Formal Semantics and Ontological Analysis for Understanding Subsetting, Specialization and Redefinition of Associations in UML

Dolors Costal[1], Cristina Gómez[1], Giancarlo Guizzardi[2]

[1]Dept. d'Enginyeria de Serveis i Sistemes d'Informació, Universitat Politècnica de Catalunya (UPC), Barcelona (Catalonia)
{dolors|cristina}@essi.upc.edu
[2]Ontology and Conceptual Modeling Research Group (NEMO), Federal University of Espírito Santo (UFES), Vitória-ES, Brazil
gguizzardi@inf.ufes.br

Abstract. The definition of the exact meaning of conceptual modeling constructs is considered a relevant issue since it contributes to their effective and appropriate use by conceptual modelers. This paper studies three related constructs that enhance the expressiveness of the UML language about associations and which still lack a complete and comprehensive study, namely, association *subsetting*, *specialization* and *redefinition*. It formalizes their semantics, analyses them from an ontological perspective and compares them. The semantic formalization is based on mapping the studied constructs to a basic UML layer which have a previous formal definition in the literature. Furthermore, the ontological analysis developed here is based on a formal theory of relations which is part of the Unified Foundational Ontology (UFO).

Keywords: conceptual modelling, ontological analysis, UML associations.

1 Introduction

During the last decade, UML has been widely adopted both in industry and academia, thus contributing to the improvement of information systems engineering practices. However, one drawback of UML that has been frequently pointed out is the lack of clear meaning for some of its constructs. While the UML metamodel [1] gives information about its abstract syntax, its meaning is described in natural language. Thus, many concepts still lack definitions precise enough to be interpreted unambiguously. The version 2.0 of UML has made a significant step towards precise definitions of concepts. But its attempt to increase the expressiveness of the language has introduced new ambiguities and there are still issues that remain open.

Associations (also termed *relationship types* or simply *relations*) are central structural elements in conceptual modelling, in general, and in UML, in particular. UML 2 has improved the expressiveness of the language with respect to associations in several manners. A significant one has been the introduction of the association *redefinition* concept. This concept allows enhancing the definition of an association by means of another association that defines it more specifically in a particular context. Association subsetting and association specialization have been included in UML since its earliest versions and share some relevant features with association redefinition. These similarities among the three constructs make it frequently difficult, especially to novice users, to: decide which one of these concepts is the best suited to model a particular situation; systematically justify their modelling choices. This situation is worsened by the fact that, despite its importance, the association construct is often regarded as one of the most problematical in conceptual modelling [2].

It seems also to be generally accepted that redefinition, specialization and subsetting of associations still need to be studied in more detail [3,4]. For instance, [3] notices that *"the distinction between subsetting and specializing an association is not clearly described in the UML2 specification"* and [4] mentions that *"it is not completely clear what else subtyping for associations should really mean"*.

The contributions of this paper are two-fold. Firstly, we provide a precise and complete description of the meaning of subsetting, specialization and redefinition of binary associations in UML, making explicit the similarities and differences that exist among these constructs. The semantic formalization is based on mapping them to a basic UML layer which has a previous formal definition in the literature. The proposed formalization provides an interpretation of the constructs under consideration which is in accordance with the OMG document statements about them [1]. Secondly, as we will demonstrate, the formal characterization of these constructs by itself is insufficient to completely differentiate them. Thus, as a second contribution, we develop here an ontological analysis of these constructs by employing an ontologically well-founded formal theory of relations. By mapping these concepts to this ontological theory we are able to: (i) provide unambiguous real-world semantics to them; (ii) provide some methodological guidelines for helping conceptual modelers to systematically choose how elements in the universe of discourse should be modeled using these concepts; (iii) explain specific characteristics of each of these concepts which are manifested in their syntactical constraints and formal semantics.

The remainder of this paper is organized as follows. The next section gives a brief introduction to these constructs in UML. Section 3 provides the formal semantics of the three constructs. Section 4 conducts the ontological analysis. Section 5 reviews related work and finally, Section 6 presents our final considerations.

2 Background

This section briefly describes the notion and syntax of subsetting, specialization and redefinition of associations according to the UML standard document [1].

Subsetting is a construct specified for association ends. It defines that the instances related by the subsetting end are a subset of the instances related by the subsetted end, taking the same departing instance at the opposite ends. One or both ends of a binary association can be subsetted.

An association specialization is a taxonomic relationship between a more specific and a more general association. The specific association inherits the features of the general one. In contrast to subsetting, specialization is a construct specified for associations themselves and not for association ends.

The purpose of association redefinition is to define an association end more specifically in a particular context. One or both ends of an association can be redefined and redefinitions always involve two associations.

The syntax and the syntactic rules, according to UML, of the three constructs are depicted in Figure 1. Note that, in case of redefinition, class A_1 cannot be the same as class A in contrast to the other constructs.

