A Formal Ontology of Properties

Nicola Guarino and Christopher Welty[†]

LADSEB-CNR
Padova, Italy
{guarino,welty}@ladseb.pd.cnr.it
http://www.ladseb.pd.cnr.it/infor/ontology/ontology.html
† on sabbatical from Vassar College, Poughkeepsie, NY

Abstract

A common problem of ontologies is that their taxonomic structure is often poor and confusing. This is typically exemplified by the unrestrained use of subsumption to accomplish a variety of tasks. In this paper we show how a formal ontology of unary properties can help using the subsumption relation in a disciplined way. This formal ontology is based on some meta-properties built around the fundamental philosophical notions of *identity, unity, rigidity,* and *dependence*. These meta-properties impose some constraints on the subsumption relation that clarify many misconceptions about taxonomies. We discuss these misconceptions by means of real examples, and show how our analysis can bring true order to taxonomies, facilitating their understanding, comparison and integration. This is a first step towards a general methodology for *ontology-driven conceptual analysis* which combines the established tradition of *formal ontology* in Philosophy with the needs of information systems design.

Introduction

Ontologies are becoming increasingly popular in practice, and the number of poor quality ontologies have made clear the need for a principled methodology for building them. Perhaps the most common problem we have seen in practice with ontologies is that, while they are expected to bring order and structure to information, their taxonomic structure is often poor and confusing. This is typically exemplified by the unrestrained use of subsumption to accomplish a variety of reasoning and representation tasks. For example, in previous work (Guarino, 1999) several unclear uses of the *is-a* relation in existing ontologies were identified, such as:

1. a physical object is an amount of matter (Pangloss)
2. an amount of matter is a physical object (WordNet)

This striking dissimilarity poses a difficult integration problem, since the standard approach of generalizing overlapping concepts would not work, and shows that even the most experienced modelers need some guidance for using subsumption consistently.

In this paper we show how a formal ontology of unary properties (corresponding to concepts in taxonomies) can help using the subsumption relation in a disciplined way. This formal ontology is based on some meta-properties built around the fundamental philosophical notions of *rigidity, identity, unity,* and *dependence*. These meta-properties impose some constraints on the subsumption relation that clarify many misconceptions about taxonomies – misconceptions that normally turn taxonomies into a tangled mess. We discuss these misconceptions by means of real examples, and show how our analysis can bring true order to taxonomies, facilitating their understanding, comparison and integration. This is a first step towards a general methodology for *ontology-driven conceptual analysis* (ODCA) which combines the established tradition of *formal ontology* in Philosophy with the needs of information systems design.

The Formal Tools of Ontological Analysis

Our methodology is based on four fundamental ontological notions, which will be discussed in this section: *identity, unity, rigidity,* and *dependence*. We shall represent the behavior of a property with respect to these notions by means of a set of *meta-properties.* Our goal is to show how these meta-properties impose some constraints on the way subsumption is used to model a domain.

Preliminaries

Let's assume we have a first-order language $\mathcal{L}_0$ (the modeling language) whose intended domain is the world to be modeled, and another first order language $\mathcal{L}_1$ (the meta-language) whose constant symbols are the predicates of $\mathcal{L}_0$. Our meta-properties will be represented by predicate symbols of $\mathcal{L}_1$. Primitive meta-properties will correspond to *axiom schemes* of $\mathcal{L}_0$. When a certain axiom scheme holds in $\mathcal{L}_0$ for a certain property, then the corresponding meta-property holds in $\mathcal{L}_1$. This correspondence can be seen as a system of *reflection rules* between $\mathcal{L}_0$ and $\mathcal{L}_1$, which allow us to define a particular meta-property in our meta-language, avoiding a second-order logical definition. Meta-properties will be used as *analysis tools* to characterize the ontological nature of properties in $\mathcal{L}_0$, and will always be defined with respect to a given conceptualization.

We shall denote primitive meta-properties by bolded letters preceded by the sign "+", "-" or "~" corresponding to *carrying* the meta-property, *not carrying* the meta-property, and *anti* the meta-property. The latter will be used to denote special restrictions that are stronger than the simple negation, and will be described in more detail, when relevant, for each meta-property. We use the notation $\varphi^{\mathbf{M}}$ to indicate that the property φ has the meta-property **M**.

We shall furthermore adopt a first order logic with identity. This will be occasionally extended to a simple temporal logic, where all predicates are temporally indexed by means of an extra argument. If the time argument is omitted for a certain predicate φ, then the predicate is assumed to be time invariant, that is $t\varphi(x, t) \rightarrow t\varphi(x, t)$. Note that the identity relation will be assumed as time invariant: if two things are identical, they are identical forever. This means that Leibniz's rule holds with no exceptions.

We also adopt a time-indexed mereological relation P(*x*,*y*,*t*), meaning that *x* is a (proper or improper) part of *y* at time *t*, satisfying the minimal set of axioms and definitions (adapted from (Simons, 1987), p. 362) shown in Table 1.

