An Ontologically Well-Founded Profile for UML Conceptual Models

Giancarlo Guizzardi[1], Gerd Wagner[2], Nicola Guarino[3], and Marten van Sinderen[1]

[1] Centre for Telematics and Information Technology
University of Twente, Enschede, The Netherlands
guizzard@cs.utwente.nl, sinderen@ctit.utwente.nl
[2] Eindhoven University of Technology
Faculty of Technology Management, Eindhoven, The Netherlands
G.Wagner@tm.tue.nl
[3] Institute for Cognitive Science and Technology
Italian National Research Council
Laboratory for Applied Ontology, Trento, Italy
guarino@loa-cnr.it

Abstract. UML class diagrams can be used as a language for expressing a conceptual model of a domain. In a series of papers [1,2,3] we have been using the General Ontological Language (GOL) and its underlying upper level ontology, proposed in [4,5], to evaluate the ontological correctness of a conceptual UML class model and to develop guidelines for how the constructs of the UML should be used in conceptual modeling. In this paper, we focus on the UML metaconcepts of classes and objects from an ontological point of view. We use a philosophically and psychologically well-founded theory of classifiers to propose a UML profile for Ontology Representation and Conceptual Modeling. Moreover, we propose a design pattern based on this profile to target a recurrent problem in *role* modeling discussed in the literature. Finally, we demonstrate the relevance of the tools proposed by applying them to solve recurrent problems in the practice of conceptual modeling.

1 Introduction

Conceptual modeling is concerned with identifying, analyzing and describing the essential concepts and constraints of a domain with the help of a (diagrammatic) modeling language that is based on a small set of basic meta-concepts (forming a metamodel). Ontological modeling, on the other hand, is concerned with capturing the relevant entities of a domain in an ontology of that domain using an ontology specification language that is based on a small set of basic, domain-independent ontological categories (forming an upper level ontology). While conceptual modeling languages are evaluated on the basis of their successful use in (the early phases of) information systems development, ontology specification languages and their underlying upper level ontologies have to be rooted in principled philosophical theories about what kinds of things exist and what their basic relationships with each other are.

Recently, it has been proposed that UML should be used as an Ontology Representation Language [6]. Moreover, in this paper the authors argue that although UML

A. Persson and J. Stirna (Eds.): CAiSE 2004, LNCS 3084, pp. 112–126, 2004.

lacks a precise definition of its formal semantics, this difficulty shall be overcome with the current developments made by the precise UML community [7]. We believe, however, that defining UML constructs only in terms of its mathematical semantics, although essential, it is not sufficient to make it a suitable ontology representation language. The position defended here is that, in order to model reality, a conceptual modeling language should be founded on formal upper-level ontologies. In other words, it should have both, formal and ontological semantics.

In a series of papers we have been employing the General Ontological Language (GOL) and its underlying upper level ontology, proposed in [4,5], to evaluate the ontological correctness of UML conceptual models and to develop guidelines that assign well-defined ontological semantics to UML modeling constructs. In [1], we have discussed the meaning of the UML metaconcepts of *classes and objects*, *power-types*, *association* and *part-whole* relations (*aggregation/composition*). The UML metaconcepts of *abstract classes* and *datatypes* are addressed in a companion paper [2]. In [3], we have employed some of the results in [1] and [2] to evaluate and improve the conceptual correctness and clarity of UML models in the area of Molecular Biology. The work presented here can be seen as a continuation of this work in which we focus on one aspect of the philosophical problem between universals and particulars (roughly, classes and instances).

Although the Class (entity type, concept) meta-construct is fundamental in conceptual modeling (being present in all major conceptual modeling languages) there is still a deficiency of methodological support for helping the user of the language deciding how to model the elements of a given domain. In practice, a set of primitives are often used to model distinctions in different types of classes (Type, Role, State, Mixin, among others). However, the choice of how the elements that denote universal properties in a domain (viz. Person, Student, Red Thing, Physical Thing, Deceased Person, Customer) should be modeled is often made in ad hoc manner. Likewise, it is the judgment of what are the admissible relations between these modeling meta-constructs. Finally, an inspection of the literature shows that there is still much debate on the meaning of these categories [8,9,10,13].

This paper proposes a philosophically and psychologically well-founded typology of classifiers, which is further used to generate a UML profile of Class types. We also propose a set of methodological guidelines that should govern the use of this profile. Moreover, we demonstrate the relevance of the tools proposed by applying them to solve recurrent problems in the practice of conceptual modeling. In particular, we address a recurrent problem in role modeling presented by Steimann in [10,11,12] and show how the techniques presented here (the profile and a design pattern based on it) account for a proposal which is philosophically better justified but requires no changes to be made in the UML meta-model.

The remaining of this article is structured as follows: Section 2 presents the theory of classifier types and its philosophical and psychological foundations. Section 3 proposes the UML profile for Class types derived from this theory along with examples of how the profile can be used to improve the conceptual quality of conceptual models. Section 4 employs the proposed modeling profile to derive a *design pattern* for the conceptual modeling of *roles*. Finally, section 5 elaborates on some conclusions and future work.