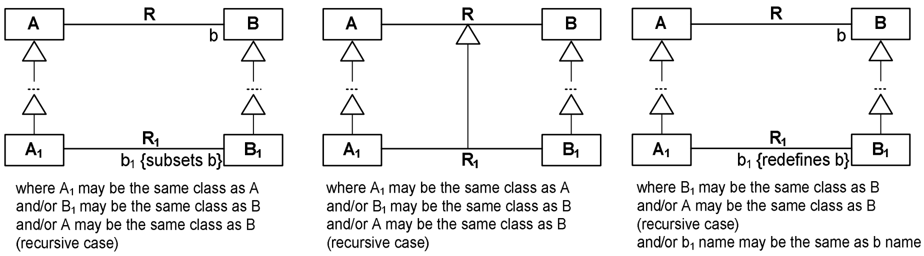

Fig.1.a-b-c Association subsetting, specialization and redefinition syntax and syntactic rules

3 Formal Semantics

The semantics of a language defines the meaning of each construct of the language. A semantic definition consists of two steps: first, a semantic domain must be defined and then, the semantic definition of a construct is done by mapping the syntax of the construct to concepts of the semantic domain. In general, several notations or languages may be used as semantic domains. Formal languages like mathematical terms (as done in [5]), mathematical structures (as done in [6]) or, even, a subset of the UML itself (as done in [7]) may be used among others.

We use a basic UML layer as a semantic domain. It includes the following basic elements: classes, binary associations, multiplicities, class specializations and general OCL constraints [1,8]. Since its elements are generally well-understood, we think that using this layer contributes to the goal of making the studied constructs understandable for any UML user. The concrete and abstract syntaxes of these basic elements are completely developed in [1,3,8] and their semantics is defined using set theory as a basis [5,8,9]. The mapping of a construct to the UML layer is defined as a translation between a generic schema with the specified construct and another generic schema using only elements of the basic UML layer [7].

3.1 UML Association Subsetting Semantics

The translation of a subsetting into our basic UML layer consists in replacing it by an inclusion constraint between the subsetting and the subsetted ends expressed in OCL.

Definition 1: *Subsetting translation definition*: Consider the schema of Figure 1.a. The translated schema is obtained from the original one by: 1) eliminating from it the adornment which denotes that the end b_1 subsets the end b, and, 2) attaching to it the OCL expression: `context A1 inv: self.b ->includesAll(self.b1).`

Since the instances of an association can be understood as bidirectional links from a conceptual modelling point of view, it is semantically identical to specify a subset between two association ends or between their respective opposite ends [10-12]. This indicates that, although subsetting is specified for association ends, it affects the whole association. The semantics of a subsetting is to establish an inclusion constraint between the subsetting association (that with the subsetting end) and the subsetted association (that with the subsetted end). This constraint implies that each instance of the subsetting association must be an instance of the subsetted one.

Consider the example of Figure 2.a. The end *es* subsets the end *s*. Part (b) gives its translation. The OCL invariant guarantees that the set of instances of *Enrols* is a subset of the set of instances of *HasPreference*.

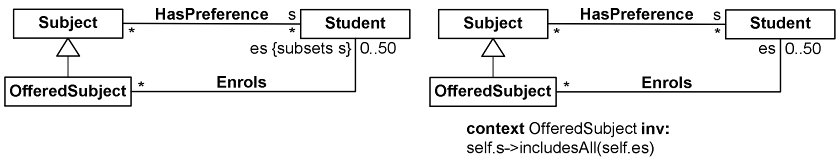

Fig.2.a-b Example of an association subsetting and its translation

3.2 UML Association Specialization Semantics

The semantics of an association specialization is to establish an inclusion constraint between the specific and the general association. Thus, its semantics is identical to the semantics of subsettings.

Definition 2: *Specialization translation definition*: Consider the schema of Figure 1.b. The translated schema is obtained from the original one by: 1) eliminating from it the specialization symbol that relates R and R_1, and, 2) attaching to it the OCL expression:

```
context A1 inv: self.b ->includesAll(self.b1)
```

In Figure 3.a, *PronouncesSentence*, that relates a court and a defendant, is specialized by *Absolves*. The pronouncements of sentences in which the defendant is found not guilty by the court correspond to instances of *Absolves*. The translation (Figure 3.b) ensures that, for all absolutions, there exists the corresponding sentence pronouncement.

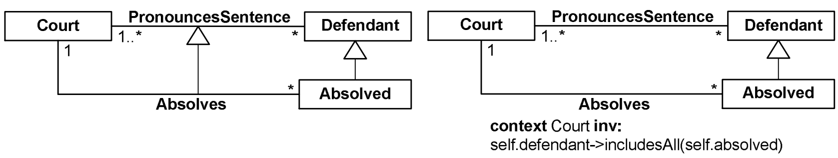

Fig.3.a-b Example of an association specialization and its translation

3.3 UML Association Redefintion Semantics

The semantics of an association redefinition is to establish a constraint that guarantees that the links involving instances of A_1 (see Figure 1.c) coincide in both associations.

Those links are redefined since they have to encompass not only the specification of the redefined association but also the specification of the redefining association.

Definition 3. *Redefinition translation definition***:** Consider the schema of Figure 1.c. The translated schema is obtained from the original one by: 1) eliminating from it the adornment which denotes that the end b_1 redefines the end b, and, 2) attaching to it the OCL expression:

```
context A1 inv: self.b = self.b1
```
in case b name $\neq$ b_1 name

```
context A1 inv: self.oclAsType(A).b = self.b1
```
in case b name = b_1 name

Figure 4.a shows two examples of association redefinitions. The end *participant* is redefined in the context of critical projects and the end *proj* is redefined in the context of junior employees. Their translation is depicted in Figure 4.b. The second invariant translates the *senPart* redefinition and guarantees that the instances that relate critical projects to employees via *Participates* are the same as those that relate them to senior employees through *ParticipatesCritical*.