$\mathrm{PP}(x,y,t) =_{\mathrm{def}} \mathrm{f}(x,y,t) \quad \neg x{=}y$	(proper part)
$\mathrm{O}(x,y,t) =_{\mathrm{def}} \quad z(\mathrm{P}(z,x,t) \quad \mathrm{P}(z,y,t))$	(overlap)
$\mathrm{P}(x,y,t) \rightarrow \mathrm{E}(x,t) \quad \mathrm{E}(y,t)$	(actual existence of parts)
$\mathrm{P}(x,y,t) \quad \mathrm{P}(y,x,t) \rightarrow x{=}y$	(antisymmetry)
$\mathrm{P}(x,y,t) \quad \mathrm{P}(y,z,t) \rightarrow \mathrm{P}(x,z,t)$	(transitivity)
$\mathrm{PP}(x,y,t) \rightarrow \quad z(\mathrm{PP}(z,y,t) \quad \neg \mathrm{O}(z,x,t))$	(weak supplementation)

Table 1. Axiomatization of the part relation.

Our domain of quantification will be that of *possibilia*. That is, the extension of predicates will not be limited to what exists in the actual world, but to what exists in any possible world (Lewis, 1983). For example, a predicate like "Unicorn" will not be empty in our world, although no instance has actual existence there. Actual existence is therefore different from existential quantification ("logical existence"), and will be represented by the temporally indexed predicate E*(x,t)*, meaning that *x* has actual existence at time *t* (Hirst, 1991).

Finally, in order to avoid trivial cases in our meta-property definitions, we shall implicitly assume the property variables as restricted to *discriminating properties* (Guarino, Carrara & Giaretta, 1994), i.e. properties *P* such that $\Diamond \quad xP(x) \quad \Diamond \quad x\neg P(x)$.

The Basic Notions

The notion of identity is at the core of our methodology. Despite its fundamental importance in Philosophy, it has been slow in making its way into the practice of conceptual modeling, although it has been recognized from time to time by various communities. In object-oriented languages, for example, uniquely identifying an object (as a collection of data) is critical, in particular when a system has persistence or distributed components (Wieringa, De Jonge & Spruit, 1994). In databases, *globally unique id's* have been introduced into most commercial systems to address this issue. These solutions approach the notion of identity we use here, but do not account for it completely, as they merely provide a framework for identifying unique *descriptions* and not for understanding the nature of the identity relationship that holds among the entities they describe.

Understanding the distinctions and similarities between *identity* and *unity* appears to be of crucial importance. These notions are different, albeit closely related and often confused under a generic notion of identity. Strictly speaking, identity is related to the problem of distinguishing a specific instance of a certain class from other instances by means of a *characteristic property*, which is unique for *it* (that *whole* instance). Unity, on the other hand, is related to the problem of distinguishing the *parts* of an instance from the rest of the world by means of a *unifying relation* that binds them together (not involving anything else). For example, asking "Is that my dog?" would be a problem of identity, whereas asking "is the collar part of my dog?" would be a problem of unity.

Both notions encounter problems when time is involved. The classical one is that of *identity through change*: in order to account for common sense, we need to admit that an individual may remain *the same* while exhibiting different properties at different times. But which properties can change, and which must not? And how can we reidentify an instance of a certain property after some time? The former issue leads to the notion of *rigidity*, discussed below, while the latter is related to the distinction between *synchronic* and *diachronic* identity. An extensive analysis of these issues in the context of conceptual modeling has been made elsewhere (Guarino & Welty, 2000).

Finally, it is important to note that, while we use examples to clarify the notions central to our analysis, *the examples are not the point of this paper.* The everyday use of these analysis tools ultimately depend on the assumptions resulting from our *conceptualization* of the world (Guarino, 1998). For example, the decision as to whether a cat remains the same cat after it loses its tail, or whether a statue is identical with the marble it is made of, are ultimately the result of our sensory system, our culture, etc. The aim of the present analysis is to clarify the formal tools that can both make such assumptions explicit, and reveal the logical consequences of them. When we say, e.g. that "having the same fingerprint" may be considered an identity criterion for *PERSON*, we do *not* mean to claim this is the universal identity criterion for *PERSON*s, but that *if this were* to be taken as an identity criterion in some conceptualization, what would that mean for the property, for its instances, and its relationships to other properties?

Rigidity

A rigid property has been defined in (Guarino, Carrara & Giaretta, 1994) as a property that necessarily holds for all its instances. For example, we normally think of *PERSON* as rigid; if *x* is an instance of *PERSON*, it must be an instance of *PERSON* in every possible world. The *STUDENT* property, on the other hand, is normally not rigid; we can easily imagine an entity moving in and out of the *STUDENT* property while being the same individual. This notion was later refined in (Guarino, 1998), as shown in Table 2, where the notion of anti-rigidity was added to gain a further restriction. The **~R** meta-property is subsumed by **–R**, but is stronger, as the former constrains all instances of a property and the latter, as the simple negation of **+R**, constrains at least one instance. Anti-rigidity attempts to capture the intuition that all instances of certain properties must possibly not be instances of that property. Consider the property *STUDENT*, for example: in its normal usage, every instance of student is not necessarily so.

Rigid	$\varphi^{+\mathbf{R}}$	φ is a necessary property for *all* its instances
Non-Rigid	$\varphi^{-\mathbf{R}}$	φ is not a necessary property of *all* its instances
Anti-Rigid	$\varphi^{\sim\mathbf{R}}$	φ is an optional property for *all* its instances

Table 2. Rigidity behavior for a property φ.

Rigidity as a meta-property is not "inherited" by subproperties of properties that carry it, e.g. if we have *PERSON*$^{+\mathbf{R}}$ and $\quad xSTUDENT(x) \rightarrow PERSON(x)$ then we know that all instances of *STUDENT* are necessarily instances of *PERSON*, but not *necessarily* (in the modal sense) instances of *STUDENT*, and we furthermore have *STU-*

DENT$^{\sim\mathbf{R}}$. In simpler terms, an instance of *STUDENT* can cease to be a student but may not cease to be a person.