2 Towards a Theory of Classifier Types for Conceptual Modeling: Philosophical and Psychological Foundations

In [14], van Leeuwen shows an important syntactical difference in natural languages that reflects a semantical and ontological one, namely, the difference between common nouns (CNs) on one side and arbitrary general terms (adjectives, verbs, mass nouns, etc…) on the other. CNs have the singular feature that they can combine with determiners and serve as argument for predication in sentences such as:

(i) (exactly) five mice were in the kitchen last night;
(ii) the mouse which has eaten the cheese, has been in turn eaten by the cat.

In other words, if we have the patterns *(exactly) five X…* and *the Y which is Z…*, only the substitution of X,Y,Z by CNs will produce sentences which are grammatical. To see that, we can try the substitution by the adjective *Red* in the sentence (*i): (exactly) five red were in the kitchen last night*. A request to 'count the red in this room' cannot receive a definite answer: Should a red shirt be counted as one or should the shirt, the two sleeves, and two pockets be counted separately so that we have five reds? The problem in this case is not that one would not know how to finish the counting but that one would not know how to start since arbitrarily many subparts of a red thing are still red.

The explanation for this feature unique of CNs lies on the function that determinates (demonstratives and quantifiers) play in noun phrases, which is to determine a certain range on individuals. Both reference and quantification requires that the thing (or things) which are referred or which form the domain of quantification are determinate individuals, i.e., their conditions for *individuation* and *identity* must be determinate. In other words, if it is not determinate how to count Xs or how to identify X that is the same as Y, the sentences in the patterns (i) and (ii) do not express determinate propositions, i.e. propositions with definite truth values.

The distinction between the grammatical categories of CNs and arbitrary general terms can be explained in terms of the ontological categories of Sortal and Characterizing universals [15], which are roughly their ontological counterparts. Whilst the latter supply only a principle of application for the individuals they collect, the former supply both a principle of application and a principle of identity. A principle of application is that in accordance with which we judge whether a general term applies to a particular (e.g. whether something is a Person, a Dog, a Chair or a Student). A principle of identity supports the judgment whether two particulars are the same, i.e., in which circumstances the identity relation holds.

In [16], Macnamara, investigates the role of sortal concepts in cognition and provides a comprehensive theory for explaining the process that a child undergoes when learning proper nouns and common nouns. He proposes the following example: suppose a little boy (Tom), which is about to learn the meaning of a proper name for his puppy. When presented to the word "Spot", Tom has to decide what it refers to. One should notice that a demonstrative such as "that" will not be sufficient to determinate the bearer of the proper name? How to decide that "that" which changes all its perceptual properties is still Spot? In other words, which changes can Spot suffer and still be the same? As Macnamara (among others) shows, answers to these questions are only

possible if Spot is taken to be a proper name for an individual, which is an instance of a Sortal universal. The principles of identity supplied by the Sortals are essential to judge the validity of all identity statements. For example, if for an instance of the sortal *Statue* loosing a piece will not alter the identity of the object, the same does not hold for an instance of *Lump of Clay*.

The statement that we can only make identity and quantification statements in relation to a Sortal amounts to one of the best-supported theories in the philosophy of language, namely, that the identity of an individual can only be traced in connection with a Sortal Universal, which provides a *principle of individuation* and *identity* to the particulars it collects [14,16,17,18]. The position advocated in this article affirms an equivalent stance for a theory of conceptual modeling. We defend that among the conceptual modeling counterparts of general terms (classifiers), only constructs that represent substance sortals can provide a principle of identity and individuation for its instances. As a consequence, the following principle can be postulated:

Postulate 1: Every Object in a conceptual model (CM) of the domain must be an instance of a CM-class representing a sortal.

As argued by Kripke [19], a proper name is a rigid designator, i.e. it refers to the same individual in all possible situations, factual or counterfactual. For instance, it refers to the individual Mick Jagger both now (when he is the lead singer of Rolling Stones and 60 years old) and in the past (when he was the boy Mike Philip living in Kent, England). Moreover, it refers to the same individual in counterfactual situations such as the one in which he decided to continue in the London School of Economics and has never pursued a musical career. We would like to say that the boy Mike Philip is identical with the man Mick Jagger that he latter became. However, as pointed out by Wiggins [20] and Perry [21], statements of identity only make sense if both referents are of the same type. Thus, we could not say that a certain Boy is the same Boy as a certain Man since the latter is not a Boy (and vice-versa). However, as Putnam put it, when a man x points to a boy in a picture and says "I am that boy", the pronoun "I" in question is typed not by Man but by a supertype of Man and Boy (namely, Person) which embraces x's entire existence [22]. A generalization of this idea amount to a thesis, proposed by Wiggins, named thesis D [20]: If an individual falls under two sortals in the course of its history there must be exactly one ultimate sortal of which both sortals are specializations. Griffin elaborates Wiggins' thesis D in terms of two correlated principles:

a) The Restriction Principle: if an individual falls under two distinct sortals F and F' in the course of its history then there is at least one sortal of which F and F' are both specializations.
b) The Uniqueness Principle: if an individual falls under two distinct sortals F and F' in the course of its history then there is only one *ultimate sortal* of which F and F' are both specializations. A sortal F is ultimate if there is no other sortal F' distinct from F which F specializes.

