# On the Meanings of Subsetting, Specialization and Redefinition 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)
[2]Ontology and Conceptual Modeling Research Group (NEMO), Federal University of Espírito Santo (UFES), Vitória-ES, Brazil

{dolors|cristina}@essi.upc.edu, gguizzardi@inf.ufes.br

## Introduction

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.

In this report, we present a preliminary empirical investigation on these constructs using as a benchmark a catalogue of model examples produced by different authors which can be considered experts in the conceptual modelling field.

For each example: 1) an ontological analysis has been performed; 2) the analysis has been used to predict which one of the three constructs should be the modelling choice of the author; and 3) our prediction has been compared to the actual choice of the author.

The ontological analysis of the examples focuses on the relator types of the involved associations and discriminates three cases: 1) the relator types are different (in that case we postulate that a subsetting should be defined if there is an inclusion constraint between their extensions); 2) one relator type specializes the other (we postulate a specialization for this case); and 3) their relator type is the same (redefinition).

The sources used to obtain the examples have been:

- Alanen, M., Porres, I.: Basic Operations over Models Containing Subset and Union Properties. MoDELS 2006, LNCS 4199, 469-483.
- Villegas, A., Olivé, A., Vilalta, J.: Improving the Usability of HL7 Information Models by Automatic Filtering, IEEE 6th World Congress on Services, Florida (USA), 2010.

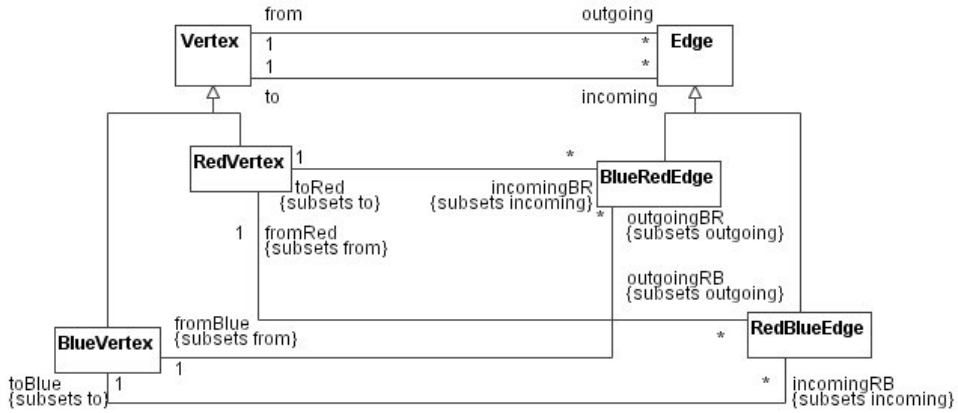
- Milicev, D.: Model-Driven Development with Executable UML. Wiley Pub. Inc, 2009.
- Olivé, A.: Conceptual modeling of information systems. Springer-Verlag, 2007.

Our investigation focuses on subsetting, specialization and redefinition of associations over base (non-derived) associations. Therefore, we have taken all the examples from the previous references where one or more of the three constructs has been used for non-derived associations. Those cases where the constructs are combined with derived associations or those where subsetting is combined with a derived union have been discarded. Then, we have selected 10 examples out from the 4 source references used. In the following, we describe the study of each example and, finally, we give some conclusions.

## Example 1

**Source:** Alanen, M., Porres, I.: Basic Operations over Models Containing Subset and Union Properties. MoDELS 2006, LNCS 4199, 469-483.

**Figure:**




**Description from the source:**

This diagram shows two classes: *Vertex* and *Edge*, and four properties: *from*, *to*, *outgoing* and *incoming*. Each property has another property as its opposite. Together they define an association that is represented as a single line. In the example, we have the *from-outgoing* and the *to-incoming* associations.

