Sweetening WordNet with DOLCE

Aldo Gangemi, Nicola Guarino, Claudio Masolo, and Alessandro Oltramari

■ Despite its original intended use, which was very different, WordNet is used more and more today as an ontology, where the hyponym relation between word senses is interpreted as a subsumption relation between concepts. In this article, we discuss the general problems related to the semantic interpretation of WordNet taxonomy in light of rigorous ontological principles inspired by the philosophical tradition. Then we introduce the DOLCE upper-level ontology, which is inspired by such principles but with a clear orientation toward language and cognition. We report the results of an experimental effort to align WordNet's upper level with DOLCE. We suggest that such alignment could lead to an "ontologically sweetened" WordNet, meant to be conceptually more rigorous, cognitively transparent, and efficiently exploitable in several applications.

The number of applications where WordNet (Fellbaum 1998) 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 referential semantics that tells us something about (our conceptualization of) "the world." One such link 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 recent years, we developed a methodology for testing the ontological adequacy of taxonomic links called OntoClean (Guarino and Welty 2002a, 2002b), which was used as a tool for a first systematic analysis of WordNet's upper-level taxonomy of nouns (Gangemi, Guarino, and Oltramari 2001). OntoClean was based on an ontology of properties (unary universals), characterized by means of metaproperties. We are now complementing OntoClean with an ontology of particulars called a DESCRIPTIVE ONTOLOGY FOR LINGUISTIC AND COGNITIVE ENGINEERING (DOLCE), which is presented here in some detail, although 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 are developing within the WonderWeb[1] project.

This article is structured as follows: After a brief discussion of our experimental setting, we discuss some ontological inadequacies of WordNet's taxonomy of nouns. Then we introduce the basic assumptions and distinctions underlying DOLCE and discuss the preliminary results of an alignment effort aimed at improving WordNet's overall ontological (and cognitive) adequacy and facilitating its effective deployment in practical applications.

WordNet's Preliminary Analysis

We applied our analysis to the noun synsets taxonomy of WordNet 1.6. To perform our investigation, we had to adopt some preliminary assumptions to convert WordNet's databases into a workable knowledge base.

```
(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|)
```

Figure 1. Example from the LOOM WORDNET Knowledge Base.

Experiment Setting

At the beginning, we assumed that the hyponymy relation could simply be 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 logic semantics (Woods and Schmolze 1992). To work with named concepts, we normalized the way synsets are referred to as lexemes in WORDNET, thus obtaining one distinct name for each synset: If a synset had a unique noun phrase, it was used as a concept name; if this noun phrase was polysemous, the concept name was numbered (for example, window_1). If a synset had more than one synonymous noun phrase, the concept name linked them together with a dummy character (for example, Equine$Equid).

Then we created a LOOM knowledge base (MacGregor 1991), containing, for each named concept, its direct superconcept(s), some annotations describing the quasisynonyms, the gloss and the synset topic partition, and its original numeric identifier in WORDNET; figure 1 is an example.

The elements processed in the LOOM WORDNET knowledge base are reported in table 1. Figure 2 is an overview of WORDNET's noun top level, as translated in our LOOM knowledge base. The nine unique beginners are shown in bold face.

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 the main ontological drawbacks we found after applying this cleaning process.

Confusing Concepts and Individuals The first critical point was the confusion between concepts and individuals. For example, if we look at the hyponyms of the unique beginner event, we 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.[2] Under Territorial_-Dominion, we find Macao and Palestine together with Trust_Territory. The Trust_Territory 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 such as Macao or Palestine. If we go deeper into the

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 Knowledge Base.