Identity

In the philosophical literature, an *identity condition* (IC) for a arbitrary property φ is usually defined as a suitable relation ρ satisfying the following formula:

$$\varphi(x) \quad \varphi(y) \rightarrow (\rho(x, y) \leftrightarrow x = y) \tag{1}$$

Since identity is an equivalence relation, it follows that ρ restricted to φ must also be an equivalence relation. For example, the property *PERSON* can be seen as carrying an IC if relations like *having-the-same-SSN* or *having-the-same-fingerprints* are assumed to satisfy (1).

As discussed in more detail elsewhere (Guarino & Welty, 2000), the above formulation has some problems, in our opinion. The first problem is related to the need of distinguishing between *supplying* an IC and simply *carrying* an IC: it seems that non-rigid properties like *STUDENT* can only carry their ICs, inheriting those supplied by their subsuming rigid properties like *PERSON*. The intuition behind this is that, since the same person can be a student at different times in different schools, an IC allegedly supplied by *STUDENT* (say, having the same registration number) may be only local, within a certain studenthood experience. It would not supply therefore a "global" condition for identity, satisfying (1) only as a sufficient condition, not as a necessary one.

The second problem regards the nature of the ρ relation: what makes it an IC, and how can we index it with respect to time to account for the difference between *synchronic* and *diachronic* identity?

Finally, deciding whether a property carries an IC may be difficult, since finding a ρ that is both necessary *and* sufficient for identity is often hard, especially for natural kinds and artifacts.

For these reasons, we introduce below a notion of identity conditions that have the following characteristics: i) they can only be supplied by rigid properties; ii) they reformulate the ρ relation above in terms of a formula that explicitly takes two different times into account, allowing the distinction between synchronic (same time) and diachronic (different times) identity; iii) they can be only sufficient or only necessary.

Definition 1 Let φ be a rigid property, and Γ(x,y,t,t') a formula containing x,y,t,t' as the only free variables, such that

$$\neg \quad xytt'(\Gamma(x,y,t,t') \leftrightarrow x{=}y) \tag{2}$$

We say that φ *carries the IC* Γ iff one of the following conditions is verified:

Definition 2 Γ is a *necessary IC carried by* φ when:

$$E(x,t) \quad \varphi(x,t) \quad E(y,t') \quad \varphi(y,t') \quad x{=}y \rightarrow \Gamma(x,y,t,t') \tag{3}$$

$$\neg \quad xy(E(x,t) \quad \varphi(x,t) \quad E(y,t) \quad \varphi(y,t') \rightarrow \Gamma(x,y,t,t')) \tag{4}$$

Definition 3 Γ is a *sufficient IC carried by* φ when:

$$E(x,t) \quad \varphi(x,t) \quad E(y,t') \quad \varphi(y,t') \quad \Gamma(x,y,t,t') \rightarrow x{=}y \tag{5}$$

$$xytt' \, \Gamma(x,y,t,t'). \tag{6}$$

In the formulas above, (2) guarantees that Γ is bound to identity under a certain sortal, and not to arbitrary identity, (4) is needed to guarantee that the last conjunct in (3) is relevant and not tautological, and (6) ensures that Γ is not trivially false.

ICs are "inherited" along a hierarchy of properties, in the sense that, if $\varphi(x) \rightarrow \phi(x)$ and, for example, Γ is a necessary IC for ϕ, then (3) above will hold for φ replacing ϕ.

Definition 4 A non-rigid property *carries* an IC Γ iff it is subsumed by a rigid property carrying Γ.

Definition 5 Any property carrying an IC is marked with the meta-property **+I** (**-I** otherwise).

Definition 6 A property φ *supplies* an IC Γ iff i) it is rigid; ii) it carries Γ; and iii) Γ is not carried by *all* the properties subsuming φ[1]. This means that, if φ inherits different (but compatible) ICs from multiple properties, it still counts as supplying an IC.

Definition 7 Any property supplying an IC is marked with the meta-property **+O** (**-O** otherwise). The letter "O" is a mnemonic for "own identity".

From the above definitions, it is obvious that **+O** implies **+I** and **+R**. For example, both *PERSON* and *STUDENT* do carry identity (they are therefore **+I**), but only the former *supplies* it (**+O**). Supplying an IC is analogous to defining the equality predicate for a class in an object-oriented language, except that ICs can not be "overridden" by a subproperty, merely augmented. See the "Identity constraints" section below for further discussion of this.

Definition 8 Any property carrying an IC (**+I**) is called a *sortal* (Strawson, 1959).

Notice that to recognize that a property is a sortal we are not forced to know *which* IC it carries: as we shall see, distinguishing between sortals and non-sortals is often enough to start bringing order to taxonomies.

Unity

Before addressing what it means for a certain property to *carry* a unity condition (UC), we must first clarify what it means for a certain object to *have* a UC, that is to be a whole.

Definition 9 Let ω be an equivalence relation. At a given time t, an object x is a *contingent whole under* ω if:

$$\neg (P(y, x, t) \quad P(z, x, t) \leftrightarrow \omega(z, y, t)) \tag{7}$$

$$y(P(y,x,t) \rightarrow \quad z(P(z,x,t) \leftrightarrow \omega(z, y,t))) \tag{8}$$

We can read the above definition as follows: *at time t, each part of x must be bound by* ω *to all other parts and to nothing else.* (7) is a non-triviality condition on ω, that avoids considering any mereological sum as a contingent whole. (8) expresses a condition of *maximal self-connectedness* according to a suitable relation of "generalized connection," ω.