It is not the case that two incompatible principles of identity could apply to the same individual x, otherwise x would not be a viable entity (determinate particular) [14]. Imagine an individual x which is an instance of both Statue and Lump of clay. Now,

the answer to the question whether loosing a piece will alter the identity of x is indeterminate since each of the two principles of identity that x obeys imply a different answer. As a consequence, we can say that if two sortals F and F' intersect (i.e. have common individuals in their extension), the principles of identity contained in them must be equivalent. Moreover, F and F' cannot supply a principle of identity for x, since both sortals apply to x only contingently and a principle of identity must be used to identify x all possible worlds. Therefore, there must be a sortal G that supplies the principle of identity carried by F and F'. This proves the restriction principle. The uniqueness of the ultimate sortal G can be argued as follows: (i) G is a sortal, since it supplies a principle of identity for all the things in its extension; (ii) if it restricts a sortal H then, since H cannot supply an incompatible principle of identity, H either is: equivalent to G (i.e. does supply the same principle of identity) and therefore should be ultimate, or does not supply a principle of identity for the particulars in its extension (see text on dispersive classifiers below). This proves the uniqueness principle. The unique ultimate sortal G that supplies the principle of identity for its instances is named a *substance sortal*.

As a consequence of the *uniqueness principle* we define a second postulate:

Postulate 2: An Object in a conceptual model of the domain cannot instantiate more than one CM-Class representing an ultimate Substance Sortal.

In the example above, the sortal Person is the *unique substance sortal* that defines the validity of the claim that Mick Jagger is the same as Mike Philip or, in other words, that Mike Philip persists through changes in height, weight, age, residence, etc., as the same individual. Person can only be the sortal that supports the proper name Mick Jagger in all possible situations because it applies necessarily to the individual referred by the proper name, i.e. instances of Person cannot cease to be so without ceasing to exist. As a consequence, the extension of a substance sortal is world invariant. This meta-property of classifiers is named *Modal Constancy* [18] or *rigidity* [23] and is formally stated as follows:

Let W be a non-empty set of possible worlds and let $w \in W$ be a specific world. The extension function $ext_w(G)$ maps a classifier G to the set of its instances in world w. Let ext(G) be an extension function mapping to the set of instances of the classifier G that exist in all possible worlds, such that

1. $ext(G) = \cup_{w \in W} ext_w(G)$

and for any classifiers F and G such that F is a specialization of G and, for all $w \in W$

2. $ext_w(F) \subseteq ext_w(G)$

if G is a substance sortal then we have that

3. $ext_w(G) = ext_{w'}(G)$, for any $w,w' \in W$ and consequently,
4. $ext(G) = ext_w(G)$, for all $w \in W$

Sortals such as Boy and Adult Man in the example above, but also Student, Employee, Caterpillar and Butterfly, Philosopher, Writer, Alive and Deceased, which possibly apply to a continuant during a certain phase of its existence, are named phased-sortal in [20]. As a consequence of the Restriction Principle we have that for every phased-sortal PS that applies to a continuant, there is a substance sortal S of which PS is a specialization.

Contrary to substance sortals, phased-sortals apply to individuals contingently and, thus, do not enjoy modal constancy. For example, for an individual John instance of Student, we can easily imagine John moving in an out of the Student type, while being the same individual, i.e. without loosing his identity. Moreover, for every instance x of Student in a world w, there is another world w' in which x is not an instance of Student. This meta-property of classifiers is named anti-rigid in [23]. Formally,

Let PS be a *phased-sortal* and S be a substance sortal restricted by PS. Let

5. $ext_w(\sim PS) = ext_w(S) / ext_w(PS)$

be the complement of the extension of PS in world w. Then for all worlds $w \in W$, there is a $w' \in$ W such that

6. $ext_w(PS) \cap ext_{w'}(\sim PS) \neq \varnothing$

Putting (2), (4) and (6) together derives another postulate:

Postulate 3: A CM-Class representing a rigid classifier cannot be a subclass a CM-Class representing an anti-rigid classifier.

To see that is the case suppose there is a rigid classifier G which specializes an anti-rigid classifier F. Let {a,b,c,d} and {a,b} be the extension of F and G in world w, respectively. By (6), there is a world w' in which $a \in ext_w(F)$ is in $ext_{w'}(\sim F)$ and thus $a \notin ext_{w'}(F)$. By (4), however, $ext_w(G) = ext_{w'}(G)$ and, by (2), $ext_{w'}(G) \subseteq ext_{w'}(F)$, ergo, $a \in ext_{w'}(F)$, which is a contradiction. We therefore conclude that there cannot be the case that a rigid classifier specializes an anti-rigid one.

If PS is a phased-sortal and S is the substance sortal specialized by PS, there is a specialization condition φ such that x is a PS iff x is a S that satisfies φ [14]. A further clarification on the different types of specialization conditions allows us to distinguish between two different types of phased-sortals which are of great importance to the practice of conceptual modeling, namely, *phases* and *roles*.

Phases (also named dynamic subclasses [13] or states [9]) constitute possible stages in the history of a substance sortal. Examples are: (a) Alive and Deceased: as possible stages of a Person; (b) Catterpillar and Butterfly of a Lepidopteran; (c) Town and Metropolis of a City; (d) Boy, Male Teenager and Adult Male of a Male Person. *Classifiers representing phases constitute a partition of the substance sortal they specialize*. For example, if <Alive, Deceased> is a *phase-partition* of a sustance sortal Person then for every world w, every Person x is either and instance of Alive or of Deceased but not of both. Moreover, if x is an instance of Alive in world *w* then there is world *w*' such that x is not an instance of Alive in *w*', which in this case, implies that x is an instance of Deceased in *w*'.

