View Operations on Objects with Roles for a Statically Typed Database Language*

Antonio Albano, Giuseppe Antognoni, and Giorgio Ghelli

Abstract

To deal with the evolution of data and applications, and with the existence of multiple views for the same data, the object data model needs to be extended with two different sets of operations: object extension operations, to allow an object to dynamically change its type, and object viewing operations, to allow an object to be seen as if it had a different structure. Object extension and object viewing operations are related, in that they are both identity-preserving operations; but different, in that object extension may modify the behavior of the original object while object viewing creates a new view for the original object without modifying its behavior. In this paper a set of object viewing operations is defined in the context of a statically and strongly typed database programming language, which supports objects with roles, and the relationships with object extension and role mechanisms is discussed. We then show how the object viewing operations can be used to give the semantics of a higher level mechanism to define views for object databases. Examples of the use of these operations are given with reference to the prototype implementation of the language Galileo 97.

Index Terms—Object databases, object data models, objects with roles, object views, database programming languages

1 Introduction

The success of the object data model is mainly due to its expressivity and its ability to deal with both structural and procedural aspects of real-world entities. The technology of object databases (ODBs) is, however, still relatively new and has some limitations, in particular with respect to the possibility of adapting a database to the evolution of application requirements. It is widely recognized that long lasting, evolving databases need support for (a) objects that may evolve over time, and may exibit a different behavior in different contexts, (b) views mechanisms to adapt the database schema to the changing needs of users, and (c) schema evolution, data evolution and versioning mechanisms to support changing application requirements.

In this paper we consider a set of operators on objects, in the context of a statically and strongly typed database programming language, to support the first two aspects of database evolution:

- to allow objects to acquire and lose types during their life, to model the dynamic and context dependent behavior of real-world entities. For example, to model situations such as that of a human being who is initially classified as a person, then becomes a student, and finally an employee. When an object acquires a new type it preserves its identity, but it may change its behavior. The definition of a type-safe extension operator implies, as will be shown later, that objects are enriched with a

*Author's address: Dipartimento di Informatica, Università di Pisa, Corso Italia, 40, 56125 Pisa, Italy. E_mail: {albano, ghelli}@di.unipi.it This work has been supported in part by grants from the C.E.C. under ESPRIT BRA No.6309 (FIDE2: Fully Integrated Data Environment), by PASTEL Esprit Working Group 22552, and by "Ministero dell'Università e della Ricerca Scientifica e Tecnologica" (MURST), Progetto INTERDATA.

role notion. In our approach, this means that every object is equipped with a set of roles which can change over time; a message is sent to an object only through one if its roles, and the answer to the message may depend on the role the message is addressed to, and on the roles possessed by the object.

- to allow the creation of *virtual objects* starting from base or virtual objects, using a set of object algebra operators. A virtual object is used to model a different interface to the same conceptual entity, and so it has the same identity as the object from which it has been constructed, but it may have a different structure and behavior which, in contrast to role acquisition, does not affect the behavior of the object from which it has been constructed. We will also show how the object algebra operators can be used to give the semantics of a higher level mechanism to define views for ODBs, i.e. query-defined collections of virtual objects.

As will be shown later, objects with roles and virtual objects are two related mechanisms, hence it is essential to compare them in order to understand their similarities and differences. Objects with roles and virtual objects have been implemented in the original Galileo 97 language, a statically and strongly typed language, thus proving that flexibility can be achieved without compromising the well known benefits of static typechecking. Galileo 97 is the result of a redesign of the Galileo language [5], and is aimed at a better integration of an object mechanism into a database programming language. The role mechanism in Galileo 97 is based on the one proposed for Fibonacci [4], [7]. The role mechanism is thus not a novel contribution of this paper, but is presented here for several reasons: (a) to compare it with the virtual objects mechanism proposed, (b) to show their different semantics and their interactions, and (c) to discuss why both mechanisms are needed to support database evolution at both instance and schema levels. The main contributions of this paper are:

- the definition of an identity-preserving object algebra which allows virtual objects to be defined and typed; the novelty of our approach is that our operators are defined in the context of a statically strongly typed language, where data and code are mixed, and we define a type system to deal with both of them (Section 4);

- the definition of a translation from a view mechanism (i.e. a mechanism to define virtual *classes*) to our virtual *object* mechanisms (Section 5). Most work in this field focusses on one aspect only, and we have never found a formal definition of the relationship between the two mechanisms.