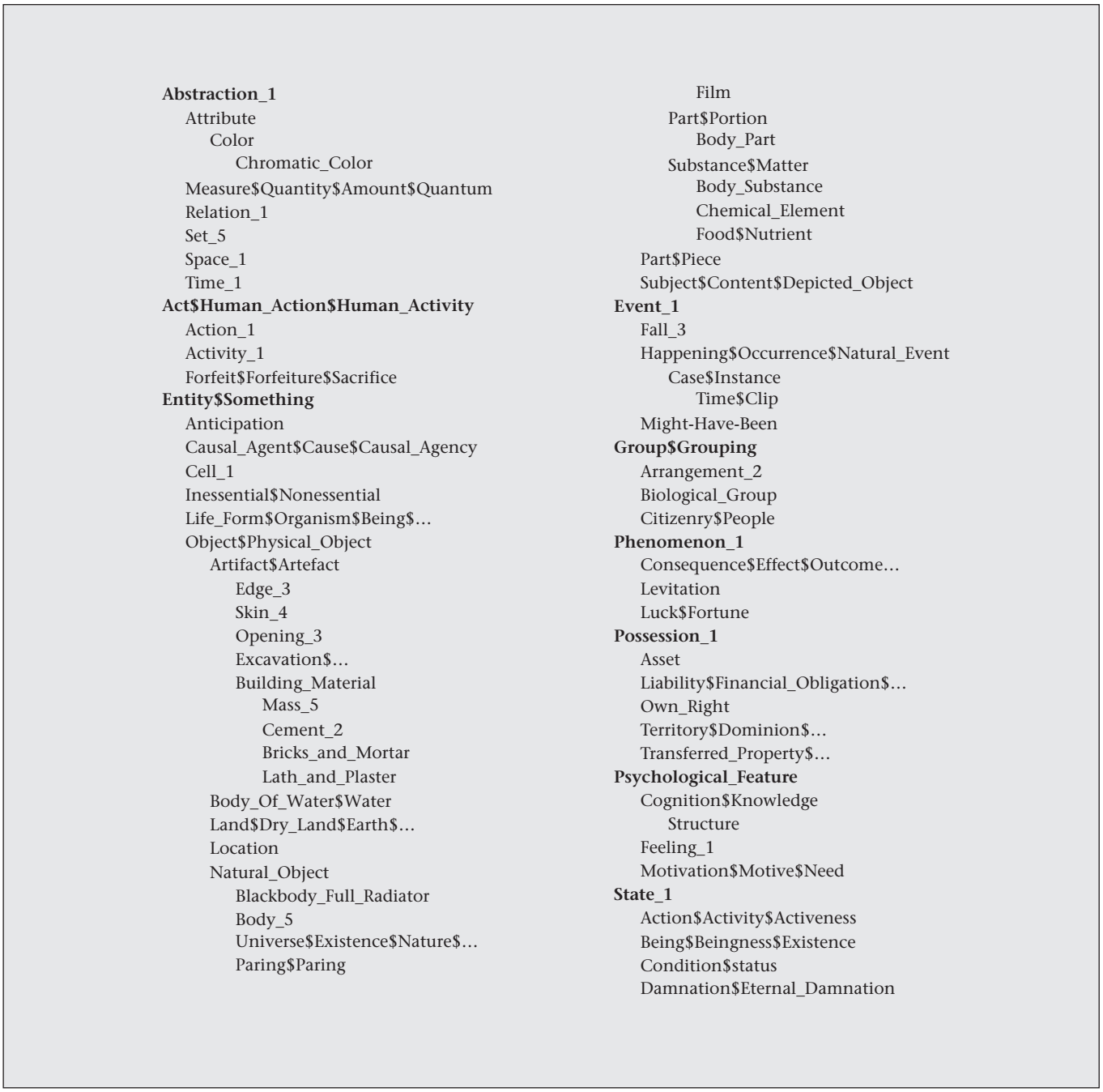

Figure 2. WordNet's Top Level.

taxonomy, we find many other examples of this sort. For example, 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, and Committee together with instances such as Irish_Republican_Army and Red Cross.

We face here a general problem: The concept-individual confusion is nothing but the

product of a lack of expressivity. In fact, if there was an instance-of relation, we could distinguish between a concept-to-concept relation (subsumption) and an individual-to-concept relation (*instantiation*).

Confusing Object Level and Metalevel The synset Abstraction seems to include both object-level concepts, such as Set, Time, and Space, and metalevel 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 with Locke's position (Lowe 1998). This meaning seems to fit the metalevel group of terms (Attribute, Relation, and possibly some hyponyms of Quantity) but not to the object-level group of terms. Moreover, it is quite natural to consider attributes and relations as metalevel concepts, although Set, Time, and Space seem to belong to the object domain.