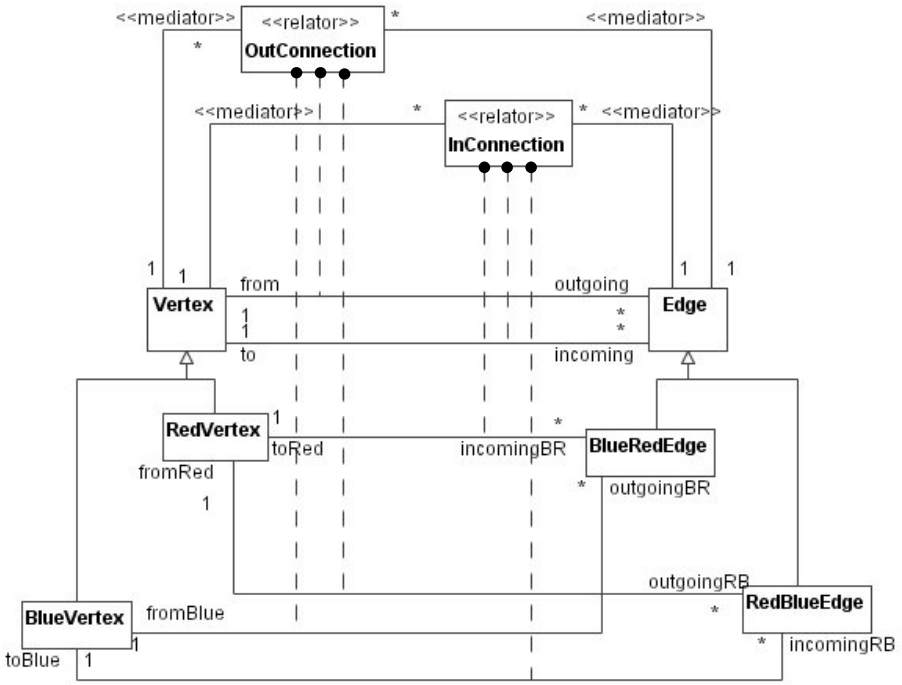
…The classes *Blue Vertex* and *Red Vertex* will now be specializations of *Vertex*. Also, the *fromRed* and *toBlue* properties will become subsets of the *from* and *to* properties, and similarly for the other properties. …The intuition behind the metamodel is as follows: an element of type *Red Vertex* has four slots that correspond to properties *outgoing*, *incoming*, *outgoingRB* and *incomingBR*. Elements of type *Edge* can be inserted into the *outgoing* or *incoming* slot and elements of type *RedBlue*

*Edge* can also be inserted into *outgoingRB*. At any moment, the contents of the slot *outgoingRB* should be a subset of the contents of the slot *outgoing*.

**Constructs according to source:**

- Association *toRed-incomingBR* and association *to-incoming*: association subsetting
- Association *fromRed-outgoingRB* and association *from-outgoing*: association subsetting
- Association *fromBlue-outgoingBR* and association *from-outgoing*: association subsetting
- Association *toBlue-incomingRB* and association *to-incoming*: association subsetting

**Ontological analysis:**




Independently of the colour of vertexes, the relations that connect a vertex with an outgoing edge are derived from the same relator *OutConnection*. Similarly, the relations that connect a vertex with an incoming edge are also derived from the same relator *InConnection*. In other words, the different ways of connecting vertexes and edges represented by the associations of the example are motivated by the vertex colour not by difference in different types of *OutConnection* or *InConnection*. Therefore:

- Association *toRed-incomingBR* and association *to-incoming*: their relator type is the same.
- Association *fromRed-outgoingRB* and association *from-outgoing*: their relator type is the same
- Association *fromBlue-outgoingBR* and association *from-outgoing*: their relator type is the same
- Association *toBlue-incomingRB* and association *to-incoming*: their relator type is the same

**Constructs according to the ontological analysis:**

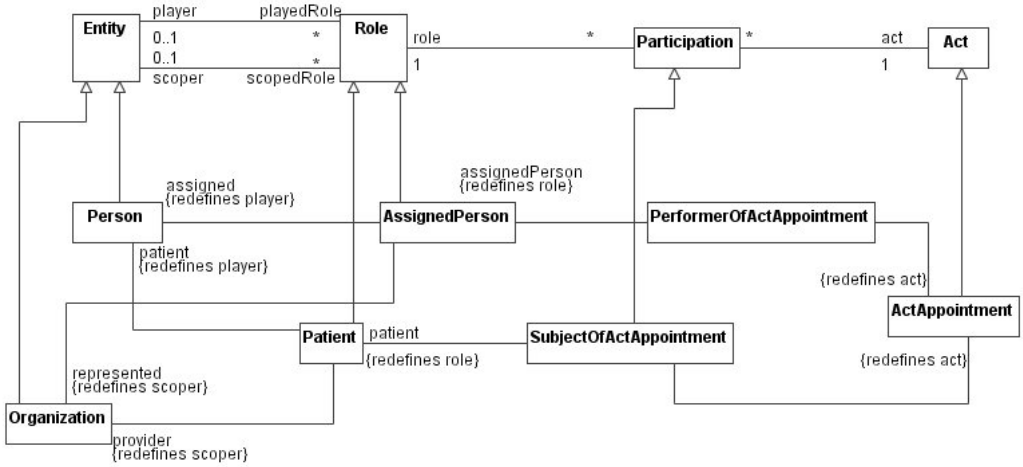
According to the ontological analysis above the construct that relates the associations are not subsettings as specified in the original example but redefinitions.