Due to a limit on the length of the paper, it has not been possible to include the formalization of both the type rules and the operational semantics of the proposed language mechanisms; it can be found in [3].

Objects with roles have recently been studied by several authors (e.g., [10], [15], [21], [23], [25], [33], [34]); a comparison of these proposals is beyond the scope of this paper but can be found in [23].

View mechanisms for ODBs have also been proposed by several authors (e.g., [12], [8], [9], [2], [27], [16], [18], [22], [24], [26], [30], [28], [17], [35],[19]), and the main approaches are compared in the survey paper [20], together with a discussion of the functionalities needed by a general mechanism for defining views in ODBs. A deep comparison between our and those approaches will be presented once our approach has been defined, in Section 6. We only mention here that most of these approaches focus on the virtual class mechanism, while we study here virtual object operators, as done in [22]; we also study the relationship between the two approaches (Section 5), while this aspect is not dealt with in other works.

The paper is organized as follows. Section 2 defines some basic terminology and introduces the concepts that will be used to explain the goals of the paper. Section 3 briefly presents the object with roles mechanism and its semantics. Section 4 describes the virtual objects mechanism and its semantics. Section 5 outlines the mechanism for defining views, whose semantics is given in terms of the virtual objects operators. Section 6 compares related works. In Section 7 some conclusions are drawn.

2 Concepts, Terminology and Issues

This section establishes some basic terminology and describes informally the notion of objects with roles and virtual objects. We will also introduce the essentials of the Galileo97 language syntax that will be used to give examples in the rest of the paper to make the presentation more concrete.

2.1 Objects and Types

An *object* is a software entity which has an internal state (*instance variables*) equipped with a setof local operations (*methods*) to manipulate that state. The request for an object to execute an operation is called a *message* to which the object can reply.

An object is an instance of a type defined with a *generative type constuctor*, i.e. each object type definition produces a new type, which is different from any other previously defined types. An object type describes the state fields and the implementation of methods of its possible instances. An object type definition introduces a constructor of its instances, and so an object can be constructed only after its object type definition has been given. The signature $\Downarrow\mathcal{T}$ of an object type $\mathcal{T}$ is the set of label-type pairs of the messages which can be sent to its instances.

Each application of an object constructor returns a new object with a different *identity* that persists over time, independently of changes to its state. The equality operator on objects is based on identity. Object identity may be implemented as a hidden field with a system-generated and system-wide unique *object-identifier* (OID). Hereafter we will not talk about OID, but we will assume that equality on objects is OID equality (but see also Section 4.6).

In the object-oriented programming context this approach to objects is called *class-based* since the description of objects is called a *class*; we prefer the term "type" since we will use "class" with a different meaning according to the database tradition.

In Galileo97 a new object type is defined by the operator `<->`[1]. An object type description is defined using the record constructor `[]` and a set of label-type and label-method pairs. Each label is called a "field" or "attribute". A label-type pair, such as `Ide: string`, represents an instance variable (a *state component*), while a label-method pair, such as `Ide:=meth ...`, represents a method. Methods can access the fields of an object using the predefined identifier `self`. Messages are sent to objects using the dot notation.

Components of the state can be declared to be updatable (e.g. `Ide:= var T`); they are updated in an object `O` with the operator `<-` (e.g. `O.Ide <- v`), and their value can be obtained with the operator `at` (e.g. `at O.Ide`). The distinction of constant and updatable attributes in the Galileo97 type system is an important feature to ensure static typechecking.

Components of the state or methods can be defined as `private` to enforce encapsulation, otherwise they are considered `public`, however for the sake of simplicity we will not consider this issue, which is orthogonal to our main subject.