OntoClean Constraint Violations A core aspect of OntoClean is the analysis of subsumption constraints induced by the identity, rigidity, and unity metaproperties. In our analysis, we only found rigidity violations. We suspect that there are two reasons why we didn't observe other kinds of violations: On the 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 registered is bound to the distinction between *roles* (such as Student) and *types* (such as Person). Roles are antirigid: Every instance of a student can possibly be a nonstudent. Types are rigid: Every instance of a person must be a person. Therefore, roles cannot subsume types. Let's review a clarifying example.

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

Someone could argue that every person is necessarily a causal agent because *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 rigid. However, in this case, it would have only hyponyms denoting things that are (essentially) causal agents, including animals, spiritual beings, and the personified Fate. Unfortunately, such is not 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, and Antifungal. Instances of these concepts are not causal agents essentially, which means that considering Causal_Agent to be rigid would introduce further inconsistencies.

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

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. 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. Subject 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).

For Animal (subsumed by Life_Form), this heterogeneity becomes clearer. Together with concepts such as Chordate, Larva, and Fictional_Animal, we find apparently more specific concepts, such as Work_Animal, Domestic_Animal, Mate, Captive, and Prey. We are induced to consider Chordate and its siblings as types and Work_Animal and its siblings 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 such as Natural_Phenomenon and Process, together with a specific concept such as Consequence, that could be modeled as a role (every event can be the consequence of a previous event, but it seems that this characteristic is not essential to the event itself).

The DOLCE Upper Ontology

The ontology we present here is a first reference module of the WonderWeb library of foundational ontologies. In contrast with *lightweight ontologies,* which focus on a minimal terminological structure (often just a taxonomy) fitting the needs of a specific community, the main purpose of foundational ontologies is to negotiate meaning, either for enabling effec-

tive cooperation among multiple artificial agents or establishing consensus in a mixed society where artificial agents cooperate with human beings. The WONDERWEB vision is to have a library of such ontologies, reflecting different ontological choices. The idea is to make the rationales and alternatives underlying such choices as explicit as possible as the result of a careful isolation of the fundamental ontological options and their formal relationships.

Basic Choices

As reflected by its acronym, DOLCE has a clear cognitive orientation, in the sense that it aims at capturing the ontological categories underlying natural language and human commonsense. Hence, we do not intend DOLCE's categories to account for the intimate nature of the world, but rather, we see them as cognitive artifacts ultimately depending on human perception, cultural imprints, and social conventions. Thus, especially with respect to natural language, our attitude is more descriptive than revisionary (Loux 1998; Strawson 1959).

DOLCE is an ontology of particulars, in the sense that its domain of discourse is restricted to them. The fundamental ontological distinction between *universals* and *particulars* can be understood informally by taking the relation of instantiation as a primitive: Particulars are entities that have no instances;[3] universals are entities that do have instances. Properties and relations (corresponding to predicates in a logical language) are usually considered as universals. Because their domains of discourse are disjoint, we take the ontology of universals as formally separated from that of particulars. Of course, universals do appear in an ontology of particulars, in so far as they are used to organize and characterize them: Simply, because they are not in the domain of discourse, they are not themselves subject to being organized and characterized (for example, by means of metaproperties). An ontology of unary universals is presented in Guarino and Welty (2000).

A basic choice we make in DOLCE is the so-called *multiplicative approach:* Different entities can be colocated in the same space and time. The reason we assume they are different is because we ascribe incompatible essential properties to them. The classical example is that of the vase and the amount of clay: Necessarily, the vase does not survive a radical change in shape or topology, but necessarily, the amount of clay does. Therefore, the two things must be different but colocated: As we see, we say that the vase is constituted by an amount of clay, but it is not an amount of clay.[4] Certain properties a particular amount of clay happened to have when it was shaped by the vase master are considered essential for the emergence of a new entity. In language and cognition, we refer to this new entity as a *genuine different thing*: for example, we say that a vase has a handle but not that a piece of clay has a handle.

Endurant
- Physical endurant
  - Amount of matter
  - Physical object
    - Agentive physical object
    - Non-agentive physical object
  - Feature
  - …
- Non-physical endurant
  - Mental object
  - Social object
  - …
- …

Perdurant
- Eventive
  - Accomplishment
  - Achievement
- Stative
  - State
  - Process

Quality
- Physical quality
- Temporal quality
- Abstract quality

Abstract
- Quality region
- Set
- Proposition
- …

Figure 3. DOLCE's Top Categories.