[1] Note that this is not a second order quantification, since the properties we are talking of are those fixed within a certain conceptualization.

Depending on the ontological nature of such a generalized connection relation, we may distinguish three main kinds of unity for concrete entities (i.e., those having a spatio-temporal location):

- *Topological unity*: based on some kind of topological connection (a piece of coal, a lump of coal)
- *Morphological unity*: based on shape (a ball, a constellation)
- *Functional unity* (a hammer, a bikini)

As the examples show, nothing prevents a whole from having parts that are themselves wholes (with a different UC). This can be the foundation of a theory of *pluralities*, which is however out of this paper's scope.

We define a stronger notion of whole by assuming that a UC must hold for an object throughout its existence, i.e. by assuming unity as an *essential* property:

Definition 10 Let ω be an equivalence relation. An object x is an *intrinsic whole under* ω if, at any time where x exists, it is a contingent whole under ω.

An important remark is that, if an object is always atomic (i.e., it has no proper parts), then it is an intrinsic whole under the identity relation. We are now in the position to state the following:

Definition 11 A property φ *carries a unity condition* (**+U**) iff there exists an equivalence relation ω such that all its instances are intrinsic wholes under ω.

Notice that the above definition does not imply a second order existential quantification in order to get the relation ω: simply, if such relation is part of our ontology, then $\varphi^{+\mathbf{U}}$.

It is important to make clear that carrying a UC does not imply carrying a necessary IC. This is due to the way Definition 2 is formulated. To see that, suppose that φ carries a UC. We may think that the *persistence* of such condition across time could be a good candidate for a necessary IC for φ, since it satisfies (3). However, it fails to satisfy (4), and does not qualify as a necessary *identity* condition: thus, a UC is a persistence condition, but not an identity condition.

As with rigidity, in some situations it maybe important to distinguish properties that do not carry a *common* UC for all its instances, from properties all of whose instances are not intrinsic wholes. As we shall see, an example of the former kind may be Legal Agent, all of whose instances are intrinsic wholes (some people, some companies), however there is not a single relation ω for all of them (since persons and companies may have different UCs). Amount of Matter is usually an example of the latter kind, since none of its instances can be intrinsic wholes. Therefore we define:

Definition 12 A property has *anti-unity* (**~U**) if every instance of the property is not an intrinsic whole.

Of course, **~U** implies -**U.**

Dependence

The final meta-property we employ as a formal ontological tool is based on the notion of dependence. This is a very general notion, whose various forms and variations are discussed in detail in (Simons, 1987). We shall introduce here a specific kind of dependence, which we call *external notional dependence* (or simply *external dependence*), based on Simons' *notional dependence*. Intuitively, we say that a property φ is *externally dependent* on a property ψ if, for all its instances x, necessarily some instance of ψ must exist, which is not a part nor a constituent of x[1]. For example, *PARENT* is externally dependent on *CHILD* (one can not be a parent without having a child), but *PERSON* is not externally dependent on heart nor on body (because any person has a heart as a part and is constituted of a body).

A formal account of this definition requires a definition of the *constitution* relationship, which in turn is based on non-extensional mereology (see again Simons' book). It will suffice here to say that x constitutes y if x and y *share the same* basic parts, and y is *existentially dependent* on x, that is, x cannot actually exist without y actually existing. For example, a castle and the lump of bricks it is constituted of are formed of the same constituent parts, and the castle cannot exist without the lump of bricks also existing (but not vice-versa). A property which is externally dependent on some other property will be marked with the metaproperty **+D**.

Constraints and Assumptions

Let us now discuss the constraints that follow from our definitions. We distinguish between four kinds of constraints, which are largely overlooked in many practical cases (Guarino, 1999). Concrete examples will be discussed at the end of this paper. In the following, we take φ and ψ to be arbitrary properties.

Rigidity constraints

$$\varphi^{\sim\mathbf{R}} \text{ can't subsume } \psi^{+\mathbf{R}} \qquad (9)$$

This constraint follows immediately from the definitions reported in Table 2. As we shall see, this means that, if $PERSON^{+\mathbf{R}}$ and $AGENT^{\sim\mathbf{R}}$, the latter cannot subsume the former.

Identity constraints

$$\varphi^{+\mathbf{I}} \text{ can't subsume } \psi^{-\mathbf{I}} \qquad (10)$$

$$\text{Properties with incompatible ICs are disjoint.} \qquad (11)$$

The first constraint follows immediately from our definitions, while the second one deserves some comment. An important point is the difference between *different* and *incompatible* ICs, related to the fact that they can be inherited and specialized along taxonomies. Consider the domain of abstract geometrical figures, for example, where the property *POLYGON* subsumes *TRIANGLE.* A necessary and sufficient IC for polygons is, "Having the same edges and the same angles". On the other hand, an *additional* necessary and sufficient IC for triangles is, "Having two edges and their internal angle in common" (note that this condition is only-necessary for polygons). So the two properties have *different* ICs (although they have one IC in common),

[1] We should also exclude the case where this instance of ψ is a *quality* of x (such as its color), and also those where this instance necessarily exists (suppose it is the whole universe, for instance). A full discussion would take to much space.

but their extensions are not disjoint. On the other hand, consider *AMOUNT OF MATTER* and *PERSON*. If we admit mereological extensionality for the former but not for the latter (since persons can replace their parts), they have *incompatible* ICs, so they must be disjoint (in this case, we can't say that a person is an amount of matter).