The following is an example of the object type `Person` definition:

```
let rec type Person <->
    [Name: string;
     BirthYear: int;
     Address: var string;
     WhoAreYou:= meth(): string is "My name is  " & self.Name & "." ];
```

The definition of an object type $\mathcal{T}$ introduces the function `mk`$\mathcal{T}$ to construct objects of type $\mathcal{T}$: the function parameter is a record type contaning the label-type pairs of $\mathcal{T}$. For example, the following expression binds the identifier `John` to an object of type `Person`:

[1] Galileo97 syntax is not very different from those of other object oriented database languages, and the query language is very similar to the ODMG-OQL syntax.

```
let John := mkPerson([Name :="John Smith";Address :=var "A street";BirthYear :=1967 ]);
```

2.2 Inheritance

Inheritance is a mechanism which allows something to be defined, typically an object type, by only describing how it differs from a previously defined one. Inheritance should not be confused with subtyping: subtyping is a relation between types such that when $\mathcal{T} \leq \mathcal{S}$, then any operation which can be applied to any value of type $\mathcal{S}$ can also be applied to any value of type $\mathcal{T}$. The two notions are sometimes confused because, in object oriented languages, inheritance is generally only used to define object subtypes, and object subtypes can only be defined by inheritance. However, we will keep the two terms distinct and will use each of them with its proper meaning.

In Galileo97 an object type $\mathcal{T}$ can be defined by inheritance from another object type $\mathcal{T}'$ as follows:

type $\mathcal{T}$ <-> is $\mathcal{T}'$ and $\mathcal{H}$

The type $\mathcal{T}$ *inherits* the $\mathcal{T}'$ attributes, i.e. both its instance variables and methods. Galileo97 allows *strict* inheritance only: an $\mathcal{T}'$ attribute $\mathcal{A}_i$, with type $\mathcal{H}_i$, may be redefined in $\mathcal{T}$ only by specializing its type, that is the new type $\mathcal{H}'_i$ of $\mathcal{A}_i$ must be a subtype of $\mathcal{H}_i$. For this reason, the inheritance mechanism in Galileo97 always produces an object subtype, i.e. in our example we will have $\mathcal{T} \leq \mathcal{T}'$. The following is an example of an object type defined by inheritance:

```
let rec type Student <-> is Person and
    [Code: string;
     Faculty: string;
     WhoAreYou := meth(): string is
                  super.WhoAreYou  & " I am a student of " & self.Faculty ];
```

In an object type defined by inheritance, the special identifer `super` in a new method, such as `WhoAreYou`, is used to invoke the old version of a method `m` from a direct supertype. As usual, any occurrence of `self` within the method `m` is interpreted with respect to the current subtype, and not with respect to the supertype.

Multiple inheritance is generally possible, i.e. inheritance from several supertypes, however for the sake of simplicity we will not consider this issue.

Given a method invocation of the form `O.m(...)`, a language dependent technique is responsible for identifying the appropriate method `m` of the object `O` that has to be executed. Let us consider the following function:

```
let print := fun(x: Person):string is x.WhoAreYou;
```

Let `John` be an object of type `Person` and `Bob` an object of type `Student`. The problem is the meaning of `x.WhoAreYou` in the body of `print` during the invocation of `print(Bob)`. According to a *static dispatch* technique (also called *early binding*), based on compile-time information about `x`, the code of `WhoAreYou` in `Person` is executed. On the other hand, according to a *dynamic dispatch* technique, based on the run-time information about `x`, the code of `WhoAreYou` in `Student` is executed. Dynamic dispatch (also called *late binding*), is found in all object oriented languages, and it is considered to be one of their defining properties.

2.3 Sequences, Classes and Subclasses

An object data model supports a mechanism to define a collection of homogeneous values to model multivalued attributes or collections of objects to model databases. Usually two different mechanisms are provided. To model multivalued attributes, type constructors are available for bags, lists (or sequences),

and sets. For the sake of simplicity we will only consider sequences. To model databases we consider a mechanism called *class*. A class is a modifiable sequence of objects with the same type. A class definition has two different effects: (a) it introduces the definition of the type $\mathcal{T}$ of its elements and a constructor for values of this type (*intensional aspect*), and (b) it supplies a name to denote the modifiable sequence of the elements of type $\mathcal{T}$ currently in the database (*extensional aspect*).