The Top Categories

The taxonomy of the most basic categories of particulars in DOLCE is depicted in figure 3. They are assumed to be mutually disjoint and to cover the whole domain of particulars. At the metalevel, they are considered as rigid properties, according to the ONTOCLEAN methodology that stresses the importance of focusing on these properties first. In the follow-

ing, we keep the discussion at an informal, introductory level.[5]

Endurants and Perdurants A fundamental distinction we assume is that between enduring and perduring entities. This distinction is almost identical, as we see, to the distinction between so-called continuants and occurrents (Simons 1987), which is still being debated both in the philosophical literature (Varzi 2000) and within ontology standardization initiatives.[6]

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 (2000) that enduring entities can be seen as equivalence classes of perduring entities as the result of some kind of abstraction mechanism.

Classically, the difference between enduring and perduring entities (which we also call endurants and perdurants) is related to their behavior in time. *Endurants* are always wholly present (that is, all their proper parts are present) at any time they are present. *Perdurants,* however, 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 (for example, their previous or future phases) might not be present. For example, the piece of paper you are reading now is wholly present, but some temporal parts of your reading are not present any more. Philosophers say that endurants are entities that are in time but lack temporal parts (put another way, all their parts flow with them in time). Perdurants, however, 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 because none of their parts keeps its identity in time. 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. However, 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.

Another way of characterizing endurants and perdurants has been proposed recently by Katherine Hawley (2001): Something is an endurant iff (1) it exists at more than one moment and (2) statements about what parts it has must be made relative to some time or other. In other words, the distinction is based on the different nature of the parthood relation when applied to the two categories: Endurants need a time-indexed parthood, but perdurants do not. Indeed, the statement "this keyboard is part of my computer" is incomplete unless you specify a particular time, but "my youth is part of my life" does not require such specification.

In DOLCE, the main relation between endurants and perdurants is that of participation: An endurant "lives" in time by participating in some perdurant(s). For example, a person (endurant) can participate in a discussion (perdurant). A person's life is also a perdurant, where a person participates throughout all its duration.

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

Qualities and Quality Spaces Qualities can be seen as the basic entities we can perceive or measure: shapes, colors, sizes, sounds, and smells as well as weights, lengths, electrical charges, and so on. *Quality* is often used as a synonym for *property* but not in DOLCE: Qualities are particulars, but properties are universals. Qualities inhere to entities: Every entity (including qualities themselves) comes with its own exclusive qualities, which exist as long as the entity exists. Thus, we distinguish between a quality (for example, the color of a specific rose) and its value (for example, a particular shade of red). The value is called *quale* and describes the position of an individual quality within a certain conceptual space (called here *quality space*) (Gärdenfors 2000), Thus, 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 they are numerically distinct qualities. Each quality is an instance of a quality type (for example, color, size, and smell), and each quality type is associated to a quality space. Quality types are universals; quality spaces are abstract particulars (see discussion later).

This distinction between qualities and qualia is inspired by Goodman (1951) and the so-called *trope theory* (Campbell 1990). Its intuitive rationale is mainly because natural language—in certain constructs—often seems to make a similar distinction. For example, 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 keeps its identity when some of its properties change.

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

Each quality type has an associated quality space with a specific structure. For example, lengths are usually associated to a metric linear space and colors to a topological two-dimensional 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 this region).

Space and time locations as special qualities: In our ontology, space and time are considered quality types such as color and weight. The spatial (temporal) individual quality of an entity is called *spatial* (temporal) *location*, but 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, the quale is a region in the temporal space. This method allows a homogeneous approach that remains neutral about the properties of the geometric and temporal space adopted (for example, one can assume a circular time). Notice that quality regions can have qualities themselves (for example, the spatial location of a certain object can have a shape).

Direct and indirect qualities: We distinguish in DOLCE between direct and indirect quality inherence. The main reason for this choice comes from the symmetric behavior of perdurants and endurants with respect to their temporal and spatial locations: Perdurants have a well-defined temporal location, but their spatial location seems to come indirectly from the spatial location of their participants; similarly, most endurants (what we call physical endurants; see discussion later) have a clear spatial location, but their temporal location comes indirectly from that of the perdurants they participate in.