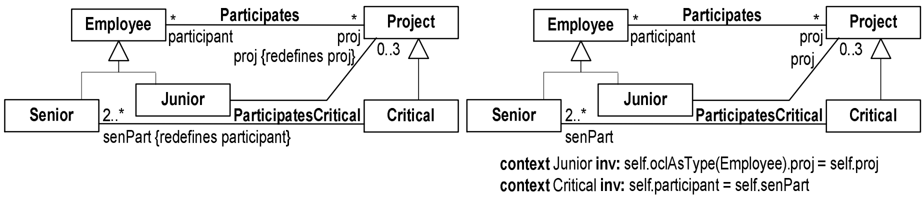

Fig.4.a-b Examples of association redefinitions and their translations

Although the semantics of a redefinition is different from that of specializations and subsettings, an inclusion constraint is also induced. Indeed, from the translation we can see that all the instances of the redefining association belong to the redefined one. This can be easily proven from the definition 3. In Figure 4, for example, all instances of the *ParticipatesCritical* association are also instances of *Participates*.

An association redefinition has implicit significant semantic effects on the redefined association. These effects are circumscribed to the instances of the redefined association that are also instances of the redefining association. We call them affected instances of the redefinition. In Figure 4, the affected instances of the *senPart* redefinition are the instances of *Participates* that involve critical projects. In the following, we describe all kinds of constraints that may be induced on a redefined association depending on the features of the redefinition.

Type is redefined when the redefining end is connected to a descendant of the class at the redefined end. Considering the Figure 1.c syntax, this occurs when class B_1 is not the same as class B. The effect of a type redefinition is to establish an additional participation constraint on the redefined association [13]. In Figure 4, the *senPart* redefinition establishes that critical projects may only have senior employees as participants and, consequently, not junior employees. This induced effect can be proven from the translation of the association redefinition and the referential integrity constraint that is implicit in the redefining association.

Multiplicity is redefined when the redefining end has a multiplicity which is more restrictive than that of the redefined end. The effect is to establish additional minimum and/or maximum cardinality constraints that restrict the cardinality for the affected instances of the redefinition [13]. In Figure 4, the *senPart* redefinition establishes that critical projects must have at least two participants. This induced effect can be proven from the translation of the association redefinition and the cardinality constraints specified by the multiplicity at the redefining end.

On the other hand, association subsettings/specializations induce minimum cardinality constraints over the subsetted/general associations when the lower bound of a multiplicity specified for the subsetting/specific association is greater than the corresponding lower bound at the subsetted/general association. The reason is the existence of the inclusion constraint between both associations due to the subsetting/specialization. However, it is never the case that maximum cardinality constraints or participation constraints are induced by a subsetting or a specialization. In Figure 2, although the offered subjects may have at most 50 enrolled students, they may have more than 50 students with a preference for them.

3.4 Comparing Subsetting, Specialization and Redefinition of Associations

Table 1 summarizes the syntactic and semantic relevant features that we have identified. We can observe that: 1) the three constructs have differing syntactic features, 2) the three constructs imply the existence of an inclusion constraint between the involved associations and 3) subsetting and specialization induce the same constraints on the subsetted/specialized association (i.e. minimum cardinality) while redefinition may induce participation and maximum cardinality constraints in addition.

Therefore, subsettings and specializations cannot be distinguished from a formal semantic point of view since they are semantically equivalent. Redefinitions can be semantically distinguished from the other two constructs when they induce participation and/or a maximum cardinality constraints but not when they only induce minimum cardinality constraints.

Table 1. Syntactic and semantic features for assoc. subsetting, specialization and redefinition

		Subsetting	**Specialization**	**Redefinition**
Syntactic	**Specified for**	Association end	Association	Association end
	Rules for association end classes	Descendant or same classes	Descendant or same classes	End opp. redefining end: desc. class Redefining end: desc. or same class
Sem.	**Inclusion constr.**	Yes	Yes	Yes
	Induced constr.	Min. card.	Min. card.	Participation Min. & max. card.

We must conclude that the formal semantics analysis contributes to capture relevant aspects of the meaning of each construct but it is not sufficient to completely characterize each one of them. To provide the conceptual modeller the criteria to decide which construct is the most adequate to model a particular domain situation we analyse, in the following section, the constructs from an ontological point of view.

4 An Ontological Analysis of Relations

In [14], we have presented an in depth analysis of domain relations from an ontological point of view. In particular, we have employed the Unified Foundational Ontology (UFO), a formal framework which has been constructed by considering a number of theories from formal ontology in philosophy, but also cognitive science, linguistics and philosophical logics. In a number of papers, UFO has been successfully employed to analyze and provide real-world semantics for conceptual modeling grammars and specifications. Here, we make a very brief presentation of this foundational ontology and concentrate only on the categories which are germane to the purposes of this article. For an in depth discussion on UFO, one should see [15].