In Galileo97 classes and sequences can be queried using the same operators, similar to those offered by the language OQL. However, classes and sequences differ in that classes are automatically updated every time an object of the corresponding type is created or deleted, while the extent of a sequence is immutable. In addition, classes only can be defined by inheritance and organized into a *subclass* hierarchy, such that if $\mathcal{C}_1$ is a subclass of $\mathcal{C}_2$, then the following properties hold: (a) the type of the elements in $\mathcal{C}_1$ is defined by inheritance from the type of the elements in $\mathcal{C}_2$; (b) the elements in $\mathcal{C}_1$ are a subset of the elements in $\mathcal{C}_2$.

Subtype, inheritance, and subset are three different kinds of partial order relations between types, definitions, and values of an object language. *Subtype* is a relation between types which implies value substitutability; *inheritance* is a relation between definitions, which means that the inheriting definition is specified "by difference" with respect to the super-definition; *subset* is a subset relation between collections of objects, which also implies a subtype relation between the types of their elements.

In Galileo97 `seq` $\mathcal{T}$ is the type of a sequence of elements of type $\mathcal{T}$. Sequences are enclosed in curly brackets and their elements are separated by semicolons. A class definition introduces a name for the class and one for the object type of its elements. In the following example, the class `Persons` is defined, whose members belong to the object type `Person`:

```
let rec Persons class Person <->
    [Name: string;
     BirthYear: int;
     WhoAreYou:= meth(): string is "My name is  " & self.Name & "."];
```

Classes of objects model sets of entities of the observed world, while relationships between such entities are represented as objects that have other objects as components. When an object with the type $\mathcal{T}$ of the elements of a class is constructed by `mk`$\mathcal{T}$ or `in`$\mathcal{T}$, then the object automatically becomes an element in that class. When an object loses the type $\mathcal{T}$ (`drop`$\mathcal{T}$), then it is also removed from the corresponding class. Here is an example of a subclass definition:

```
let rec Students subset of Persons class
Student <-> is Person and
    [Code: string;
     Faculty: string;
     WhoAreYou := meth(): string is
                  super.WhoAreYou & " I am a student of " & self.Faculty ];
```

2.4 Objects with Roles

We call *object extension* the operation which allows an object to assume a new type without changing its identity. This operation is necessary to model the behavior of real world entities. It is also useful in the context of database evolution: when a new subtype is added to an object type hierarchy, it is often useful to make some existing objects acquire the new type.

With the object mechanism described so far, object extension is not allowed. Moreover, an object is always an instance of a single *minimal type*, that is a type $\mathcal{T}$ such that all the other types to which it belongs are supertypes of $\mathcal{T}$. The minimal type of an object is the one that it receives at construction time, and is the one which dictates which method the object uses to answer a message.

When object extension is allowed, it becomes possible for an object to acquire several minimal types, with possibly conflicting state component and method definitions (see Section 3 for an example). This

problem should not be solved by having one minimal type which prevails on the other ones, since in general every context where an object has been extended with a minimal type expects that minimal type to be the prevailing one. For this reason, several authors have proposed to adopt a notion of object with *role* to support objects that can evolve over time by changing type and appearance. In these proposals, an *object role* (or simply a *role*) is one of the perspectives of an object as an instance of a type, and it defines a particular context for method execution. One object may possess one different role for each of its minimal types, or even one for each of its types (as in Galileo97). In any case, whenever an object is extended with a new minimal subtype it acquires a new role. Note that in this approach a new type is only acquired when a new role is acquired and viceversa, hence we do not have two distinct 'type acquisition' and 'role acquisition' operations.

The notion of objects with roles is essential to support object extension, but is also useful to model situations where one real world entity may exhibit a different behavior in different contexts. An analysis of these approaches can be found in [23]. Three important issues in the design of a language that supports objects with roles are the technique to choose the role which will answer a specific message, the method lookup algorithm to adopt, and the semantics of the operator to drop types from an object. They will be discussed in Section 3 in the context of Galileo97.

2.5 Views for ODBs

Relational database systems provide a well-known view mechanism to define derived (or computed) tables as the result of a query on other real (or derived) tables. These derived tables can be queried as if they were real tables, but updating a derived table is generally not allowed, since it may not be possible to map the operation onto the real tables.