Another reason for this distinction concerns complex qualities such as colors, which, according to Gärdenfors (2000), exhibit multiple dimensions (hue, luminosity, and so on). We model this case by assuming that such dimensions are qualities of qualities: Each color quality has a specific hue that directly inheres to it.

Abstract Entities The main characteristic of abstract entities is that they do not have spatial or temporal qualities, and they are not qualities themselves. The only class of abstract entities we consider in the current version of DOLCE is that of quality regions (or simply regions). Quality spaces are special kinds of quality regions, acting as mereological sums of all the regions related to a certain quality type. The other examples of abstract entities reported in figure 3 (sets and propositions) are only indicative.

Further Distinctions

Before discussing further distinctions within the basic DOLCE categories, let us informally introduce some useful definitions based on three notions: (1) mereological invariance (obvious for time-indexed parthood), (2) unity (discussed informally in Guarino and Welty [2002]) and formalized in Gangemi et al. [2001]), and (3) ontological dependence (adapted from Thomasson [1999]).

An endurant is *mereologically constant* iff all its parts remain the same during its life and *mereologically invariant* iff they remain the same across all possible worlds. For example, as we see, amounts of matter are taken as mereologically invariant (all their parts are essential parts).

A particular *x* is an *essential whole* if there is a suitable relation *R* such that necessarily, *x* is a maximal mereological sum of entities that are all bound by *R*. For example, a piece of matter is a *topological whole* whose parts are bound together by a relation of topological connection; a bikini is a functional whole, whose parts are bound together by a functional relationship.

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 his / her 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 constantly be generically dependent on the property of having a heart.

Kinds of Endurant Within endurants, we distinguish between physical and nonphysical endurants, according to whether they have direct spatial qualities. Within physical endurants, the main categories are amount of matter, physical object, and feature.

Amounts of matter: The common trait of amounts of matter—"stuffs" referred to by mass nouns such as *gold, iron, wood, sand,* and *meat*—is that they are endurants with no unity (none of them is an essential whole). They are also mereologically invariant because they change their identity when they change some of their parts.

Physical objects: The main characteristic of physical objects is that they are endurants with unity. However, they have no common unity because different subtypes of objects can have different unity criteria. Different from amounts of matter, (most) physical objects change some of their parts but keep their identity; therefore, they can have temporary parts. Often, objects (indeed, all endurants) are ontologically independent from occurrences (discussed later). However, if we admit that every object has a life, it is hard to exclude a mutually 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.

Features: Typical examples of features are "parasitic entities," such as holes, boundaries, surfaces, or stains, which generically are constantly dependent on physical objects (their hosts). All features are essential wholes, but as in the case of objects, no common unity criterion can exist for all of them. However, typical features have a topological unity because they are singular entities. Some features can be relevant parts of their host, such as a bump or an edge, or places such as a hole in a piece of cheese, the underneath of a table, the front of a house, which are not parts of their host.

It might be interesting to note that we do not consider body parts such as heads or hands as features because 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.

The Agentive-Nonagentive Distinction: Within physical objects, we give special recognition to intentions, beliefs, and desires. These are called *agentive* as opposed to *nonagentive.* Intentionality is understood here as the capability of heading for, or dealing with, objects or states of the world. This area of ontological investigation is important because we haven't properly explored it, so our suggestions are still rather preliminary.

In general, we assume agentive objects as constituted by nonagentive objects: Persons are constituted by organisms, robots are constituted by hardware, and so on. Among nonagentive physical objects, we have, for example, houses, body organs, pieces of wood, and so on.

Nonphysical objects are divided into *social objects* and *mental objects,* according to whether or not they are generically dependent on a community of agents. A private experience is an example of a mental object.

Social objects are further divided into agentive and nonagentive. Examples of agentive social objects are social agents such as "the president of United States." We might think that the president, besides depending generically on a community of U.S. citizens, also depends generically on "George Bush *qua* legal person" (because the president can be substituted), which in turn depends specifically on "George Bush *qua* human being." Social agents are not constituted by agentive physical objects (although they depend on them), but they can constitute societies, such as the Italian National Research Council and Mercedes-Benz. Examples of nonagentive social objects are laws, norms, shares, and peace treaties, which are generically dependent on societies.