A fundamental distinction in this ontology is between the categories of *Objects* and *Tropes*. *Objects* are existentially independent entities. Examples include ordinary objects of everyday experience such as an individual person, an organization, an organ. The word *Trope,* in contrast, denotes, what is sometimes named an individualized (objectified) property or property in particular. A trope is an individual that can only exist in other individuals (in the way in which, for example, electrical charge can exist only in some conductor, or that a covalent bond can only exist if those connecting atoms exist). Typical examples of tropes are a colour, a connection, an electric charge, a symptom, a covalent bond. As discussed in [15], there is solid evidence for tropes in the literature. On one hand, in the analysis of the content of perception, tropes are the immediate objects of everyday perception. On the other hand, the idea of tropes as *truthmakers* underlies a standard event-based approach to natural language semantics. Existential dependence can also be used to differentiate intrinsic and relational tropes: intrinsic tropes or *qualities* are dependent on one single individual (e.g., a symptom, a skill, a belief); relational tropes or *relators* depend on a plurality of individuals. In this paper, we focus on relators.

Another important distinction in the UFO ontology is within the categories of relations. Following the philosophical literature, it recognizes two broad categories of relations, namely, *material* and *formal* relations. *Formal relations* hold between two or more entities directly, without any further intervening individual. In principle, it includes those relations that form the mathematical superstructure of our framework. Examples include historical and existential dependence (*ed*), subtype-of, part-of, subset-of, instantiation, among many others [15]. Domain relations such as *working at*, *being enrolled at*, and *being the husband of* are of a completely different nature. These relations, exemplifying the category of *Material relations*, have material structure of their own. Whilst a formal relation such as the one between Paul and his headache x holds directly and as soon as Paul and x exist, for a material relation of *being treated in* between Paul and the medical unit MU_1 to exist, another entity must exist which *mediates* Paul and MU_1. These entities are termed *relators*.

Relators are individuals with the power of connecting entities. For example, a medical treatment connects a patient with a medical unit; an enrollment connects a student with an educational institution; a covalent bond connects two atoms. The notion of relator is supported by several works in the philosophical literature [16] and, they play an important role in answering questions of the sort: what does it mean to say that John is married to Mary? Why is it true to say that Bill works for Company X but not for Company Y? Again, relators are special types of tropes which, therefore, are existential dependent entities. The relation of *mediation* (holding between a relator *r* and the entities that *r* connects) is a sort of existential dependence relation [15].

An important notion for the characterization of relators (and, hence, for the characterization of material relations) is the notion of *foundation*. Foundation can be seen as a type of *historical dependence* [15], in the way that, for instance, an instance of *being kissed* is founded on an individual *kiss,* or an instance of *being connected to* between airports is founded on a particular flight connection. Suppose that John *is married to* Mary. In this case, we can assume that there is an individual relator m_1 of type *marriage* that mediates John and Mary. The foundation of this relator can be, for instance, a wedding event or the signing of a social contract between the involved parties. In other words, for instance, a certain event e_1 in which John and Mary participate can create an individual marriage m_1 which existentially depends on John and Mary and which mediates them. The event e_1 is the foundation of relator m_1.

The relator m_1 in this case is said to be the *truthmaker* of propositions such as “John is the husband of Mary”, and “Mary is the wife of John”. In other words, material relations such as *being married to*, *being legally bound to*, *being the husband of* can be said to hold for the individuals John and Mary because and only because there is an individual relator marriage m_1 mediating the two. Thus, as demonstrated in [14], material relations are purely linguistic/logical constructions which are founded on and can be completely derived from the existence of relators. In fact, in [14], we have defined a formal relation of derivation (symbolized as *der*) between a relator type (e.g., Marriage) and each material relation which is derived from it.

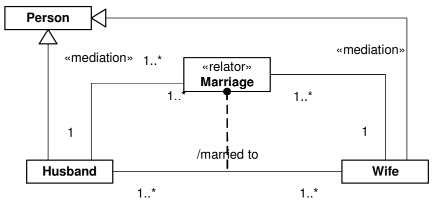

Fig.5. Model with an explicit representation of: a Relator Type and a Material Relation derived from it.

Figure 5 above depicts a model expressed in the OntoUML language which summarizes our discussion on relators and material relations. OntoUML is a well-founded version of UML whose modelling primitives (stereotypes in this Figure) are derived from the ontological categories underlying UFO [14]. This model captures that the material relation *married to* is derived from the relator type Marriage (the derivation relation *der* is symbolized as ●- - - - in the language). As a consequence, we have that a tuple such as <John,Mary> is an instance of this relation iff there is an instance of Marriage m_1 that mediates the elements of the tuple.

The explicit representation of the relator type in OntoUML solves a number of problems associated with the traditional representation of relations in Conceptual Modelling, including the ambiguity of cardinality constraints of material relations (and only material relations!) caused by the so-called problem of collapsing *Single-tuple* and *Multiple-tuple* cardinality constraints [17]. The reader should notice that the relator type is not a UML association class as it actually addresses an ontological problem caused by the Association Class construct termed *non-lucidity at the language level* [18]. These aspects are discussed in depth in [14]. Here we only make use of this representation to make explicit the relevant relator types and the material relations derived from them.