Unity constraints

$$\varphi^{+U} \text{ can't subsume } \psi^{-U} \tag{12}$$

$$\varphi^{-U} \text{ can't subsume } \psi^{+U} \tag{13}$$

Properties with incompatible UCs are disjoint. (14)

Again, constraints ((12)-(13)) trivially follow from our definitions. As an example of (13), suppose we wonder if *VASE* is subsumed by *AMOUNT OF CLAY*. We may think of a vase as an amount of clay that has "just" the property of being a whole, satisfying a suitable UC for vases. In this case however it would not be an *intrinsic* whole, since it would remain the same amount of clay after the vase is crashed. This analysis of UCs brings to light a very common misuse of the subsumption relation, the fact is that vases are *constituted* of amounts of clay, not subsumed by them.

Dependence constraints

$$\varphi^{+D} \text{ can't subsume } \psi^{-D} \tag{15}$$

This constraint trivially follows from our definitions.

Assumptions

Finally, we make the following assumptions regarding identity, adapted from (Lowe, 1989):

- *Sortal Individuation.* Every domain element must instantiate some property carrying an IC (+**I**). In this way we satisfy Quine's dictum "No entity without identity" (1969).
- *Sortal Expandability.* If two entities (instances of different properties) are the same, they must be instances of a property carrying a condition for their identity.

Property Kinds

We now explore the various combinations of meta-properties discussed in the previous section in order to characterize some basic kinds of properties that usually appear in taxonomies. We shall limit our analysis to three groups of mutually independent meta-properties: identity (**I**,**O**), rigidity (**R**), and dependence (**D**). Unity will not be taken into account for this classfication since we found that its role is more useful for a a "fine-grained" analysis.

A systematic analysis

Analyzing properties based exclusively on the meta-properties discussed in the previous section (excluding **U**) gives us 24 potential categories (**I, O, D** are boolean, **R** has three values). Since **+O**→**+I** and **+O**→**+R** we reduce the number to 14, shown in Table 3, that collapse into the 8 relevant classifications discussed below.

+O	**+I**	**+R**	**+D**	Type	Sortal
			-D		
-O	**+I**	**+R**	**+D**	Quasi-type	
			-D		
-O	**+I**	**~R**	**+D**	Material role	
-O	**+I**	**~R**	**-D**	Phased sortal	
-O	**+I**	**-R**	**+D**	Mixin	
			-D		
-O	**-I**	**+R**	**+D**	Category	Non-sortal
			-D		
-O	**-I**	**~R**	**+D**	Formal Role	
-O	**-I**	**~R**	**-D**	Attribution	
		-R	**+D**		
			-D		
+O	**-I**			incoherent	
		~R			
		-R			

Table 3: Formal ontological property classifications.

The taxonomic structure of these classifications is shown in Figure 1. At the top level (the left), we distinguish between *sortal* and *non-sortal* properties, based on the presence or absence of ICs (the meta-property +**I**). *Roles* group together anti-rigid, dependent properties (**~R+D**), and split into *formal roles* (-**I**) and *material roles* (+**I**). Sortals are divided into *rigid* (**+R**) and *non-rigid* (**-R**), and non-rigid sortals have a further specialization for *anti-rigid* (**~R**). This taxonomy refines and extends the work presented in (Guarino, 1992) and (Guarino, Carrara, & Giaretta 1994).

Types. Types are rigid properties that supply their own identity. Examples include *PERSON*, *CAT*, and *WATER*. Types are the most important properties in an ontology, being the only ones that *supply* identity. As a consequence of the Sortal Individuation assumption, it turns out that every domain element instantiates at least one type property. Types can only be subsumed by categories and strictly non-rigid (i.e. not anti-rigid) formal attributions. We discourage the latter case, but see the discussion of formal attributions above.

Quasi-Types. These are sortals that do not supply identity, but nevertheless are rigid. Examples may be *INVERTEBRATE-ANIMAL*, or *HERBIVORE*. They are always subsumed by at least one type. Often these properties tend to introduce new necessary and sufficient *membership* conditions, which brings up an important point: it is easy to confuse membership and ICs. Membership conditions tell you what instances of a property have in common, ICs tell you when two instances are the same instance, i.e. they define the equality relation. In practice, ontologies have many merely rigid sortals, but careful analysis can often turn them into types (by finding ICs for them), and we believe this is desirable.

Material Roles. The basic notion of "role" is captured by antirigidity and dependence (compare this with (Guarino, 1992)). The intuition is that roles are dependent because they result from an event or relationship. They are anti-rigid to ensure that each entity carrying that property does not

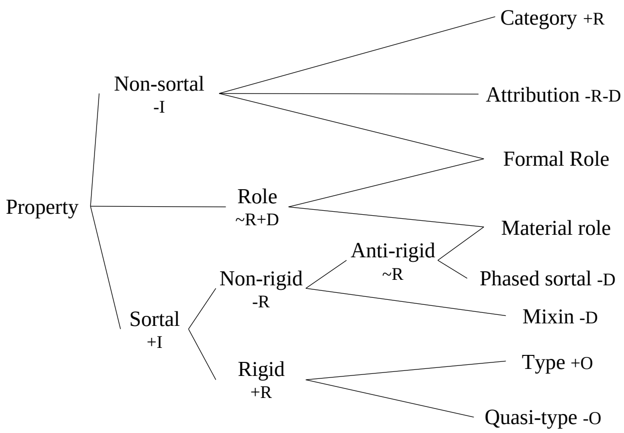

Figure 1: Taxonomy of properties.

necessarily carry it, and therefore roles can not subsume rigid sortals. In addition, *material* roles do carry identity, being subsumed by types. These are the roles that normally specialize formal roles (but not necessarily) and have specific meaning in a domain. Examples include *STUDENT* and *FOOD*.