- Association *toRed-incomingBR* and association *to-incoming*: association redefinition
- Association *fromRed-outgoingRB* and association *from-outgoing*: association redefinition
- Association *fromBlue-outgoingBR* and association *from-outgoing*: association redefinition
- Association *toBlue-incomingRB* and association *to-incoming*: association redefinition

## Example 2

**Source:** Villegas, A., Olivé, A., Vilalta, J.: Improving the Usability of HL7 Information Models by Automatic Filtering, IEEE 6th World Congress on Services, Florida (USA), 2010.

**Figure:**



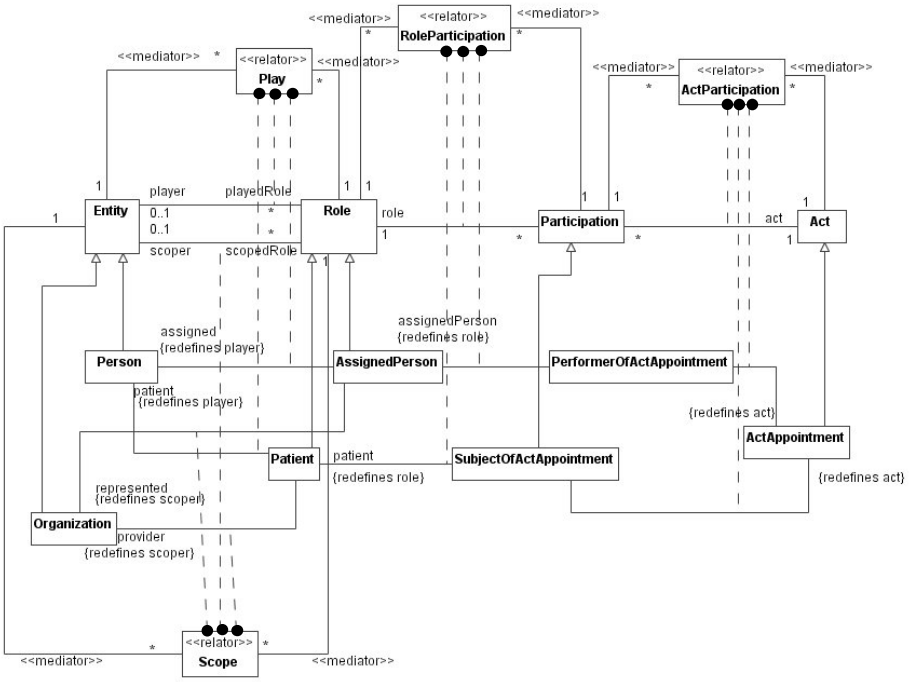
**Description from the source:**

RIM comprises …backbone classes: *Act*, *Participation*, *Entity*, *Role*… Figure … shows a few refinements related to the *ActAppointment* class. The instances of this class are appointments (a particular kind of *Act*). There may be several kinds of participations in an appointment. Figure … shows only two of them: *PerformerOfActAppointment* and *SubjectOfActAppointment*. …The overall semantics of these redefinitions is that the performer of an appointment is a *Person* that plays the role *AssignedPerson* and that the subject of an appointment is a *Person* that plays the role *Patient*. …Figure … also shows the redefinitions of associations *player-playedRole* and *scoper-scopedRole* between *Entity* and *Role*. The *player* and the *scoper* of an *AssignedPerson* and of *Patient* must be a *Person* and an *Organization*, respectively.

**Constructs according to source:**

- Association *player-playedRole* and association *assigned-assignedPerson*: association redefinition
- Association *player-playedRole* and association *patient-patient*: association redefinition
- Association *scoper-scopedRole* and association *represented-assignedPerson*: association redefinition
- Association *scoper-scopedRole* and association *provider-patient*: association redefinition
- Association *role-participation* and association *assignedPerson-performerOfActAppointment*: association redefinition
- Association *role-participation* and association *patient-subjectOfActAppointment*: association redefinition
- Association *act-participation* and association *actAppointment-performerOfActAppointment*: association redefinition
- Association *act-participation* and association *actAppointment-subjectOfActAppointment*: association redefinition