Contrary to phases, roles do not necessarily form a partition of substance sortals. Moreover, they differ from phases in terms of the specialization condition φ. For a phase P, φ represents a condition that depends solely on intrinsic properties of P. For instance, one might say that if Mick Jagger is a Living Person then he is a Person who has the property of being alive or, if Spot is a Puppy then it is a Dog who has the property of being less than a year old. For a role R, conversely, φ depends on extrinsic (relational) properties of R. For example, one might say that if John is a Student then John is a Person who is enrolled in some educational institution or that, if Peter is a Customer then Peter is a Person who buys a Product y from a Supplier z. In other words, an entity plays a role in a certain context, demarcated by its relation with other entities. In general, we can state the following: Let R be role that specializes a sortal S (named the *allowed type* for R [9]) and φ_r be a n-ary relation defined between R and the (n-1) universals on which R is *externally dependent* [23]. For instance, $\varphi_{enrollment} \subseteq$ Student × School, $\varphi_{purchase\text{-}from} \subseteq$ Customer × Supplier or $\varphi_{Marriage} \subseteq$ Husband × Wife. Moreover, let the domain of a relation in world *w* (*Dom*$_w$) be defined as follows: $Dom_w(\varphi_r) = \{x \mid \langle x,y \rangle \in ext_w(\varphi_r)\}$. Then for all worlds $w \in W$ we have that

7. $ext_w(R) \subseteq Dom_w(\varphi_r)$

Although Frege argued at length that "one cannot count without knowing what to count", in artificial logical languages inspired by him, natural language general terms such as CNs, adjectives and verbs are treated uniformly as predicates. For instance, if we want to represent the sentence "there are tall men", in the fregean approach of classical logic we would write $\exists x\ (Man(x) \wedge Tall(x))$. This reading puts the count noun Man (which denotes a Sortal) on an equal logical footing with the predicate Tall. Moreover, in this formula, the variable x is interpreted into a "supposedly" universal kind Thing. So, the natural language reading of the formula should be "there are things which have the property of being a man and the property of being tall". Since, by postulate 1, all individuals must be instances of a substance sortal we must conclude that Thing is a unique universal ultimate sortal which is able to supply a principle of identity for all elements that we consider in our universe of discourse. Moreover, by postulate 2, this principle of identity must be unique. Can that be the case?

In [25], Hirsch argues that concepts such as Thing, (Entity, Element, among others) are *dispersive*, i.e. they cover many concepts with different principles of identity. For instance, in the extension of Thing we might encounter an individuals x which is a cow and an individual y which is a watch. Since the principles of identity for Cows and Watches are not the same we conclude that Thing cannot supply a principle of identity for its instances. Otherwise, x and y would obey incompatible principles of identity and, thus, would not be determinate individuals. Therefore, as defended in [14,17,18,25], dispersive concepts do not denote sortals (despite the fact that they are considered CNs in natural languages) and *therefore cannot have direct instances.* More than that, since a principle of identity supplied by a substance sortal G is inherited by all classifiers that specialize G or, to put in another way, all subtypes of G carry the principle of identity supplied by G. Thus, all subclasses of a sortal are themselves sortals, ergo,

Postulate 4: A CM-Class representing a dispersive universal cannot be a subclass of a CM-Class representing a Sortal.

3 An Ontologically Well-Founded Profile for UML Class Diagrams

The Unified Modeling Language (UML) has built-in extension mechanisms that allow one to modify the language elements to suite certain modeling needs. Extensions to the language can be performed in two different ways: (i) by specializing the UML metamodel (layer 2) to add new semantics to UML modeling elements; (ii) by changing the MOF model (layer 3) to add new elements to the UML metamodel. The former mechanism is named *lightweight extension* and the latter *heavyweight extension*. A coherent set of such extensions, defined accordingly to a specific purpose or domain, constitutes a *UML profile* [26].

In this section we propose *lightweight extension* to UML that represents finer-grained distinctions between different types of classifiers. The proposed *profile* contains a set of *stereotyped* classes (specializations of the meta-construct class) that support the design of ontologically well-founded conceptual models according to the theory proposed in section 2.

It is important to emphasize that the particular classes chosen to exemplify each of the proposed categories are used for illustration purposes only. For example, when stereotyping class *Person* as a *Kind* we are not advocating that Person should be in general considered as a kind in conceptual modeling. Conversely, the intention is to make explicit the consequences of this modeling choice. The choice itself, nonetheless, is always left to the model designer.

3.1 Kinds and Subkinds

A UML class stereotyped as a « kind » represents a *substance sortal* that *supplies* a principle of identity for its instances. Kinds can be specialized in other *rigid* sybtypes that inherit their supplied principle of identity named subkinds. For instance, if we take Person to be a kind then some of its subkinds could be Man and Woman. In general, the stereotype « subkind » can be omitted in conceptual models without loss of clarity.