A similar functionality may be provided for object databases. For example, the following Galileo97 expression defines a sequence containing all the young persons currently present in the `Person` class:

```
let AllYoungPersons := derived  Persons where BirthYear > 1985;
```

`AllYoungPersons` is a computed value of type `seq Person`. A `derived` sequence (a) is computed every time it is used, (b) can be queried like any other sequence or class, (c) its elements can be updated in the same way as objects of a class can be updated and the effects are the same, and (d) its elements change if objects are added or removed from the class `Persons`, as one would expect. This is the only way to change the number of elements of a derived sequence.

The `AllYoungPersons` example shows that a derived sequence is adequate for representing a view whose elements are a subset of a class, while problems arise when one needs to change the structure of the class elements in the view. The only way to achieve the effect with the Galileo97 operators seen so far is to build a new value starting from a class element: it may be either a record (using the record constructor []) or a new object which does not have the same identity as the original one. Using the terminology proposed in [31], the first kind of view is called a *relational semantics view*, and the second an *object-generating semantics view*. For example:

```
let ItalianRecordPersons := derived
       select [ Nome:= Name; AnnoNascita:= BirthYear]
       from Persons;

let type Italian <-> [ Nome: string; AnnoNascita: int];

let ItalianObjectPersons := derived
       select mkItalian([Nome:= Name; AnnoNascita:= BirthYear])
       from Persons;
```

However, two things are lacking:

- the ability to define *object-preserving semantics views*, i.e. views whose elements are virtual objects with the same identity as the corresponding base object;
- the ability to place the defined view into a subset hierarchy.

To solve the first problem, an object language should have an identity preserving algebra on objects which allows one to define different interfaces for the same objects (called *virtual objects*), so that different users may see the same objects with a different structure and behavior [31].

To solve the second problem, the language should have a *virtual class* mechanism, to allow the programmer to define collections which are: (a) virtual, i.e. computed starting from a base collection, as in our examples above; (b) classes, i.e. which are placed into the subset hierarchy and which collect all existing values of their associated type.

Object algebras have been proposed by several authors (e.g., [32] [26] [29] [16] [19] [24]), and the basic algebraic operators are *selection* to define subsets of a class, *projection* to see a subset of the object attributes or methods, and *extension* to add new attributes and methods. Some of these papers combine the virtual objects and virtual classes mechanisms, while in this paper they are studied separately, and we show how the virtual objects operators can be used to give the semantics of a virtual class mechanism.

There are particular problems which must be addressed in order to include a set of object algebraic operators into a statically and strongly typed object-oriented programming language:

- What is the type of a virtual object, and how is it related to the type of the object it is based on?
- When the mechanism allows new methods to be added to objects, can the implementation of these methods have access to the state of the base objects? If a method can be overriden in the virtual object, what impact is there on the method lookup algorithm?
- What are the differences between the role mechanism and the virtual object mechanism?
- If a join operator is supported by the object algebra to combine two objects, what is the identity of the resulting objects?

A solution to these problems will be given in the context of Galileo97 in Section 4, and a comparison with other proposals will be presented later in Section 6.

3 Objects with Roles in Galileo 97

In Galileo97, besides the operator `mk`$\mathcal{T}$ to construct objects of type $\mathcal{T}$, the operator `in`S exists to extend dynamically an object with a new subtype S of $\mathcal{T}$, without changing its identity, but with the possibility of changing its behavior. The operator `in`S adds a new *role* to the object, and returns a reference to this new role of that object. In Galileo97 an object expression always denotes one specific role of an object.