4.1 An Ontological Analysis of Subsetting

Let us start with the example depicted in Figure 2. The relation *Enrolls* is a stereotypical example of a material relation derived from the relator *Enrollment*. So, the truthmaker of the predicate *Enrolls(x,y)* is the existence of a particular enrollment *z* connecting *x* and *y*. What about *HasPreference*? The relevant question again is to inquiry about the truthmaker of *HasPreference(x,y)*, i.e., when is it true to say that a *student x has preference for subject y* in the conceptualization underlying this model? In order for these preferences to become public (and then recorded in an information system), we can imagine an explicit manifestation of interest of Student for a list of Subjects. Notice that, the social object which records this list of interests is indeed a relator (existentially dependent on both a student and on the list of subjects). This situation is depicted in Figure 6.a below.

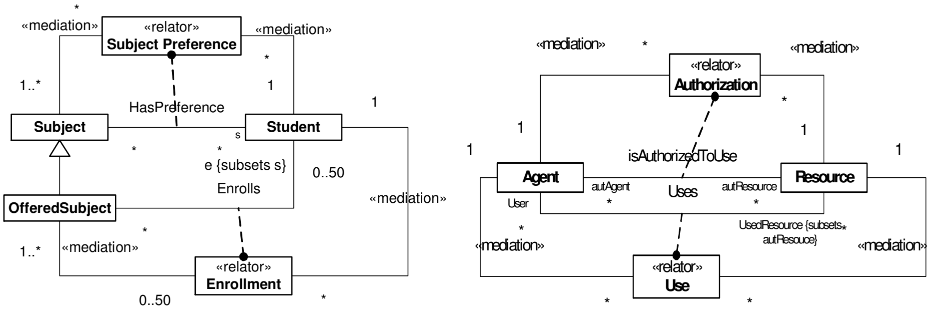

Fig.6.a-b Explicit representations of the relator types from which the relations in Figure 2 and [12] are derived, respectively.

One should notice that the two relations of Figure 6.a are derived from relator types of different kinds and are based on different foundational events (different social speech acts). In other words, there is no necessary connection between the truthmakers of these two relations. It is just a matter of accident that the extension of *Enrolls* is included in the extension of *HasPreference*, and one can easily imagine an alternative conceptualization in which this constraint is abandoned, i.e., in which students are allowed to enroll in subjects regardless of their preferences. This accidental inclusion of the extension of one relation in the extension of the other is intuitively captures in the statement in the UML specification [1, pg.39]: "*subsetting is a relationship in the domain of extensional semantics*". In summary, we postulate that:

Postulate 1: *a subsetting relation should be defined between two material relations* R_2 *and* R_1 (R_2 $_{subsets}$ R_1) *iff these relations are derived from relators of disjoint kinds and there is an inclusion constraint that includes the extension of* R_2 *in the one of* R_1.

In Figure 6.b, we present another example of the result of an ontological analysis of subsetting extracted from [12]. Here again, our analysis is able to explain the modelling choice adopted by the author. Once more both material relations are founded on relators of disjoint kinds, and once more that it is merely accidental that (in this conceptualization) resources must be authorized before used.

4.2 An Ontological Analysis of Specialization

Now, let us examine the example of Figure 3 with two material relations: *PronouncedSentence* and *Absolves*. The former is derived from a social relator *Sentence* which records the outcome of a given event (the sentence pronunciation). However, a further analysis of this relator type reveals that this type is an *abstract* one, i.e., there is no general instance of *Sentence* which is not one of its specific kinds (e.g., Conviction, Absolution). Thus, we have that the relator type associated with the *Absolves* relation, i.e., *Absolution*, is a specialization of the *Sentence* relator type (Figure 7.a). In other words, to be absolved is a specific kind of being sentenced.

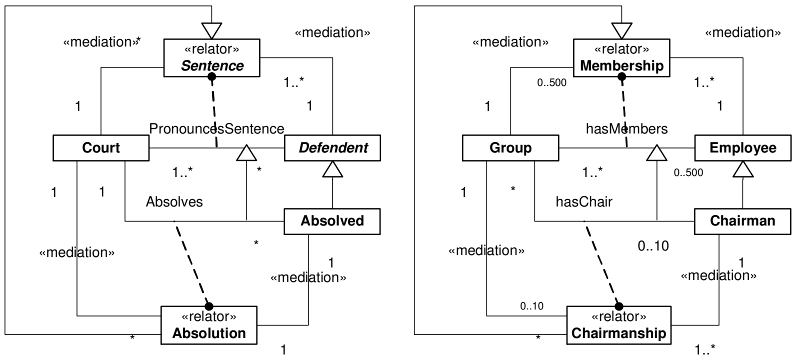

Fig.7.a-b Explicit representations of the relator types from which the relations in Figure 3 and [12] are derived, respectively.