Phased Sortals. Phased sortals are a unique kind of property that comes from combining a requirement for carrying identity with anti-rigidity and independence. Although they do not supply a *global* IC, they supply a *local* IC, corresponding to a certain temporal phase of their instances (Wiggins, 1980). For example, an individual may at one time be a *CATERPILLAR* and at another time be a *BUTTERFLY*. Some local ICs change across these phases, but it is still the same entity according to the global ICs. According to the Sortal Expandability principle, we have that phased sortals must be subsumed by a type, because it must be possible to determine that they are the same entity at these two times. To this we add two methodological points, that it is useful to group phases of the *same entity* below a single type in the taxonomy, and that outside of biology, phased sortal properties are rare. Properties that may appear to be phased sortals should be targets for even more careful analysis.

Mixins. These are properties that carry identity and are non-rigid (but not anti-rigid). For the most part, properties that fit into these classifications are artificial disjunctions of rigid and non-rigid properties. For example, the property *CAT-OR-WEAPON*, subsuming the type *CAT* and the role *WEAPON*, is non-rigid; some of its instances (instances of *CAT*) are necessarily so, others (instances of *WEAPON* and not *CAT*) are not. Other examples include *MALE-PERSON* and *RED-FLOWER* (assuming that persons can change gender and flowers can change color).While in general we discourage using such properties, in some cases they can be useful for organizing large amounts of similar information in a non-rigid way.

Categories. Categories are properties that are rigid but do not carry identity. Since they can not be subsumed by sortals, categories are normally the highest level rigid properties in an ontology, they are properties that carve the domain into useful segments, and they are often primitive, in the sense that no necessary and sufficient membership conditions can be defined for them. The archetypal category may be *THING*, other examples may be *LOCATION* and *ENTITY*.

Formal Roles. Formal roles, like all roles, are anti-rigid and dependent. Formal roles, since they can not have identity, can not be subsumed by sortals, and therefore are the top-level properties in role taxonomies. Examples include *PART*, *PATIENT*, and *ACTOR*.

Attributions. These are non-sortal properties that are non-rigid. They are normally anti-rigid, and intuitively represents values of *attributes* like *COLOR-OF, GENDER* etc. (hence the name). Examples are *RED or MALE.* The possibility exists that attributions should be always anti-rigid, but we have left it open for now, pending further analysis of cases where types have attributions as necessary properties. One might say that e.g. instances of the type *HAMMER* necessarily have the property *HARD*, whereas other types such as *SPONGE* have the property conditionally (a dry sponge is hard, a wet sponge is soft).

A Taxonomy Cleaning Example

We present now an example of how these meta-properties can be used to *make modeling assumptions clear*, and to produce *well-founded taxonomies.*

Figure 2 shows a messy taxonomy, which has mostly been drawn from examples of overloaded *is-a* relationships in existing ontologies. Our methodology proceeds as follows:

1. *Make clear the ontological assumptions about each property in the taxonomy to in terms of the relevant meta-properties.* To save space, this step is already shown in Figure 2. The assignment of meta-properties was made based on deliberately naive – but believable – assumptions regarding the most common meanings of the terms.

In the next steps, the consistency of these assumptions will be validated on the basis of our meta-properties and their constraints. We give here a necessarily brief account of these assumptions, recalling that the point here is not to claim that they are correct (though we believe them to be reasonable), but to explore the consequences of making these assumptions within a particular ontology.

Locations can be spatial or temporal regions. Since they can be either connected or not, we don't assume a unity

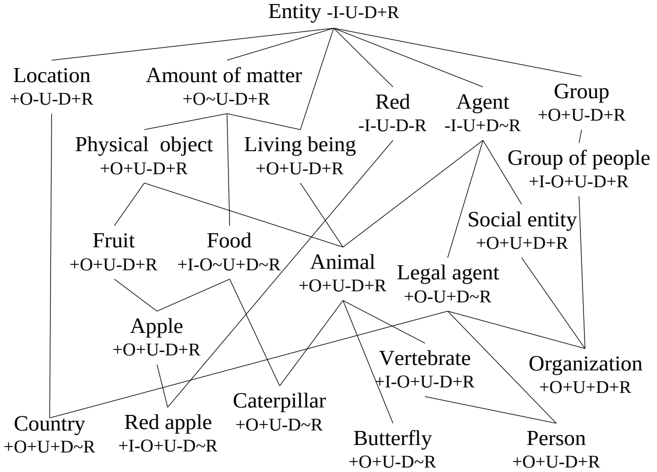

Figure 2: A messy taxonomy.