Kinds of Perdurant Perdurants (also called occurrences) comprise what are variously called events, processes, phenomena, activities, and states. They can have temporal parts or spatial parts. For example, the first movement of (an execution of) a symphony is a temporal part of it. However, 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 cannot be parts of occurrences but, rather, that they participate in them.

In DOLCE, we distinguish among different kinds of occurrences mainly on the basis of two notions, both extensively discussed in the linguistic and philosophic literature: (1) homeomericity and (2) cumulativity. Homeomericity is discussed, for example, in Casati and Varzi (1996); cumulativity has been introduced in Goodman (1951) and refined in Pelletier (1979).

Intuitively, we say that an occurrence is *homeomeric* if and only if all its temporal parts are described by the very expression used for the whole occurrence. Every temporal part of the occurrence "John sitting here" is still described by "John sitting here." However, if we consider "a walk from Ponte dei Sospiri in Venice to Piazza S. Marco," there are no parts of such an event that constitute a walk from these two places. In linguistic, as well as in philosophical, terminology, the notion of the homeomericity of an occurrence is often introduced with re-

spect to a property characteristic of (or exemplified by) the occurrence itself. If such property holds for all the temporal parts of the occurrence, then the occurrence is homeomeric. In our axiomatization, this presupposes a finite list of occurrence-types (occurrents) that are declared in advance.

An occurrence type is *stative* or *eventive* according to whether it holds the mereological sum of two of its instances; that is, if it is cumulative or not. A sitting occurrence is stative because the sum of two sittings is still a sitting occurrence. Within stative occurrences, we distinguish between *states* and *processes* according to homeomericity: Sitting is classified as a state, but running is classified as a process because there are (very short) temporal parts of a run that are not themselves runs.

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

Kinds of Quality We assume that qualities belong to disjoint quality types according to the kinds of entity they directly inhere to. That is, *temporal qualities* are those that directly inhere to perdurants, *physical qualities* are those that directly inhere to physical endurants, and *abstract qualities* are those that directly inhere to nonphysical perdurants. We are aware that unfortunately, this terminology is very problematic: For example, it should be clear that abstract qualities are not abstracts because they have a temporal location.

Mapping WordNet into DOLCE

Let us consider now the results of integrating the WordNet top concepts into our upper level. According to the OntoClean methodology, we have concentrated first on the so-called backbone taxonomy, which only includes the rigid properties. Formal and material roles have therefore been 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, such as Anticipation, Imaginary_Place, and Inessential. Such synsets have only a few children, which have already been excluded in our analysis.

Some examples of our merging work are sketched in figure 4. Some problems encountered for each category are discussed later.

Endurants

Entity is a very confused synset. A lot of its hyponyms have to be rejected; in fact, there are roles (Causal_Agent, Subject_4), unclear synsets (Location),[7] and so on. This Unique Beginner maps partly to our Amount of Matter and partly to our Physical Object category. Some hyponyms of Physical_Object are mapped to our top concept feature.

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

Possession_1 is a role, and it includes both roles and types. In our opinion, the synsets marked as types (Asset, Liability, and so on) should be moved toward lower levels of the ontology because their meanings seem to deal more with a specific domain—the economic one—than with a set of general concepts. Thus, the remainder branch also has to be eliminated from the top level because of its overall antirigidity (the peculiarity of roles).

Perdurants

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

Qualities and Abstracts

Abstraction_1 is the most heterogeneous Unique Beginner: It contains abstracts such as Set_5; quality regions such as Chromatic_Color; qualities (mostly from the synset Attribute); and a hybrid concept (Relation_1) that contains social objects, concrete entities (as Substance_4),[8] and even metalevel categories. Each child synset has been mapped appropriately.

Psychological_feature contains both mental objects (Cognition)[9] and events (Feeling_1). We consider Motivation a material role; thus, it can be added to lower levels of the taxonomy of mental objects.