**Ontological analysis:**




- Association *player-playedRole* and association *assigned-assignedPerson*: their relator type is the same
- Association *player-playedRole* and association *patient-patient*: their relator type is the same
- Association *scoper-scopedRole* and association *represented-assignedPerson*: their relator type is the same
- Association *scoper-scopedRole* and association *provider-patient*: their relator type is the same
- Association *role-participation* and association *assignedPerson-performerOfActAppointment*: their relator type is the same
- Association *role-participation* and association *patient-subjectOfActAppointment*: their relator type is the same
- Association *act-participation* and association *actAppointment-performerOfActAppointment*: their relator type is the same
- Association *act-participation* and association *actAppointment-subjectOfActAppointment*: their relator type is the same

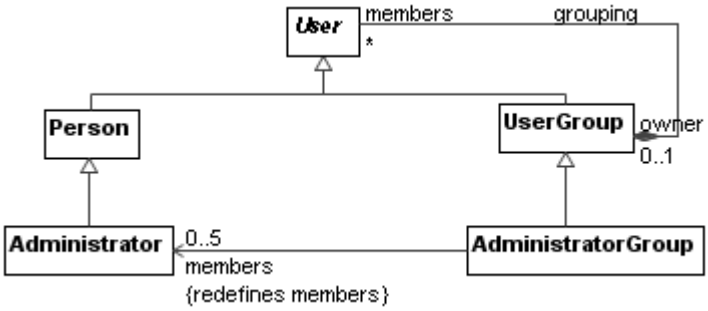
**Constructs according to the ontological analysis:**

The same constructs as the source states.

## Example 3

**Source:** Milicev, D.: Model-Driven Development with Executable UML. Wiley Publishing, Inc, 2009, (page 317).

**Figure:**



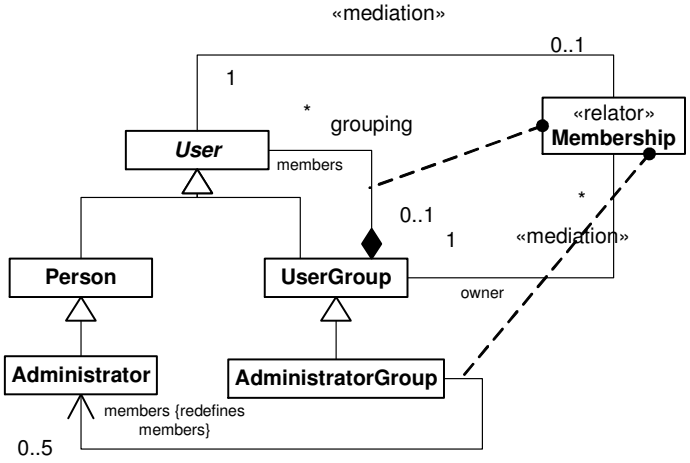

**Description from source:**

The association grouping specifies that a User Group, in general, can have an arbitrary number of Users as its members. However, an Administration Group, which is a special kind of User Group, can have only Administrators as its members, and at most five of them. Of course, an Administrator is also a kind of User. Note that an Administrator (being also a Person and a User) can still be a member of a general User Group because it has not be redefined the property owner.

**Construct according to the source:**

- Association *grouping* and association *administratorGroup-administrator*: association redefinition

**Ontological analysis:**



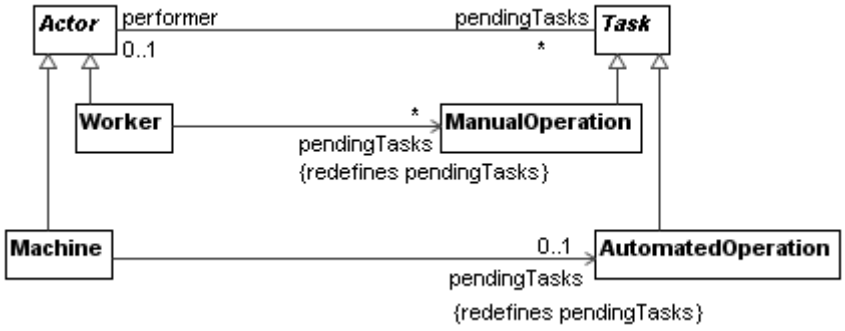
**Construct according to our analysis:**

The same construct as the source states since, as the previous figure shows, the redefined and redefining relations are derived from the same relator type (*Membership*) and the same foundation. 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.