condition for them. For amounts of matter, we assume that none of them has a unity condition, thus **~U**. Agent is assumed to be anti-rigid to capture the intuition that something is an agent only while it is involved in an action (we are therefore not thinking of *potential* agents). Physical objects are taken to be isolated material entities (so an apple is a physical object, but an undetached part of it is not). Vertebrates are thought of as vertebrate *animals*, not just as arbitrary things having a spine. Social entities are thought of as pluralities of living beings exhibiting some kind of "social unity". Legal agents are conceived as being involved in a legal contract. Finally, countries are (quite naively) conceived as geographic regions that have a (not necessarily permanent) political status. We now continue with the next step of our methodology:

2. *C heck the consistency of each set of meta-properties*. We have seen that our meta-properties are not independent, since "own identity" (**+O**) is only defined for rigid (**+R**) properties. In our case, it seemed natural to assume for *COUNTRY* both **+O** and **~R**, but this is inconsistent with Definition 7**.** The inconsistency forces a closer inspection which reveals that two senses of *country* (political region and geographical region) have been merged into one, while their identity conditions are pretty different. This property is therefore split into two rigid properties, *GEOGRAPHICAL REGION* and *COUNTRY*, carrying their own identity and unity. *COUNTRY* is classified under *SOCIAL ENTITY* and *LEGAL ENTITY* (Figure 3).
3. *Remove all properties except for the rigid ones*. The result of this process, shown in Figure 3, is the first step towards identifying the *backbone taxonomy*, that involves only categories, types, and quasi-types. The Sortal Indviduation assumption guarantees that this backbone covers the entire domain, describing its basic structure.
4. *Check the constraints imposed on each taxonomic relationship* as a consequence of the meta-properties assigned to its arguments. The two links connecting *PHYSICAL OBJECT* and *LIVING BEING* to *AMOUNT OF MATTER* have been deleted because they violate the unity constraints. The link between *ANIMAL* and *PHYSICAL OBJECT* is removed because of incompatible ICs: when an animal dies it ceases to exist, however the physical body remains. *ORGANIZATION*s, similarly, are more than just a group of people, since the same group of people can make different organizations. The result of these operations gives the preliminary *backbone taxonomy* reported in Figure 4.

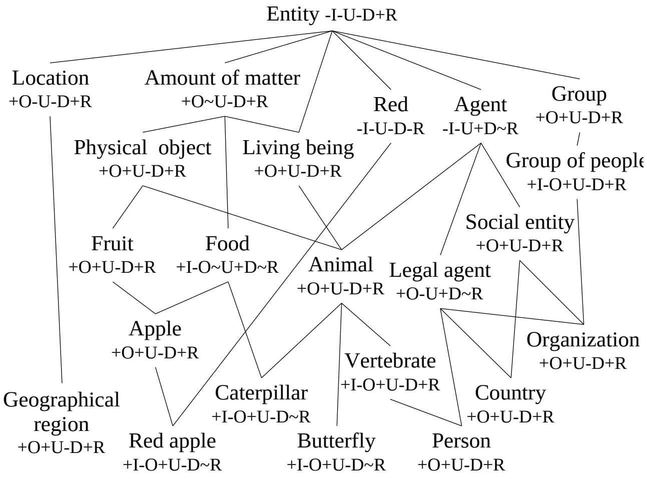

Figure 3: The *COUNTRY* case fixed.

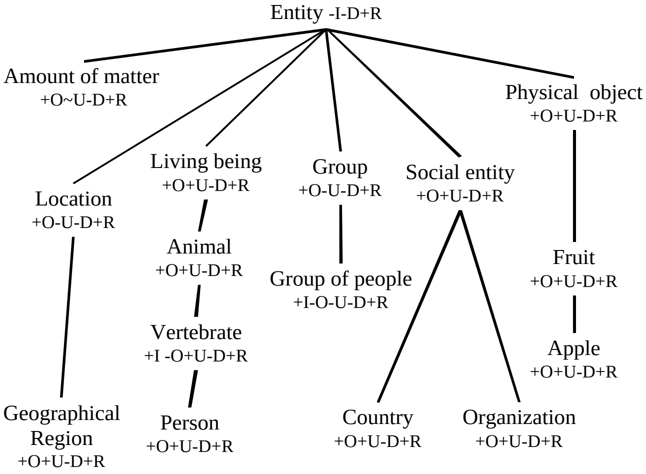

Figure 4: The preliminary backbone taxonomy

5. *Add other properties, checking for possible constraint violations.* In this case, we find that *AGENT* and *LEGAL-AGENT* are not allowed to subsume *PERSON*, *ORGANIZATION*, and *COUNTRY*, since **~R** can not subsume **+R**. It may appear that this results in lost information: previously we had that a *PERSON* is a *LEGAL-AGENT*, where did this go? The answer is that, although *PERSON* is a valid kind of *LEGAL-AGENT*, it *is not the case* that all persons are legal agents. The is-a relation is not the proper way to represent this, a partition of the *LEGAL-AGENT* property, or a specific relation would be more appropriate. The same analysis holds for *FOOD*. The result of this step is shown in Figure 5.
6. *Check for missing concepts.* We have seen that identity incompatibilities imply disjoint sortals. The reverse may be not necessarily true; for instance, in our case, we know that *BUTTERFLY* and *CATERPILLAR* are disjoint from *PERSON* although no identity incompatibility accounts for that. Moreover, we know that caterpillar and butterfly are not disjoint, since the same insect (a lepidopteran) can be a caterpillar at an earlier stage and a butterfly at a later stage. We have therefore good reasons to add a new concept, *LEPIDOPTERAN*, which subsumes both of them. It supplies its own IC, which will be different from those of persons.