Every object in a conceptual model using this profile must be an instance of a Kind, directly or indirectly (postulate 1). Moreover, it cannot be an instance of more than one ultimate Kind (postulate 2). Figure 1-a shows an excerpt of a conceptual model that violates the second postulate (extracted from the CYC[1]). Here, we assume that the kinds *Social Being* and *Group* supply different principles of identity. Moreover, it is considered that *Group* supplies an extensional principle of identity, i.e. two groups are the same iff they have the same members. This is generally incompatible with a principle supplied by *Social Being*: we can change the members of a company, football team or music band and still have the same social being. Moreover, the same group can form different social beings with different purposes. One should notice that if “The Beatles” would be an instance of both Kinds, it would not be a determinate object (an answer to the question whether it was still the same thing when Ringo Star replaced Pete Best, is both affirmative and negative!). Figure 1-b shows a version of the model of fig.1-a that obeys the constraints of this profile.

[1] http://www.opencyc.org/

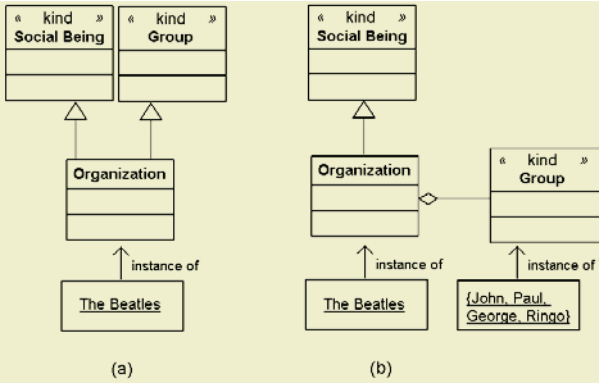

Fig. 1. (a) Example of an instance with conflicting principles of identity; (b) an ontologically correct version of the same model

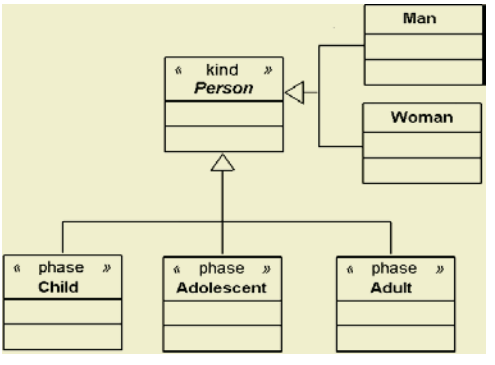

Fig. 2. Two partitions of the same kind: a subkind-partition and a phase-partition

By postulate 3 (sec.2), rigid classes cannot be supertyped by anti-rigid ones. Therefore, kinds cannot appear in conceptual models as subtypes of phases, roles (3.3), and role mixins (3.4).

3.2 Phases

UML classes stereotyped as « phase » represent the phased-sortals *phase*. Figure 2 depicts an example with the kind Person, its subkinds Man and Woman and the phases Child, Adolescent and Adult. The classes connected to one single hollow arrowhead symbol in UML (concrete syntax for the subtyping relation) define a generalization set [27]. A generalization set constitutes a partition of the class pointed by the symbol (superclass). A class with an italicized name is an abstract class, i.e. a class that cannot have direct instances.

3.3 Roles

UML classes stereotyped as « role » represent the phased-sortals *role*. Roles and Phases are anti-rigid universals and cannot appear in a conceptual model as a super-

class of a Kind (postulate 3). Sometimes subtyping is wrongly used in conceptual modeling to represent alternative *allowed types* that can fulfill a role. For instance, in figure 3-a, the intention of the model is to represent that customers are either persons or organizations. Another example is shown in figure 3-b. However, in general being a customer is assumed to be a contingent property of person, i.e. there possible worlds in which a Person is not a customer but still the same person. Likewise, a participant can stop participating in a Forum without ceasing to exist. Figure 3-b contains yet another conceptual problem. In this model, a participant can take part in zero-to-many forums. It is common in Database and Object-Oriented Design to use a minimum cardinality equal to zero to express that in certain state of the system, for example, an object of type Participant is not related to any object of type Forum. However, from a conceptual viewpoint, the involvement in this relation is part of definition of the role type. In this example, the association *participation* is a specialization condition (sec.2), which is part of the content of the concept Participant, i.e. a *Participant* is a *Person* or *SIG* that takes part in a *Forum*. As a consequence of formula (7)-section 2, the following constraint must hold for classes stereotyped as « role »:

Let X be a class stereotyped as « role » and r be an association representing X's restriction condition. Then the minimum cardinality of X.r must be at least 1 (#X.r ≥1)

In section 5, in discussing some related work w.r.t. role modeling, we present a design pattern that can be used to produce ontologically correct versions for the models of figure 3-a and 3-b.