## Example 4

**Source:** Milicev, D.: Model-Driven Development with Executable UML. Wiley Publishing, Inc, 2009, (page 317).

**Figure:**



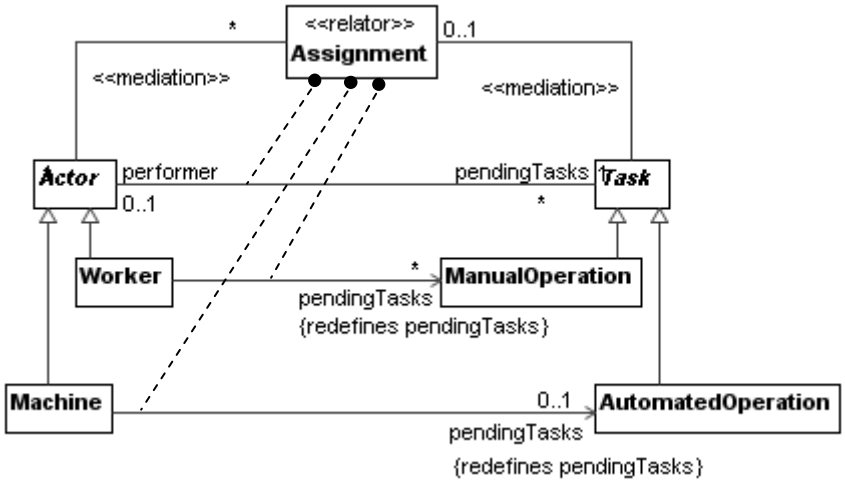

**Description from source:**

An Actor, in general, can have an arbitrary number of pending Tasks. However, a Worker, as a kind of Actor, can have only Manual Operations as its pending Tasks, while a Machine can have at most one Automated Operation as its pending Task. It is interesting to note that, as long as the class ManualOperation does not redefine the property performer, it can be assigned as pending Task of an Actor of a different kind than Worker (or Machine, since Machine constrains its pending Tasks to Automated Operations only)

**Construct according to the source:**

- Association *performer-pendingTasks* and association *worker-pendingTasks*: association redefinition
- Association *performer-pendingTasks* and association *machine-pendingTasks*: association redefinition

**Ontological analysis:**



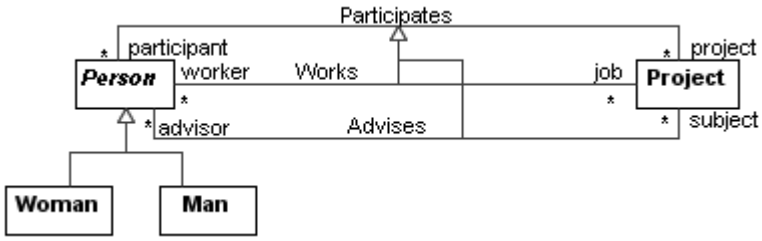

**Construct according to our analysis:**

Our analysis is able to explain the choice adopted by the author. In this case all the material relations are derived from the same relator type (*Assignment*).

## Example 5

**Source:** Olivé, A.: Conceptual modeling of information systems. Springer-Verlag, 2007, (page 168-169).

**Figure:**



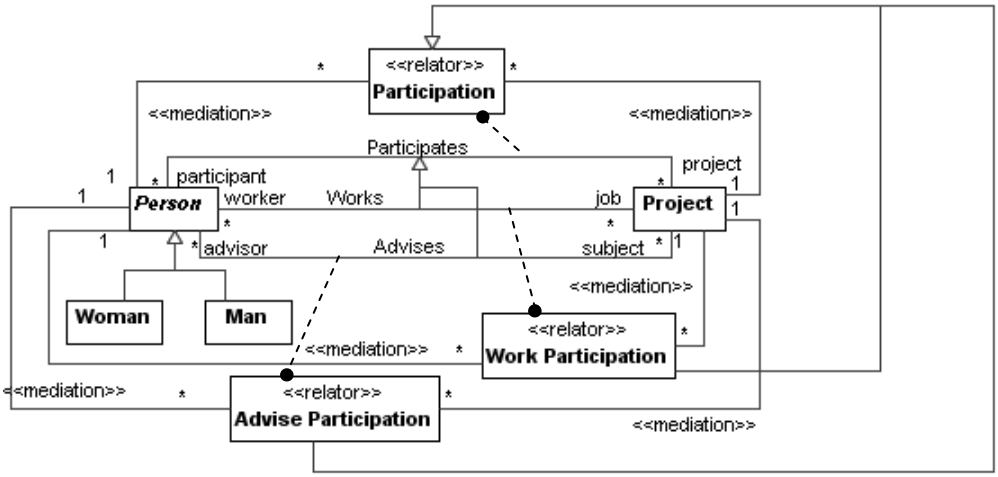

**Description from source:**

The population of *Participates* is the union of *Works* and of *Advises*.