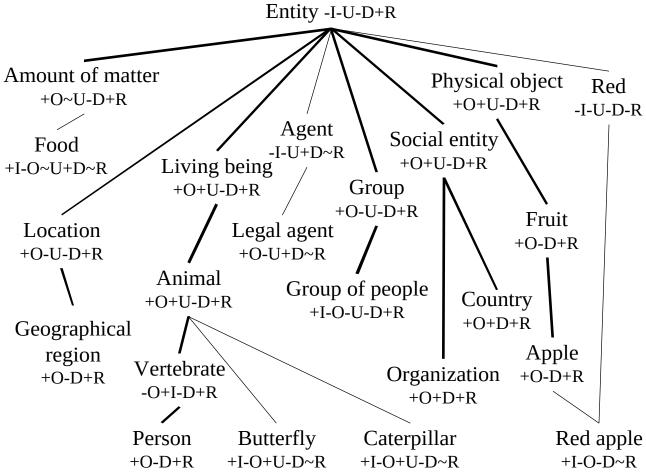

Figure 5: Adding other properties – backbone highlighted.

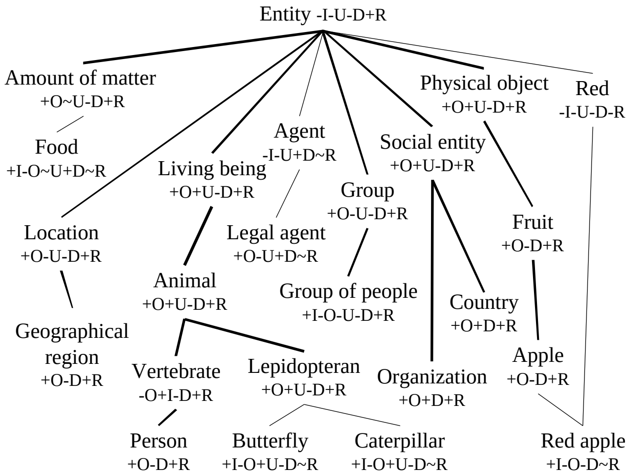

Figure 6: The final taxonomy with highlighted backbone.

The final cleaned taxonomy is shown in Figure 6. Note that one result of this “cleaning” process is the removal of many occurrences of multiple inheritance. This is not necessarily a specific goal, however it naturally follows from the fact that, as discussed in (Guarino, 1999), multiple inheritance is often used as a tool to represent more than simply subsumption – as we found in this example. We believe that these cases make taxonomies confusing; if the purpose of an ontology is to make the meaning clear, then the meaning should not be clouded by using the same mechanism to signify more than one thing, since there is no way to disambiguate the usage. Furthermore, there is at least some empirical evidence derived from studies of programmers who maintain object-oriented programs that multiple inheritance is confusing and makes taxonomies difficult to understand (Huitt and Wilde, 1992).

Conclusions

We have presented here the basic steps of a methodology for ontology design founded on a formal ontology of properties built on a core set of meta properties, which exploits the basic notions of identity, rigidity, and dependence. We have seen how a rigorous analysis based on these notions offers two main advantages to the knowledge engineer:

- It results in a cleaner taxonomy, due to the semantic constraints imposed on the *is-a* relation;
- The backbone taxonomy is identified.
- It forces the analyst to make ontological commitments explicit, clarifying the intended meaning of the concepts used and producing therefore a more reusable ontology.

Acknowledgments

We are indebted to Bill Andersen, Massimiliano Carrara, Pierdaniele Giaretta, Dario Maguolo, Claudio Masolo, Chris Partridge, and Mike Uschold for their useful comments on earlier versions of this paper.

References

Guarino, N. 1992. Concepts, Attributes and Arbitrary Relations: Some Linguistic and Ontological Criteria for Structuring Knowledge Bases. Data & Knowledge Engineering, 8(2): 249-261.

Guarino, N., Carrara, M., and Giaretta, P. 1994. An Ontology of Meta-Level Categories. *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR94).* Morgan Kaufmann.

Guarino, N. 1998. Some Ontological Principles for Designing Upper Level Lexical Resources. *Proceedings of LREC-98.*

Guarino, N. 1999. The Role of Identity Conditions in Ontology Design. In *Proceedings of IJCAI-99 workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends*. Stockholm, Sweden, IJCAI, Inc.: 2-1 2-7.

Guarino, N., and Welty, C. 2000. Identity, Unity, and Individuality: Towards a Formal Toolkit for Ontological Analysis. To appear, *Proceedings of ECAI-2000.* Available from http://www.ladseb.pd.cnr.it/infor/ontology/Papers/OntologyPapers.html.

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

Huitt, R., and Wilde, N. 1992. Maintenance Support for Object-Oriented Programs. *IEEE Transactions on Software Engineering*. **18**(12).

Lewis, D. 1983. New Work for a Theory of Universals. *Australasian Journal of Philosophy,* **61**(4).

Lowe, E. J. 1989. *Kinds of Being. A Study of Individuation, Identity and the Logic of Sortal Terms.* Basil Blackwell, Oxford.

Quine, W. V. O. 1969. *Ontological Relativity and Other Essays*. Columbia University Press, New York, London.

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

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

Wieringa, R., De Jonge, W., and Spruit, P. 1994. Roles and dynamic subclasses: a modal logic approach. In *Proceedings of European Conference on Object-Oriented Programming*. Bologna.