In contrast with subsetting, specialization has an intentional relation between types: an absolution has all the properties of a general sentence (e.g., date of pronunciation) and it is founded on the very same individual event (the sentence pronunciation). This explains the intuition discussed in [12] that: (i) the specializing relation necessarily shares properties with the general one and typically includes additional properties; (ii) there is an analytical connection between these two relations. In fact, if *Sentence* is an abstract type then *Absolution* not only inherits all its properties but should also have at least of property to differentiate it from other (disjoint) kinds of sentences. Moreover, if the definition of to be absolved is to receive a sentence *z* such that *z* is an absolution, and the definition of to be sentence is to receive a sentence *z* (of any kind), then indeed the definition of *being sentenced* is part of the very definition of *being absolved*. Thus, one cannot be absolved without being sentenced in the same way that one cannot be a bachelor without being unmarried. Thus, we postulate that:

Postulate 2: *a specialization relation should be defined between two material relations R_2 and R_1 ($R_1 \triangleleft\!\!-\!\!- R_2$) if these two relations are derived from relator types RR_1 and RR_2 such that RR_2 specializes RR_1 ($RR_1 \triangleleft\!\!-\!\!- RR_2$).*

Another example of specialization is depicted in Figure 7.b. It was taken from [12, pg.242] without any change in the relations except a change in the constraints max_b and max_{b1}, in order to illustrate the following point. Notice that in all these cases there is always the (sometimes implicit) existence of at least one additional type RR_3 (e.g., *RegularMember*, *Conviction*) such that RR_3 is distinct from RR_2 (e.g., *Chairman*, *Absolution*) and RR_3 specializes RR_1. This means that the maximum cardinality constraints of R_1 (max_b) govern relations between individuals, which are mediated by both types of relators (RR_2 and RR_3). For instance, suppose that a group is allowed to have a maximum of 500 members. Thus, the maximum of Chairmen can, in principle, be ≤ 500 in that period. However, if there is a limit of 10 chairmen for the Group, there could still be up to 490 regular members. This feature explains why the maximum cardinality constraints of R_2 (max_{b1}) are not imposed on the maximum cardinality constraints of R_1 (max_b - see discussion in section 3.3).

4.3 An Ontological Analysis of Redefinition

We now turn to the example of Figure 4 with three material relations: (i) *Participates* defined between the types *Employee* and *Project*; (ii) *ParticipatesCritical* defined between the types *Senior* and *Critical*; and (iii) *Participates* defined between the types *Junior* and *Project*. In this case, we have that all these relations are derived from the same relator type and the same foundation, namely, an allocation event and the resulting *Allocation* contract. In the situation represented by this model, the different ways of participating in a project (which entail the different relations of participation) are defined by the different types associated with the association end opposite to the redefining end of these relations. In other words, the differences in the ways a junior or senior employee participate in a project are motivated by difference in properties of these different types of employees not by difference in different types of *Allocation*.

For instance, let us suppose the case that the property *"having less than 10 years of experience"* differentiates junior from senior employees. Now, suppose an individual *John* which by virtue of having this *(intrinsic) quality* instantiates *Junior*. It is because of this quality that *John* is constrained to participate in at most 3 critical projects, not because he is connected to these projects by a relator of specific kind of *Allocation*. Moreover, notice that, in this case, *John* instantiates *Junior* (the specializing class) prior to establishing any relation to a project, i.e., the nature of this relation is constrained by the specific type *John* instantiates. In contrast, if one examines the specializing type *Absolved*, one shall realize that an individual instantiates *Absolved* because of the specific type of *Sentence* that mediates him. Thus, in the latter case, it is the specific type this individual instantiates which is determined by the specific nature of the relator that mediates him. In other words, in the case of redefinition, the type the relata (instances connected to the association end) instantiate is defined *a priori* and the participation constraints in the relation follows from that; in the case of specialization, the type the relata instantiates in that relation is defined *a posteriori* entailed by the type of relator binding them. The remodelling of this situation with an explicit representation of the founding relator type is depicted in Figure 8.a. Another example of association redefinition, taken from [19], is depicted in Figure 8.b where it is also remodelled to represent the founding relator type.

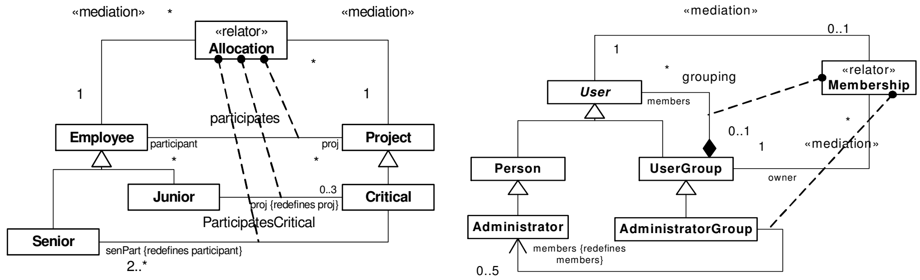

Fig.8.a-b Explicit representations of the relator types from which the relations in Figure 3 and [19] are derived, respectively.

Notice that the ontological differences between these concepts also explains the difference discussed in section 3.3 on how these two relations influence the maximum cardinality constraints of the original (redefined or specialized) relation. As discussed there, only in the case of redefinition, multiplicity is redefined when the redefining end has a multiplicity which is more restrictive than that of the redefined end. This can be observed, for instance, in Figure 8.a when a junior employee can participate in at most 3 critical projects. The reader can contrast it with Figure 7.b in which a Group can have at most 10 chairs and still an upper bound of 500 employees. Again, in the latter case, the way entities participate in these relations is defined by the relator type: to be a Chairman is to have a relational bond to this group of a specific nature, namely, a Chairman Membership (Chairmanship). However, employees can participate in groups in different ways, for instance, as regular members. In the former case, conversely, it is the specific type the employee instantiates which *a priori* determinates her constraints in participating in the relation. For this reason, there is no other manner a junior employee can participate in this relation with a critical project. In other words, the specific (redefining) way of participating in this relation is the only way of participating in the general (redefined) relation. As a consequence, the multiplicity constraints of the former should be valid for the general case.