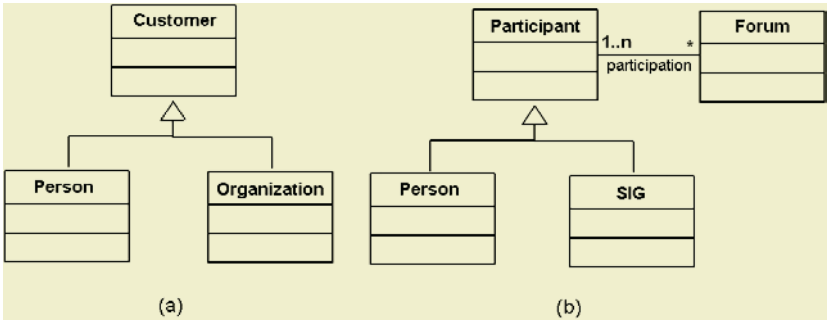

Fig. 3. (a left) and (b) Problems on modeling of roles and their *allowed types*; (b) mistaken cardinality specification for roles

3.4 Mixins

Mixins represent dispersive universals and are perceived to be of great importance in structuring conceptual models [28,29,30]. They can represent top-types such as Thing, Entity, Element (discussed in section 2) but also concepts such as *RationalEntity*, which represent an abstraction of properties that are common to different classes (fig4-a). In this case, the mixin *RationalEntity* can be judged to represent an essential property that is common to all its instances and it is itself a rigid class. We use the stereotype «category» to represent a rigid mixin that subsumes different kinds.

In contrast, some mixins are anti-rigid and represent abstractions of common properties of roles. These classes are stereotyped as «roleMixin» and represent dependent anti-rigid non-sortals. Examples of role mixins include formal roles such as *whole* and *part* and *initiatior* and *responder*. Further examples are discussed in the design pattern proposed in section 5.

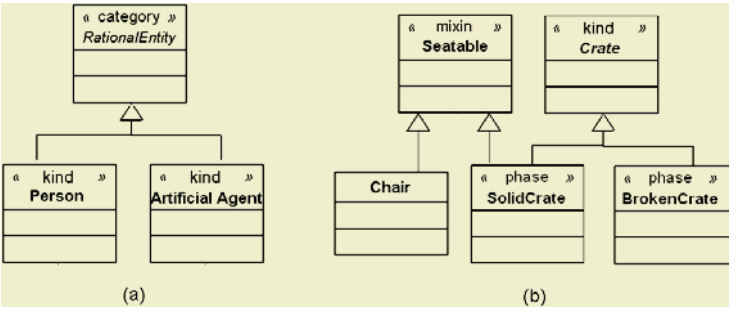

Fig. 4. (a left) Examples of categories and (b) semi-rigid mixins

Moreover, some mixins represent properties which are essential to some of its instances and accidental to others. In [23], this meta-property is named semi-rigidity (as opposed to anti-rigidity). An example is the mixin *Seatable* (fig4-b), which represents a property that can be considered essential to the kinds Chair and Stool but accidental to Crate, Paper Box or Rock. We use the stereotype « mixin » (without further qualification) to represent semi-rigid non-sortals.

Finally, by postulate 4, we have that mixins cannot appear in a conceptual model as subclasses of kinds, phases or roles. Moreover, since they cannot have direct instances, a mixin must always be depicted as an abstract class in a UML conceptual model.

Table 1 below summarizes the profile proposed in this section.

Table 1. Summary of proposed profile for different types of classifiers

Stereotype	Constraints
« kind »	supertype is **not** a member of {« subkind », « phase », « role », « roleMixin »}
« subkind »	supertype is **not** a member of {« phase », « role », « roleMixin »}
« phase »	Always defined as part of partition.
« role »	Let X be a class stereotyped as « role » and r be an association representing X's restriction condition. Then, #X.r ≥ 1
« category »	supertype is **not** a member of {« kind », « subkind », « phase », « role », « roleMixin »}
«roleMixin»	supertype is **not** a member of {« kind », « subkind », « phase », « role »}. Let X be a class stereotyped as « roleMixin » and r be an association representing X's restriction condition. Then, #X.r ≥ 1
« mixin »	supertype is **not** a member of {« kind », « subkind », « phase », « role », « roleMixin »}

4 A Design Pattern for Modeling Roles

In figure 3-a, the role *Customer* is defined as a supertype of *Person* and *Organization*. As previously mentioned, this modeling choice violates postulate 3 and produces an ontologically incorrect conceptual model. Firstly, not all persons are customers, i.e. it is not the case that the extension of *Person* is necessarily included in the extension of *Customer* (formula 2, sec.2). Moreover, an instance of *Person* is not necessarily a *Customer*. Both arguments are also valid for *Organization*.

In a series of papers [10,11,12], Steimann discusses the difficulties in specifying admissible types for Roles that can be filled by instances of disjoint types. As a conclusion, the author claims that the solution to this problem lies in the separation of role and type hierarchies which leads to a radical revision of the UML meta-model (a heavyweight extension).

In the remaining of this section we intend to show that this claim is not warranted. Moreover, we propose a *design pattern* based on the profile introduced in section 3 that can be used as an ontologically correct solution to this recurrent problem. Finally, this solution has a smaller impact to UML than the one proposed by the author, since it does not demand heavyweight extensions to the language.