**Construct according to the source:**

- Association *Participates* and association *Works*: association specialization
- Association *Participates* and association *Advises*: association specialization

**Ontological analysis:**



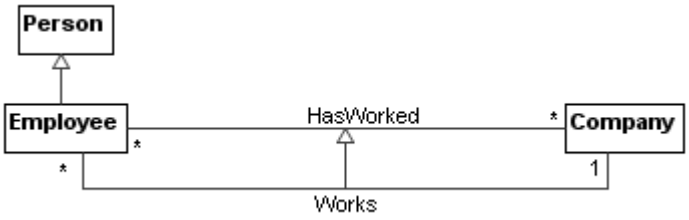

**Construct according to our analysis:**

The same constructs as the source states since, as the previous figure shows, the relator types (*Work Participation* and *Advise Participation*) of the specific relations are subtypes of the relator type (*Participation*) of the general relation.

## Example 6

**Source:** Olivé, A.: Conceptual modeling of information systems. Springer-Verlag, 2007, page (176).

**Figure:**



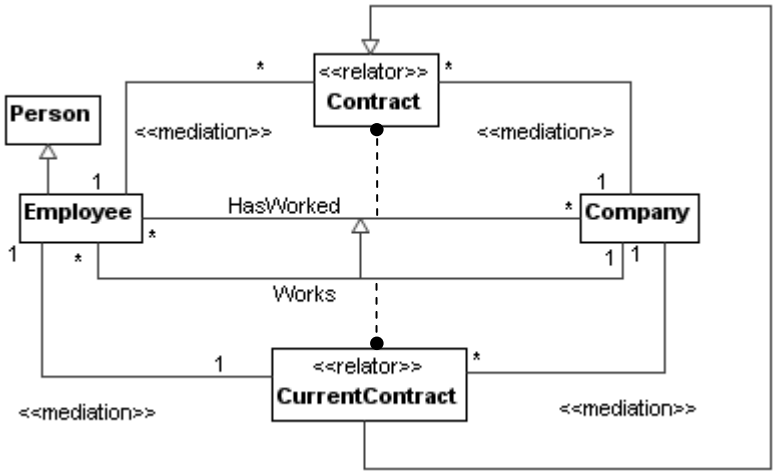

**Description from source:**

The population of *HasWorked* is the union of *Works* and the set of relationships explicitly classified as *HasWorked*.

**Construct according to the source:**

- Association *HasWorked* and association *Works*: association specialization

**Ontological analysis:**



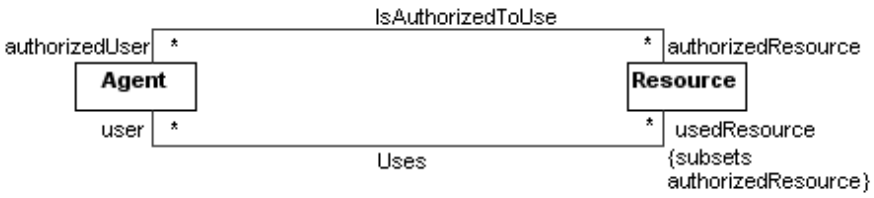

**Construct according to our analysis:**

The same construct as the source states since, as the previous figure shows, the relator type of the specific relation is a subtype of the relator type of the general relation.

## Example 7

**Source:** Olivé, A.: Conceptual modeling of information systems. Springer-Verlag, 2007, (page 199).

**Figure:**



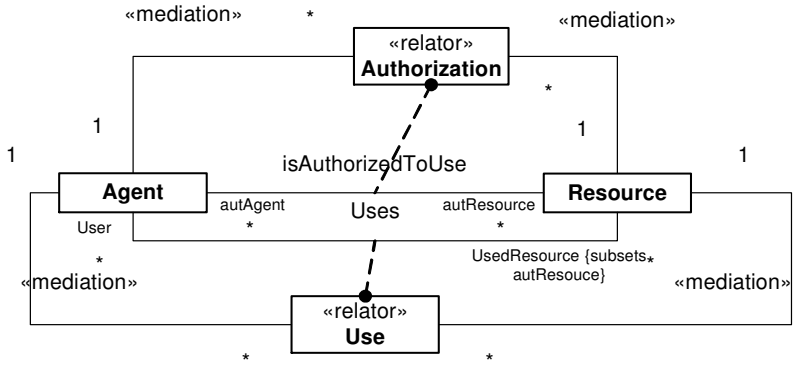

**Description from source:**

The model represents agents and resources. An agent may use a resource only if it is authorized to use it.