Finally, notice that this analysis also explains the syntactic constraint depicted in Figure 4 which differentiates redefinition from subsetting and specialization, namely, that in the case of redefinition the type A_1 in the opposite end of the redefining association end of association R_1 must be a subclass of type A. This is due to the fact that the difference in ways that instances of A participate in association R is determined by differences in intrinsic properties of these instances. Thus, if we have at least of property *p* that some instances of A must have in order to participate in R in a specific way (e.g., R_1) then we can define the type A_1 such that *$A_1(x)$ iff A(x) and (x possesses p)*. Conversely, notice that without a single property that differentiates particular subtypes of A then we cannot explain why they participate in association R in different manners and subject to different constraints. After all, in that case, all instance of A would have exactly the same properties including the specific relator type that binds then to instances of B. As a result of this analysis, we postulate that:

Postulate 3: *a redefinition relation should be defined between two material relations* R_2 *and* R_1 (R_2 $_{redefines}$ R_1) *if: (i) these two relations are derived from the same relator type RR; (ii) there is a type* A_1 *connected to one of the association ends of* R_2 *such that* A_1 *is a specialization of A* ($A \triangleleft\!\!-\!\!- A_1$) *and A is connected to the association end of* R_1 *equivalent to that of* A_1.

5 Related Work

Some studies attempt to clarify or formalize the meaning of one or more of the three constructs under consideration. They can be classified in two groups: first, works that provide informal definitions, explanations or examples and, second, works that formalize the semantics of one or more of the constructs.

In the first group, we can mention [4] developed by Stevens, that deals with the association specialization of UML 1.4. She points out the difficulty of distinguishing it from subsettings and argues that specializations in UML 1.4 may be used to represent inclusion constraints and participation constraints over associations. We must note that this latter case is better represented in UML 2 as a redefinition. Moreover, Milicev [19] defines the subsetting as an implicit constraint attached to association ends. He also defines that end redefinitions may redefine the type and the multiplicity of the redefined end. Costal and Gómez [20] describe how to use redefinitions and analyze their interactions with taxonomic constraints. Olivé [12] states that inclusion constraints between associations can be represented by means of subsettings or specializations. He indicates that specializations must be used when the specific association has the defining properties of the general one together with others and subsettings in the rest of cases. Olivé indicates that redefinitions can be used to represent participant refinements and to strength cardinality [21]. From this first set of works, the only one that sketches a distinction between subsetting and specialization is that of Olivé [12] by means of intuitive explanations and well-chosen examples.

In the second group, that presents semantic formalizations, [10] by Alanen and Porres, deals with association subsetting. In addition, Kleppe and Rensink [6] formalize subsetting and redefinition as formal extensions of a type graph (not for all scenarios). Amelunxen and Schürr [11] present a set theoretic formalization of graph transformations that is used to define the semantics of a set of UML features including subsetting, specialization and redefinition. Bildhauer [22] describes the semantics of subsetting, redefinition and specialization by providing formal definitions for specific examples. All these works, give consistent interpretations of the constructs, also consistent with our semantic formalization. However, ours is the only one that explicitly characterizes the specific participation, minimum and maximum cardinality constraints implied by the constructs. Moreover, none of these works develops an ontological analysis and, thus, they are not able to distinguish subsetting from specialization and from some cases of redefinition. There exist some ontological analyses of relations, such as [23], but they do not cover subsetting, specialization or redefinition. Hence, to our knowledge, ours is the first contribution presenting a comprehensive view by specifying both the formal semantics and ontology-based semantics and able to differentiate the three constructs.

6 Final Considerations

The purpose of a conceptual model is to support users in tasks such as communication, understanding and problem-solving about a certain universe of discourse. For this reason, two fundamental quality attributes for a conceptual modeling language are: (i) *semantic discrimination*, i.e., users of the language must understand the semantic distinctions between language's constructs and the semantic consequences of applying these constructs; (ii) *ontological clarity*: the users of the language must understand what these constructs represent in terms of phenomena in reality. In this paper, we make a contribution in both these directions by specifying the *formal semantics* and the (ontology-based) *real-world semantics* of three important, yet poorly understood, UML association constructs, namely, subsetting, specialization and redefinition.

As a future work, we intend to test the ontological analysis put forth here by conducting empirical studies. In particular, we would like to further test the ability of this theory to predict the outcome of modeling choices. However, we should point out that one significant challenge of conducting such a study at this point, namely, that, despite their importance, these constructs are still unknown to a large community of modelers. This is specially the case of redefinition, which has only been introduced in the newest version of UML.