In the example above, Customer has in its extension individuals that belong to different kinds and, thus, that obey different principles of identity. Customer is hence a dispersive type (a non-sortal) and, by definition, cannot supply a principle of identity for its instances. Since an (determinate) individual must obey one and only one principle of identity, every instance of Customer must be an instance of one of its subtypes (forming a partition) that carry that principle of identity. For example, we can define the sortals PrivateCustomer and CorporateCustomer as subtypes of Customer. These sortals, in turn, carry the (incompatible) principles of identity supplied by the kinds Person and Organization, respectively. In sum, if x is a Customer (abstract class) then x must be an instance of exactly one of its subtypes (e.g., PrivateCustomer) that carries the principle of identity supplied by an appropriate substance sortal (e.g., Person). Figure 5 shows how this solution can be incorporated in a conceptual modeling design pattern. In this picture the abstract class A is the role mixin that covers different role types (e.g., Customer, Participant). Classes B and C are the disjoint subclasses of A that can have direct instances, representing the sortal roles that carry the principles of identity that govern the individuals that fall in their extension. Classes D and E are the ultimate substance sortals (kinds) that supply the principles of identity carried by B and C, respectively. The association r represents the common specialization condition of B and C, which is represented in A. Finally, class F represents a type that A is *externally dependent* on.

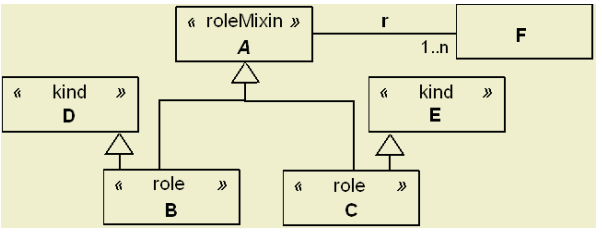

Fig. 5. A design pattern for the problem of specifying roles with multiple disjoint allowed types

An application of this pattern is illustrated in figure 6 in which it is used to produce ontologically correct versions of the models presented in figures 3-a and 3-b, respectively. In both cases, the entity the role mixin depends on, and the association representing the specialization condition are omitted for the sake of brevity.

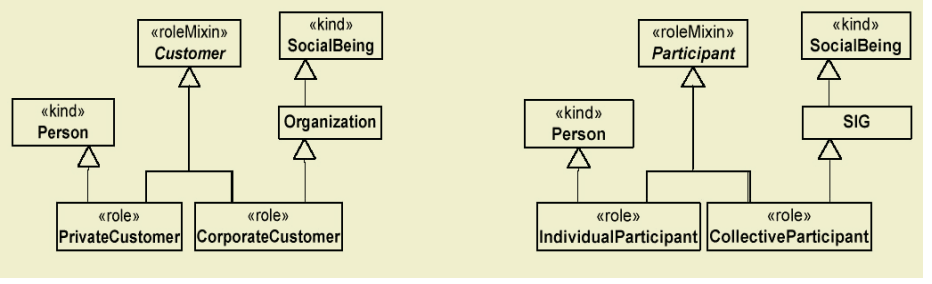

Fig. 6. Ontologically correct versions of the models of fig. 3-a and fig. 3-b obtained by the application of the Design Pattern

5 Final Considerations

The development of a well-grounded, axiomatized upper level ontology is an important step towards the definition of real-world semantics for conceptual modeling diagrammatic languages. In this paper, we use a philosophically and psychologically well-founded theory of universals to address the problem of classifiers in conceptual modeling.

The work presented in section 2 has been strongly influenced by the pioneering work of the OntoClean methodology, which proposes a number of guidelines to evaluate the conceptual correctness of generalization relationships [23,24]. Another key influence is the series of psychological claims proposed by cognitive psychologist John Macnamara in [17]. Mcnamara defends that some universals are conceptually more salient and psychologically more privileged than others and that there is a logic underlying the fact that we can understand certain propositions. A position analogous to the one defended by Chomsky, i.e., that there is a close fit between the mind's linguistic properties and properties of natural languages and, that natural languages have the properties they do because they can be recognized and manipulated by infants without the meta-linguistic support, which is available to second-language learners.

Still in section 2, we have sketched a formalization of the categories proposed by using extension sets indexed by worlds. The idea was to purposely avoid a modal logic approach with unrestricted quantification. In a subsequent article, we shall present the semantics of the proposed categories in a logic of sortals (modal logic with quantification restricted to sortal universals) in the spirit of Gupta's logic of Common Nouns [20] or Montague's systems as presented in [16].

In section 3, this theory is used in the definition of a UML profile for Ontology Representation and Conceptual Modeling. The profile comprises of: (i) a set of stereotypes representing distinctions on types of classifiers proposed by the theory (e.g., Kind, Role, Phase, Category, Mixin); (ii) Constraints on the possible relations to be

established between these elements, representing the postulates of the theory. By using this profile, we were able to propose a design pattern to target a recurrent problem in *role* modeling discussed in the literature. We believe that these results contribute to the task of defining sound *engineering* tools and principles for the practice of conceptual modeling. Nevertheless, the profile should not be regarded as a final proposal. In particular, we recognize that further discussion and elaboration on the issue of role modeling is required, a topic that shall be addressed in a future paper.

Acknowledgements

This work has been supported by the Telematica Instituut in the context of the ArCo project. Part of the work has been developed during an extended visit of the first author to ISTC/LOA (Trento, Italy) and in collaboration with the OntoMed research group (Leipzig, Germany). We would like to thank Heinrich Herre, Claudio Masolo, Laure Vieu, Alessandro Oltramari, Stefano Borgo, Emanuele Botazzi, Roberta Ferrario and Luis Ferreira Pires for fruitful discussions and for providing valuable input to the issues of this article.