The classification of qualities deals mainly with adjectives. This article focused on the WordNet database of nouns; nevertheless, our treatment of qualities also foreshadowed a semantic organization of the database of adjec-

Endurant
- **Physical Endurant**
  - **Amount of matter**
    - body_substance
    - chemical_element
    - mixture
    - compound$chemical_compound
    - mass_5
    - fluid_1
  - **Physical Object**
    - **Agentive Physical Object**
      - life_form$organism$being$…
      - citizenry
      - sainthood
      - ethnic group
    - **Non-agentive Physical Object**
      - body_of_water$water
      - landdry_landearth$…
      - body$organic_structure
      - artifact$artefact
      - biological_group
      - kingdom
      - collection
      - blackbody$full_radiator
      - body_5
      - universe$existence$nature$creation
  - **Feature**
    - edge_3
    - skin_4
    - paring$parings
    - opening_3
    - excavation$hole_in_the_ground
- **Non-physical Endurant**
  - **Mental Object**
    - cognition
    - motivation
  - **Social Object**
    - **Non-agentive Social Object**
      - rule$prescript
        - law
      - circuit_5
    - **Agentive Social Object**
      - social_group

Perdurant
- **Eventive**
  - **Accomplishment**
    - accomplishment$achievement
- **Stative**
  - **State**
    - condition$status
    - cognitive_state
    - existence
    - death_4
    - degree
    - medium_4
    - relationship_1
    - relationship_2
    - conflict
  - **Process**
    - decrement_2
    - increment
    - shaping
    - activity_1
    - chelation
    - execution
    - activity_1

Quality
- **Physical Quality**
  - position$place
  - chromatic_color
- **Temporal Quality**
  - time_interval$interval

Abstract
- **Quality Region**
  - space_1
  - time_1
  - time_interval$interval
  - chromatic_color
- **Set**
  - set_5
- **Proposition**
  - statement_1
- symbol

Figure 4. Mapping WordNet into DOLCE.

tives, which is a current desideratum in the WordNet community.

Conclusion

The final results of our mapping are sketched in figure 4. As one can see, a substantial taxonomy rearrangement was performed. The application of OntoClean's taxonomy evaluation methodology provided a first guideline, but the explicit distinctions of dolce helped clarify the meaning of WordNet senses. We believe that strong (and explicit) ontological distinctions should also help reduce the risk of classification mistakes in the ontology development process and simplify the update and maintenance processes.

WordNet is largely used because of its coverage and has proven to be a key resource in many strategic applications. Whether a "principled"' restructuring, such as the one we have proposed, will have some positive impact on the performance of these applications is still to be fully assessed experimentally. However, a recent refinement of "dolce-restructured WordNet" has successfully been used for a terminology integration project with UN-FAO in the domain of fishery.[10] In addition, some preliminary experiments on the application of this enhanced WordNet to information-retrieval tasks seem encouraging.

Acknowledgments

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

Notes

1. wonderweb.semanticweb.org/.

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

3. More exactly, we should say that they can't have instances, which coincides with saying that they have no instances if—as we do—we include possibilia (possible instances) in our domain of discourse.

4. One of the purposes of OntoClean is indeed to help the user evaluate ontological choices such as this one.

5. The reader can refer to Masolo, C.; Borgo, S.; Gangemi, A.; Guarino, N.; Oltramari, A.; and Schneider, L. 2002. The WonderWeb Library of Foundational Ontologies: Part 1, The dolce Ontology. Commission of the European Communities, project IST-2001-33052, Deliverable D17. wonderweb.semanticweb.org for a formal axiomatization.

6. See, for example, the extensive debate about the three-dimensional (3D) versus the 4D approach at suo.ieee.org.

7. Referring to Location, we find roles (There, Here, Home, Base, Whereabouts), instances (Earth), and geometric concepts (such as Line and Point).

8. "The stuff of which an object consists."

9. "The psychological result of perception and learning and reasoning."

10. www.fao.org/agris/aos.html.

References

Campbell, K. 1990. *Abstract Particulars*. Oxford, United Kingdom: Blackwell.

Casati, R., and Varzi, A., eds. 1996. *Events*. Hanover, N.H.: Dartmouth Publishing.

Fellbaum, C., ed. 1998. *WordNet—An Electronic Lexical Database*. Cambridge, Mass.: MIT Press.

Gangemi, A.; Guarino, N.; and Oltramari, A. 2001. Conceptual Analysis of Lexical Taxonomies: The Case of WordNet Top Level. In Formal Ontology in Information Systems. Proceedings of FOIS2001, eds. C. Welty and S. Barry, 285–296. New York: Association of Computing Machinery.