**Construct according to the source:**

- Association *IsAuthorizedToUSe* and association *Uses*: association subsetting

**Ontological analysis:**



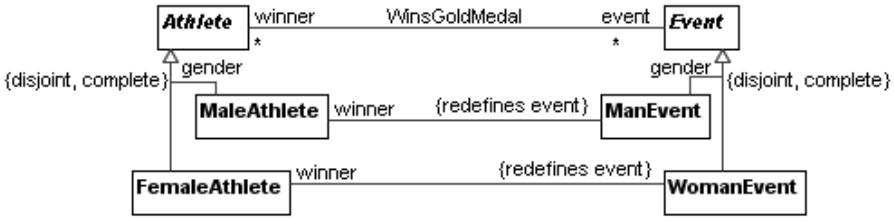

**Construct according to our analysis:**

Here, our analysis is able to explain the modelling choice adopted by the author. Both material relations are founded on relators of disjoint kinds (*Authorization* and *Use*) and the set of resources used by an agent is a subset of the set of the authorized resources to use by the agent. Note that, in this case, it is merely accidental that (in this conceptualization) resources must be authorized before used.

## Example 8

**Source:** Olivé, A.: Conceptual modeling of information systems. Springer-Verlag, 2007, (page 230).

**Figure:**



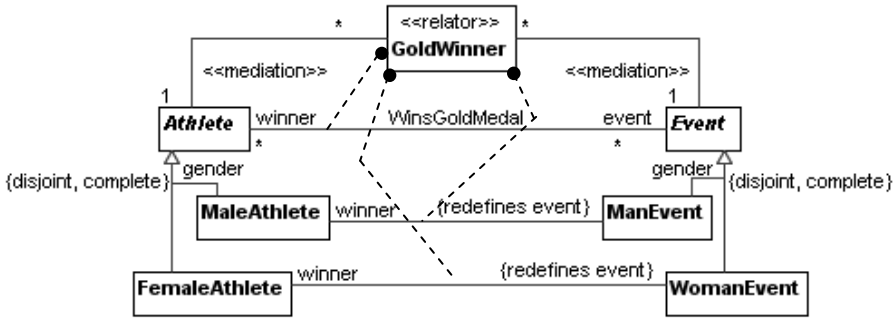

**Description from source:**

The model represents athletes that win gold medals in events. If the winner is an instance of *MaleAthlete*, then the event must be an instance of *ManEvent*. If the winner is an instance of *FemaleAthlete*, then the event must be an instance of *WomanEvent*.

**Construct according to the source:**

- Association *WinsGoldMedal* and association *winner-manEvent*: association redefinition
- Association *WinsGoldMedal* and association *winner-womanEvent*: association redefinition

**Ontological analysis:**



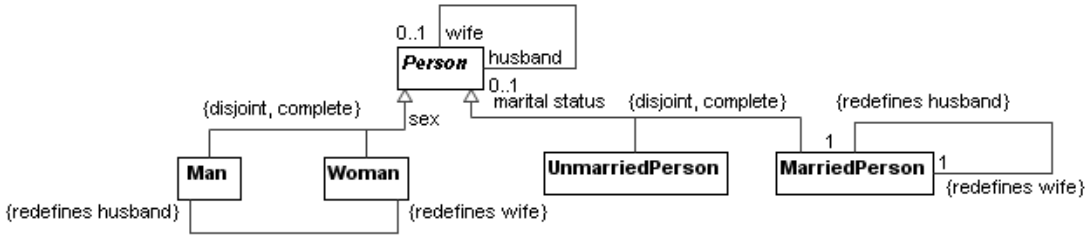

**Construct according to our analysis:**

The same constructs as the source states. Our analysis concludes that all the material relations are founded on the same relator type (*GoldWinner*).