The following example shows the definition of the type `Person`, with a method `WhoAreYou`, and two subtypes `Student` and `Athlete` defined by inheritance that have a `code` field with a different type. Such a definition is allowed in any object-oriented language, since the validity of an object type definition only depends on its object supertypes, but cannot be limited by the definition of its *cousin* (i.e. neither descendents nor ancestors) types:

```
let rec type Person <->
    [Name: string;
     BirthYear: int;
     WhoAreYou:= meth(): string is "My name is  " & self.Name & "."];
```

```
let rec type Student <-> is Person and
    [Code: string;
     Faculty: string;
     WhoAreYou := meth(): string is
                  super.WhoAreYou  & " I am a " & self.Faculty & "student" ];

let rec type Athlete <-> is Person and
    [Code: int;
     Sport: string;
     WhoAreYou:= meth(): string is
                 super.WhoAreYou & " I play " & self.Sport ];
```

The following expression builds an object, with one role only, of type `Person`:

```
let John := mkPerson ([ Name := "John Smith"; BirthYear := 1967 ]);
```

The answer to the message `John.WhoAreYou` is `"My name is John Smith"`. The object with role `John` can now be extended with the subtype `Athlete` as follows:

```
let JohnAsAthlete := inAthlete(John, [ Code := 245; Sport := "tennis" ]);
```

As a consequence of this extension, we now have two different ways to access the same object. `John` refers to the `Person` role while `JohnAsAthlete` refers to the `Athlete` role of the same object. As a consequence, method lookup generally gives different results, depending on the role used.

As a consequence of this extension, `John` now has two different methods to answer the `WhoAreYou` message, and Galileo 97 allows both of them to be accessed, using the different notations `John.WhoAreYou` and `John!WhoAreYou`.

In both cases method lookup starts from the `Person` role of `John`. With the "." notation, the method is first looked for in the subtypes of `Person` (*downward lookup* phase). If this phase fails, the method is looked for in `Person` and in its supertypes (*upward lookup* phase). This whole process is called *double lookup*, and finds the `Athlete` implementation of `WhoAreYou`. With the "!" notation, only upward lookup is performed, thus finding the `Person` implementation of `WhoAreYou`. If the method is found in a supertype, `self` stands for the role that receives the message, otherwise `self` stands for the role that answers the message.

Both techniques are statically guaranteed to find a method of the right type, and both are instances of the late binding mechanism, since they do not depend on the static type of the receiver, but on its dynamic types and on the role through which it is accessed (this is not evident in this case, since the `Person` role of `John` corresponds to its static type; this is not the case in our next example).

Another, more important, consequence of the extension is that now we have two different roles to access the same object, which behave differently. For example if the `WhoAreYou` message is sent with the upward lookup notation, the two roles `John` and `JohnAsAthlete` answer in two different ways.

The object with role `John` can also be extended with the type `Student`:

```
let JohnAsStudent := inStudent(John, [ Code := "0123"; Faculty := "Science" ]);
```

We say that `John`, `JohnAsStudent` and `JohnAsAthlete` are three different roles of the same object, of type `Person`, `Student` and `Athlete`, respectively.

This extension has the following effects that show how in Galileo 97 messages can be addressed to every role of an object, and the answer may depend on the role addressed:

- the answer to the message `Code` sent to `JohnAsStudent` is a string, while the answer to the same message sent to `JohnAsAthlete` is an integer;

- the answer to the message `WhoAreYou` sent to `JohnAsStudent` is `"My name is John Smith. I am a Science student"`, while the answer to the same message sent to `JohnAsAthlete` is `"My name is John Smith. I play tennis"`;
- the answer to the message `WhoAreYou` sent to `John` with the dot notation changes and becomes `"My name is John Smith. I am a Science student"`.

While upward and double lookup are two different forms of dynamic binding, static binding to the method of type $\mathcal{T}$ can be obtained through the `(O As T)!msg` idiom, where `O As T` is the operator which allows one to access the role with type $\mathcal{T}$ of an object `O`. Let us consider the following function:

```
let foo := fun(x:Person): seq string is
           {x.WhoAreYou;  x!WhoAreYou; (x As Person)!WhoAreYou};
```

Let `JohnAsStudent` be bound to a value of type `Student`, which has then been extended with a role of type `ForeignStudent`, subtype of `Student` which redefines the method `WhoAreYou`. The value returned by `foo(JohnAsStudent)` is a sequence of three answers produced by the method defined in type `Foreign-Student` (dynamic binding with double lookup), by the method defined in type `Student` (dynamic binding with upward lookup), and by the method defined in type `Person` (static binding).