References

1. Guizzardi, G., Herre, H., Wagner G.: Towards Ontological Foundations for UML Conceptual Models. In Proc. of 1st Intl. Conf. on Ontologies, Databases and Applications of Semantics (ODBASE 2002), Lecture Notes in Computer Science, Springer-Verlag, Berlin, (2002)
2. Guizzardi, G., Herre, H., Wagner G.: On the General Ontological Foundations of Conceptual Modeling. In Proc. of 21st Intl. Conf. on Conceptual Modeling (ER 2002). Springer-Verlag, Berlin, Lecture Notes in Computer Science (2002)
3. Guizzardi, G., Wagner G.: Using Formal Ontologies to define Real-World Semantics for UML Conceptual Models. In 1st Workshop on Application of Ontologies to Biology, European Media Laboratory, Heidelberg, Germany (2002)
4. Degen, W., Heller B., Herre H., Smith, B.: GOL: Towards an axiomatized upper level ontology. In B. Smith and N. Guarino (eds.), Proc. of 2nd Intl. Conf. of Formal Ontologies and Information Systems (FOIS'01), Ogunquit, Maine, USA, October 2001. ACM Press (2001)
5. Heller, B., Herre, H.: Formal Ontology and Principles of GOL. Onto-Med Report No. 1/2003. Research Group Ontologies in Medicine, Univ. of Leipzig (2003)
6. Cranefield, S., Purvis M.: UML as an ontology modelling language. In Proc. of the Workshop on Intelligent Information Integration, 16th Intl. Joint Conf. on Artificial Intelligence (IJCAI 1999), Germany, Univ. of Karlsruhe (1999) 46-53
7. Evans A., France R., Lano K., Rumpe B.: Developing the UML as a formal modelling notation. In P.-A. Muller and J. Bezivin (eds.), Proc. of UML'98 Intl. Workshop, Mulhouse, France, June 3 -4, 1998, ESSAIM, Mulhouse, France (1998) 297-307
8. Evermann, J. and Wand, Y.: Towards ontologically based semantics for UML constructs. In H. Kunii, S. Jajodia, and A. Solvberg (eds.), Proc. of 20th Intl. Conf. on Conceptual Modeling (ER 2001), Lecture Notes in Computer Science, Springer (2001).
9. Odell, J., Bock, C.: A More Complete Model of Relations and their Implications: Roles. Journal of OO Programming, May, 1998, 51–54

10. Steimann, F.: On the representation of roles in object-oriented and conceptual modeling. Data & Knowledge Engineering 35:1 (2000) 83–106
11. Steimann, F.: Role = Interface: a merger of concepts. Journal of Object-Oriented Programming 14:4 (2001) 23–32
12. Steimann, F.: A radical revision of UML's role concept. In: A Evans, S Kent, B Selic (eds.), UML 2000 Proc. of the 3rd Intl. Conf., Springer (2000) 194–209
13. Wieringa, R.J. de Jonge, W., Spruit, P.A.: Using dynamic classes and role classes to model object migration. Theory and Practice of Object Systems, 1(1) (1995) 61-83
14. Van Leeuwen, J.: Individuals and sortal concepts : an essay in logical descriptive metaphysics, PhD Thesis, Univ. of Amsterdam (1991)
15. Strawson, P. F.: Individuals. An Essay in Descriptive Metaphysics. London and New York: Routledge (1959)
16. McNamara, J. A Border: Dispute, the Place of Logic in Psychology. Cambridge, M.I.T. (1986)
17. McNamara, J.: Logic and Cognition. In McNamara, J.; Reyes, G. (eds.), The Logical Foundations of Cognition, Vancouver Studies in Cognitive Science, Vol. 4 (1994)
18. Gupta, A.: The Logic of Common Nouns: an investigation in quantified modal logic. Yale University Press, New Haven (1980)
19. Kripke, S.: Naming and Necessity. Harvard University Press (1982)
20. Wiggins, D.: Sameness and Substance Renewed. Cambridge University Press (2001).
21. Perry, J.: The same F. Philosophical Review (1970)
22. Putnam H.: Logic and Psychology. In McNamara, J.; Reyes, G. (eds.), The Logical Foundations of Cognition,Vancouver Studies in Cognitive Science, Vol 4 (1994).
23. Welty, C., Guarino, N.: Supporting Ontological Analysis of Taxonomic Relationships. Data and Knowledge Engineering, 39(1), (2001) 51-74
24. Guarino, N.; Welty, C.: Evaluating Ontological Decisions with OntoClean. Communications of the ACM, 45(2) (2002)
25. Hirsch, E.: The Concept of Identity. Oxford University Press, New York, Oxford (1982).
26. Object Management Group: UML 2.0 Infrastructure Specification, Doc.# ptc/03-09-15, Sep. 2003
27. Object Management Group: UML 2.0 Superstructure Specification, Doc.# ptc/03-08-02, Aug. 2003
28. Booch, G., Rumbaugh, J., Jacobson, I.: The Unified Modeling Language User Guide. Addison-Wesley (1998)
29. Booch, G.: Object-Oriented Analysis and Design. Benjamin-Cummings (1994)
30. Hendler, J.: Enhancements for Multiple Inheritance. SIGPLAN Notices, vol.21 (10) (1986) 100-100