Gangemi, A.; Guarino, N.; Masolo, C.; and Oltramari, A. 2001. Understanding Top-Level Ontological Distinctions. Paper presented at the Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01) Workshop on Ontologies and Information Sharing, 4–5 August, Seattle, Washington.

Gärdenfors, P. 2000. *Conceptual Spaces: The Geometry of Thought*. Cambridge, Mass.: MIT Press.

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

Guarino, N. 1998. Some Ontological Principles for Designing Upper-Level Lexical Resources. In Proceedings of the First International Conference on Language Resources and Evaluation, 527–534. Granada, Spain: European Language Resources Association.

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

Guarino, N., and Welty, C. 2002b. Identity and Subsumption. In *The Semantics of Relationships: An Interdisciplinary Perspective,* eds. R. Green, C. Bean and S. Myaeng, 111–126. Amsterdam, The Netherlands: Kluwer Academic.

Guarino, N., and Welty, C. 2000. A Formal Ontology of Properties. In *Knowledge Engineering and Knowledge Management: Methods, Models, and Tools, Proceedings of the Twelfth International Conference, EKAW2000,* eds. R. Dieng and O. Corby, 97–112. Berlin, Germany: Springer Verlag.

Hawley, K. 2001. *How Things Persist*. Oxford, United Kingdom: Clarendon.

Loux, M. J. 1998. *Metaphysics, A Contemporary Introduction*. New York: Routledge.

Lowe, E. J. 1998. *The Possibility of Metaphysics*. Oxford, United Kingdom: Clarendon.

MacGregor, R. M. 1991. Using a Description Classifier to Enhance Deductive Inference. In Proceedings of the Seventh Institute of Electrical and Electronics Engineers Conference on AI Applications, 141–147. Washington, D.C.: IEEE Computer Society.

Mourelatos, A. 1996. Events, Processes, States. In *Events,* eds. R. Casati and A. Varzi, 457–476. Hanover, N.H.: Dartmouth Publishing.

Pelletier, F. J. 1979. Nonsingular References: Some Preliminaries. In *Mass Terms: Some Philosophical Problems,* ed. F. J. Pelletier, 1–14. Dordrecht, The Netherlands: Reidel.

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

Simons, P. 1987. *Parts: A Study in Ontology.* Oxford, United Kingdom.: Clarendon.

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

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

Varzi, A., ed. 2000. *The Monist* (Special Issue on Temporal Parts) 83(3).

Woods, W. A., and Schmolze, J. G. 1992. The KL-ONE Family. In *Semantic Networks in Artificial Intelligence,* ed. F. W. Lehmann, 133–177. Oxford, United Kingdom: Pergamon.

Aldo Gangemi graduated in philosophy from the University of Rome "La Sapienza" in 1989 and currently is a research scientist in the Rome unit of the Laboratory for Applied Ontology of the Institute of Cognitive Sciences and Technology of the Italian National Research Council (ISTC-CNR). His interests include knowledge management, the semantic web, terminology, linguistics, and philosophy of sciences. He has been working on ontology-based conceptual modeling since 1992 and has pioneered the application of ontological techniques for terminology reengineering. He is involved in several projects funded by both the European Union and industrial companies.

Nicola Guarino leads the ISTC-CNR Laboratory for Applied Ontology in Trento. He graduated in electrical engineering at the University of Padova in 1978. Since 1991, he has played a leading role in promoting the study of ontological foundations of knowledge engineering and conceptual modeling. He is general chair of the Formal Ontology in Information Systems conference series, is associate editor of the *International Journal of Human-Computer Studies,* and has been guest editor of several journal special issues related to formal ontology and information systems.

Claudio Masolo received his Ph.D. in computer science from the Universities of Padova and Toulouse (cotutorship) in 2000. He is a research scientist at the ISTC-CNR Laboratory for Applied Ontology in Trento. His research interests include ontological foundations of knowledge representation with special emphasis on axiomatic theories of physical objects and qualitative spatiotemporal representation.

Alessandro Oltramari graduated in philosophy from the University of Padova in 2000 and currently has a research position at the ISTC-CNR Laboratory for Applied Ontology in Trento. His interests include philosophy of mind, linguistics, natural language processing, and theory of agents.