The language also provides other operators on objects and roles which allow one to discover which roles can be played by an object (*RoleExpr* `isalso` $\mathcal{T}$), and to remove the role with type $\mathcal{T}$ and its subroles from an object (`drop`$\mathcal{T}$`(`*ObjExpr*`)`). If a role of type $\mathcal{T}$ is an element of a class, when an object loses the type $\mathcal{T}$ (`drop`$\mathcal{T}$), then it is also removed from the corresponding class.

3.1 A Storage Model for Objects with Roles

This section presents a simple storage model for objects in order to give an informal semantics of objects with roles. For a discussion of extensions to this model toward a more realistic case see [6]. A formal semantics can be found in [3].

The behavior of objects of standard class-based languages can be explained in terms of the simple storage model of Figure 1 [1]. In this model, an object is represented as a record which contains the fields of the object state and its methods. The special identifier `self` in a method `m` is always bound to the object that, according to this simple storage model, contains the method `m`. Usually the storage model is more complex than this one, and methods are not embedded into objects but are factored into the object type descriptor and shared by the objects of the same type. Even though a more complex storage model would be justified for reason of efficiency, the behavior of objects can be explained by this simple model.

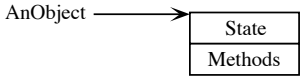

Figure 1: A simple storage model for objects.

Figure 2 shows how this simple storage model for objects might be modified to deal with roles. In this model there are two structures, the *object history*, and the *role structure*.

The object history is used to store the history of the roles that an object has acquired. This structure is a sequence of pairs (type identifier, reference to the corresponding role structure), one pair for each type acquired by the object. The set of pairs is in temporal order. The object history is used (a) for method lookup, (b) to implement the operators `As` and `isalso`, (c) to find all the roles associated with a subtype of `T` which must become invalid when the operator `dropT` is executed, and (d) to implement object identities (e.g. two objects are the same when their object history is the same).

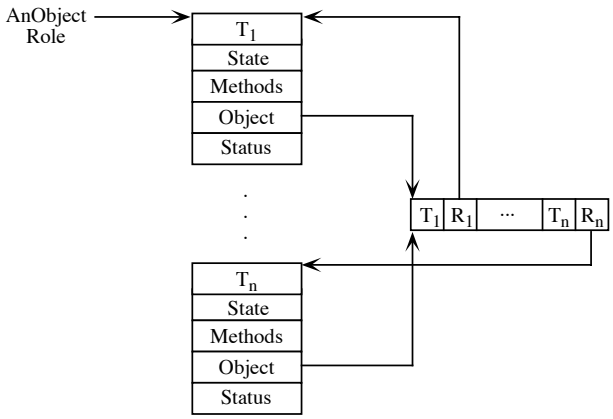

Figure 2: A simple storage model for objects with roles.

Each role structure stores information about the methods and the fields which are defined for the first time, or redefined, in the corresponding role type. In particular, a role structure contains the following information:

- the role type `T`$_i$;
- a status which is `valid` if the object has the role with type `T`$_i$, and is `removed` if the object has lost that role;
- a reference to the object history;
- the methods and the fields defined for the first time, or redefined, in the corresponding role type.

When a new object is created, a new role structure is created, and a reference to it is returned. When an object is extended with a new role type, a new role structure is created and is connected to the object, and a reference to this new structure is returned; an object is not allowed to be extended with a role type `T`, if `T` is already in the object history.

When an object is created with a subtype `T`$_j$ of the root type `T`$_i$, the effect is the same as the creation of the object with the root type `T`$_i$, and then its extension with the subtype `T`$_j$. For example, the object in Figure 3 can be obtained by creating a `Person` and then by extending it with the type `Student`, or by creating directly a `Student`. The difference between the two cases is that, in the former the system returns a reference to a role structure of type `Person` and then a reference to a role structure of type `Student`, while in the latter the system returns only a reference to a role structure of type `Student`, and the access to the other role is possible only by using the operator `As`.

When an object loses the type `T`$_i$, the following actions are executed:

1. the status of the role structure `R` with type `T`$_i$ becomes `removed`;
2. the pair `T`$_i$,`R` is removed from the object history;
3. steps 1 and 2 are repeated for every role in the object history whose type is a subtype of `T`$_i$.