## Example 9

**Source:** Olivé, A.: Conceptual modeling of information systems. Springer-Verlag, 2007, (page 235).

**Figure:**



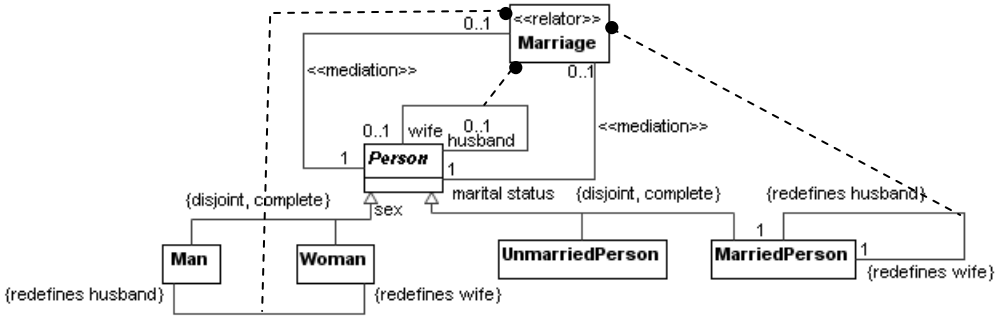

**Description from source:**

The model represents people that may be married or unmarried. Two redefinitions have been defined: one to indicate that the *husband* must be a *Man* and that the *wife* must be a *Woman* and another to indicate that the *husband* and the *wife* must be *MarriedPeople*.

**Construct according to the source:**

- Association *wife-husband* and association *man-woman*: association redefinition
- Association *wife-husband* and association *marriedPerosn-marriedPerson*: association redefinition

**Ontological analysis:**



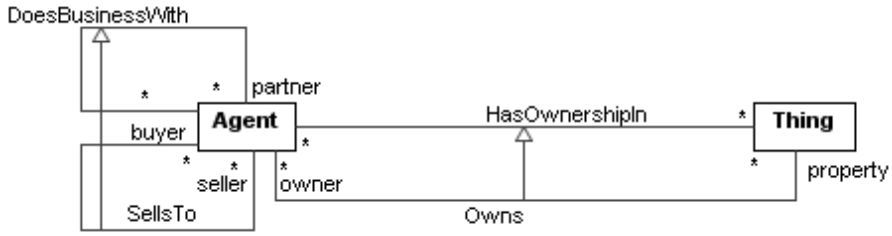

**Construct according to our analysis:**

Our analysis recommends the same constructs as the author used. In this scenario, all the material relations are derived from the same relator type (*Marriage*) and the same foundation. Moreover, 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.

## Example 10

**Source:** Olivé, A.: Conceptual modeling of information systems. Springer-Verlag, 2007, (page 241).

**Figure:**



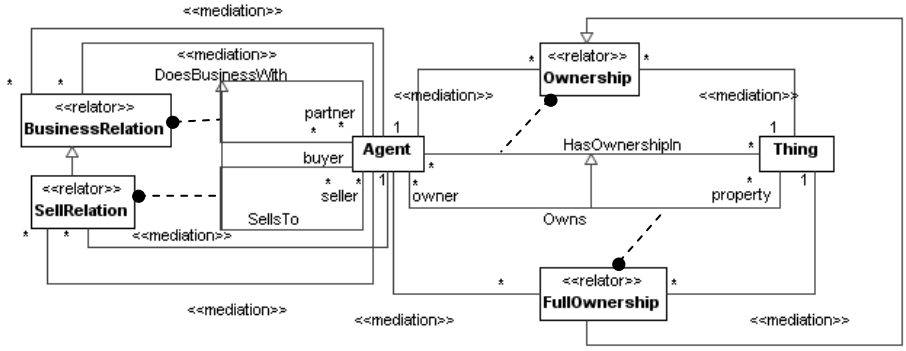

**Description from source:**

An Agent has ownership in a thing if it owns part of it or the whole thing. An agent owns a thing if it has full ownership of the thing (taken from the Cyc ontology). Obviously, multiplicity of *owner* role could be more restrictive (0..1).

**Construct according to the source:**

- Association *HasOwnershipIn* and association *Owns*: association specialization
- Association *DoesBusinessWith* and association *SellsTo*: association specialization

**Ontological analysis:**




**Construct according to our analysis:**

The same constructs as the source states since, as the previous figure shows. The relator type (*FullOwnership*) of the specific relation (*Owns*) is a subtype of the relator

type (*Ownership*) of the general relation (*HasOwnershipIn*). Similarly, the relator type (*SellRelation*) of the specific relation (*SellTo*) is a subtype of the relator type (*BusinessRelation*) of the general relation (*DoesBusinessWith*).

## Conclusions

In this report, we have described a preliminary empirical investigation on association subsetting, association specialization and association redefinition using as a benchmark a catalogue of model examples produced by different authors who can be considered experts in the conceptual modelling field. We have studied 10 examples out from 4 different sources. There are 3 examples which use association specializations, 2 using subsettings and 5 using redefinitions. There are 9 examples in which the constructs predicted by the ontological analysis are the same constructs chosen by the author and 1 case in which the prediction is different from the actual choice of the author (see example 1). Therefore, our postulates have been able to predict the modeling choices made by the authors in 90% of the cases.