Nonetheless, we have conducted a preliminary empirical investigation on this theory using as a benchmark a catalog of models produced by different people using these constructs. This catalog has been obtained from the papers reviewed in our related work section, thus, comprising a set of ten examples produced by different authors (which can be considered experts in the field). The result of this study can be found in [24]. As demonstrated there, our theory was able to predict the modeling choices made by the authors in 90% of the cases. We take this to be preliminary evidence that the ontological theory developed here together with the modeling postulates derived from it constitute a *descriptive theory for explaining and predicting*, as well as a *prescriptive theory for design and action*.

Acknowledgements. This work has been partly supported by the *Ministerio de Ciencia y Tecnologia* and FEDER under projects TIN2008-00444/TIN and TIN2010-19130-C02-01, *Grupo Consolidado*, and by the funding agencies CNPq (Project Grant #481906/2009-6 and Productivity Grant #309382/2008) and FAPES (Project Grant #45444080/09).

References

1. Object Management Group: OMG Unified Modeling Language (OMG UML), Superstructure, V2.3, (formal/2010-05-05). (2010)
2. Batra, D., Hoffler, J.A., Bostrom, R. P.: Comparing representations with relational and EER models. Communications of the ACM 33, 126-139 (1990)
3. Rumbaugh, J., Jacobson, I., Booch, G.: The unified modeling language reference manual, 2nd ed. Addison-Wesley (2005)

4. Stevens, P.: On the interpretation of binary associations in the Unified Modelling Language. Software and Systems Modeling 1, 68-79 (2002)
5. Szlenk, M.: Formal Semantics and Reasoning about UML Class Diagram. In: DEPCOS-RELCOMEX'06, pp. 51-59. IEEE Computer Society (2004)
6. Kleppe, A., Rensink, A.: On a Graph-Based Semantics for UML Class and Object Diagrams. Electronic Communications of the EASST 10 (2008)
7. Gogolla, M., Richters, M.: Expressing UML Class Diagrams Properties with OCL. In: Clark, A., Warme, J. (eds.) Object Modeling with the OCL 2002. LNCS, vol. 2263, pp. 85-114. Springer, Heidelberg (2002)
8. Object Management Group: Object Constraint Language (OCL), Version 2.2. Available Specification (formal/2010-02-01) (2010)
9. Richters, M., Gogolla, M.: On Formalizing the UML Object Constraint Language OCL. In: Ling, T.W., Ram, S., Lee, M.L. (eds.) ER 1998. LNCS, vol. 1507, pp. 449-464. Springer, Heidelberg (1998)
10. Alanen, M., Porres, I.: Basic Operations over Models Containing Subset and Union Properties. In: Nierstrasz, O., Whittle, J., Harel, D., Reggio, G. (eds.) MoDELS 2006. LNCS, vol. 4199, pp. 469-483. Springer, Heidelberg (2006)
11. Amelunxen, C., Schürr, A.: Formalising model transformation rules for UML/MOF 2. IET Software 2, 204-222 (2008)
12. Olivé, A.: Conceptual modeling of information systems. Springer-Verlag, Heidelberg (2007)
13. Nieto, P., Costal, D., Gómez, C.: Enhancing the semantics of UML association redefinition. Data Knowl. Eng. 70, 182-207 (2011)
14. Guizzardi, G., Wagner, G.: What's in a Relationship: An Ontological Analysis. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) ER 2008. LNCS, vol. 5231, pp. 83-97. Springer, Heidelberg (2008)
15. Guizzardi, G.: Ontological Foundations for Structural Conceptual Models. Universal Press, The Netherlands, ISBN 90-75176-81-3 (2005)
16. Heller, B., Herre, H.: Ontological Categories in GOL. Axiomathes 14, 71-90 (2004)
17. Bock, C., Odell, J.: A More Complete Model of Relations and Their Implementation: Relations as Object Types. Journal of Object-Oriented Programming 10(3) (1997)
18. Gurr, C.A.: Effective Diagrammatic Communication: Syntatic, Semantic and Pragmatic Issues. Journal of Visual Languages and Computing, 10, 317-342 (1999)
19. Milicev, D.: Model-Driven Development with Executable UML. Wiley Pub. Inc (2009)
20. Costal, D., Gómez, C.: On the Use of Association Redefinition in UML Class Diagrams. In: Embley, D.W., Olivé, A., Ram, S. (eds.) ER 2006. LNCS, vol. 4215, pp. 513-527. Springer, Heidelberg (2006)
21. Costal, D., Olivé, A., Teniente, E.: Relationship Type Refinement in Conceptual Models with Multiple Classification. In: Kunii, H.S., Jajodia, S., Sølvberg, A. (eds.) ER 2001. LNCS, vol. 2224, pp. 397-411. Springer, Heidelberg (2001)
22. Bildhauer, D.: On the Relationships Between Subsetting, Redefinition and Association Sepecialization. In: DB&IS'10. (2010)
23. Evermann, J.: The Association Construct in Conceptual Modelling - An Analysis Using the Bunge Ontological Model. In: Pastor, O., Cunha, J.F. (eds.) CAiSE 2005. LNCS, vol. 3520, pp. 33-47. Springer, Heidelberg (2005)
24. Costal, D., Gómez, C., Guizzardi, G.: On the Meanings of Subsetting, Specialization and Redefinition in UML, Technical Report, (2011),http://hdl.handle.net/2117/12827