single main participant, while chelation, economic_process, execution and some hyponyms of Natural_Process (a direct hyponym of Process) seem to denote relational occurrences. Under Act_1 we find in general events of two kinds: processes (see activity_1 and its hyponyms) and accomplishments (see the homonymous synset under action_1). For sake of simplicity, we consider the hyponyms of Act_1 as being both relational and non-relational, depending on the context in which they are used. Event_1 has a too much generic composition in order to be partitioned clearly in terms of our approach (see, for instance, the beginning of §2.2.1): 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 integration effort are sketched in Figure 3. Our results show that a serious taxonomy rearrangement is needed. The blind application of OntoClean's taxonomy evaluation methodology provides a first guideline, but stronger ontological commitments seem 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.

Our research is still in progress: we hope we have paved

[15] Referring to Location, we find roles (There, Here, Home, Base, Whereabouts), instances (Earth), and 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".

the way for future work and possible cooperation.

6 Acknowledgements

We would like to thank Stefano Borgo and Luc Schneider 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).

7 References

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

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

Gangemi, A. *et al.* (2001). Understanding top-level ontological distinctions: In Proceedings of IJCAI-01 Workshop on Ontologies and Information Sharing (26-33). Seattle, USA, AAAI Press.

Gangemi, A. *et al.* (2001). Conceptual Analysis of Lexical Taxonomies: The Case of WordNet Top-Level. In C. Welty & S. Barry (Eds.), Formal Ontology in Information Systems. Proceedings of FOIS2001 (285-296). , ACM Press.

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

Guarino, N. (1998). Some Ontological Principles for Designing Upper Level Lexical Resources. In A. Rubio *et al.* (Eds.), Proceedings of First International Conference on Language Resources and Evaluation (527-534). Granada, Spain, ELRA - European Language Resources Association.

Guarino, N. & Welty, C. (2000). A Formal Ontology of Properties. In R. Dieng & O. Corby (Eds.), Knowledge Engineering and Knowledge Management: Methods, Models and Tools. 12th International Conference, EKAW2000 (97-112). France, Springer Verlag.

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

Guarino, N. & Welty, C. (2002). Identity and subsumption. In R. Green *et al.* (Eds.), The Semantics of Relationships: an Interdisciplinary Perspective . , Kluwer (in press).

Loux, M. J. (1998). Metaphysics, a Contemporary Introduction. , Routledge.

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

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 & A. Varzi (Eds.), Events (457-476). Aldershot, Dartmouth Publishing Company.

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

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. München, Philosophia Verlag.

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

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

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

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

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, …
Arbitrary Collection			
Object	ENTITY$SOMETHING*	Anticipation, Causal_Agent, Imaginary_Place, Substance	
Physical Object			
Body	Natural_Object*	Dead_Body, Constellation, Stone, Nest, …	
Ordinary Object	Physical_Object*, Group*	Finding, Catch, Vagabond; Arrangement, Social_Group, …	
Mental Object	PSYCHOLOGICAL_FEATURE*	Feeling_1, Motivation_1	Own_Right (!), Social_Group
Feature			
Relevant Part	Part$portion*, Fragment	Substance_4	Edge_3, Skin_4, Paring$Parings, …
Place			Opening_3, Excavation$hole_in_the_Ground, …
Occurrence	STATE_1*, PHENOMENON_1*, ACT*	Utopia, Dystopia, Nature, Consequence, Stay_1, …	
State	STATE_1*	Utopia, Dystopia, Nature	
Non-relational	Condition$status, Cognitive$State, Existence, Death_4, Degree, …		
Relational	Medium_4, Relationship_1, Relationship_2, Order, Disorder, Hostility, Conflict, …		
Process	Process, Activity_1		
Non-relational	Decrement_2, Increment, Shaping		
Relational	Chelation, Execution, …		
Accomplishment	Accomplishment$achievement		
Non-relational			
Relational			
Abstract			Statement_1, Cognition, Arrangement_2,
Proposition	Proposition_1		
Set …	Set_5		

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
    - fluid_1
  - **Arbitrary collection**
    - …
- **Object**
  - **Physical Object**
    - **Body**
      - blackbody$full_radiator
      - body_5
      - universe$existence$nature$creation
      - …
    - **Ordinary Object**
      - collection$aggregation
      - biological_group
      - kingdom
      - …
      - body_of_water$water
      - landdry_landearth$…
      - body$organic_structure
      - artifact$artefact*
      - life_form$organism$being$…
  - **Mental Object**
    - cognition$knowledge
      - structure
      - …
    - own_right
    - social_group
    - …
- **Feature**
  - **Relevant Part**
    - edge_3
    - skin_4
    - paring$parings
    - …
  - **Place**
    - opening_3
    - excavation$hole_in_the_ground
    - …
- **Occurrence**
  - **State**
    - **Non-relational**
      - condition$status
      - cognitive_state
      - existence
      - death_4
      - degree
      - …
    - **Relational**
      - medium_4
      - relationship_1
      - relationship_2
      - conflict
      - …
  - **Process**
    - **Non-relational**
      - decrement_2
      - increment
      - shaping
      - activity_1
      - …
    - **Relational**
      - chelation
      - execution
      - activity_1
      - …
  - **Accomplishment**
    - **Non-relational**
      - accomplishment$achievement
      - …
    - **Relational**
      - accomplishment$achievement
      - …
- **Abstract**
  - statement_1
    - proposition
    - …
  - symbol
  - set_5
  - …

Figure 3: WordNet cleaned up: mapping WordNet into the OntoClean top-level.