When a message is sent to a role `r`, its status is first checked in the corresponding structure, and then its run-time type `RT` is read and the object history is used to retrieve the role structures where the method should be looked for.

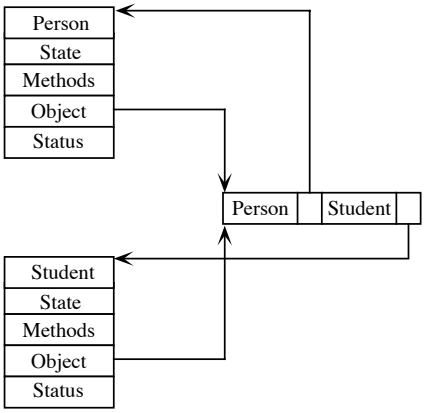

Figure 3: The structure of an object with role types `Person` and `Student`.

If the status of `r` is `valid`, downward lookup is performed by scanning the history from the last acquired role back to the `RT` role, looking for those roles whose type is a subtype of `RT` (the subroles of `r`). Upward lookup is then performed, by scanning the history from the `RT` role back to the first role, looking for the superroles of `r`.

If the status of `r` is `removed`, a failure is raised only if the object referred by the role `r` has lost the role corresponding to the static type `ST` of `r`. Otherwise, when `r` denotes a dropped role of an object which still has the role with type `ST`, method lookup is still performed, but only the non-dropped roles are considered. To implement this semantics, every message has one extra parameter, the static type `ST` assigned by the compiler to `r`, and a failure is raised if no `ST` role is found in the object history. If the `ST` role exists, then upward lookup is performed by scanning the history from the last role back to the first role, only considering the superroles of `r`. The presence of the `ST` role, and of its superroles, guarantees that a method with a suitable type will be found.

4 Virtual Objects

The operators on objects presented so far allow objects to be built, and roles to be added and dropped without affecting object identity. These operators allow one to model the dynamic behavior of real-world entities, and are also useful for dealing with the most common kind of schema evolution, i.e. attribute addition or specialization. In this situation, in fact, it is possible to introduce a new subtype of the old type and to extend the old values with the new information. The type correctness of the preexisting applications and data structures will not be affected, and it is even possible to decide to partially modify the behavior of an application by specializing the behavior of some methods.[2]

However, the role mechanism cannot cope with the related problem of giving different views of the same object without affecting its behavior. This is because object extension actually modifies the object, and object extension can only modify an object in a very limited way (only field addition and specialization). If we call "real objects" those that have been explicitly constructed using the `mk` or `in` operators, what is needed is the possibility to define "virtual objects" by starting with objects and changing their interface while preserving their identity.

The virtual object mechanism we are going to describe has the following features:

- virtual objects are first class, statically and strongly typed values;

[2]More precisely, a preexisting application is affected by method specialization only if the application exploits the double lookup ($obj.msg$) form of message passing.

- a virtual object has the same identity as the object it is based on; when it is based on the combination of several objects, then its identity is a combination of the identities of the objects it is based on;
- a virtual object can add, remove, and rename fields of its base object; moreover a virtual object can have its own instance variables, which are accessed by its own methods;
- a virtual object can be based on more than one real object;
- the behavior of a real object is not affected by the existence of a related virtual object;
- a virtual object can be used exactly like a real object and vice versa, at least as long as the type rules described later on are satisfied;
- virtual objects allow one to update those components of the state of the real object which the virtual object allows to be view;
- a message to a virtual object to execute a method imported from an object returns an answer which is the same as if the message were directly sent to the real object.

We will now show how virtual objects are defined, and discuss their types and how these types are related to object and record types. Our approach is based on a set of basic operators that can be applied to objects, real or virtual, in order to produce other objects (virtual) with an identity preserving semantics. These operators can also be applied to records, which will be seen as a special case of virtual objects, although we will not stress this point. These object operators are then extended to collections of objects to define collections of virtual objects which are similar to views built over relations.

As we have already pointed out, similar operators for collections of objects have been proposed by other authors, and the main contribution of our approach is to introduce these operators at an instance level in the framework of statically and strongly typed programming language, to clarify the interactions between virtual objects and objects with roles, and to show the different semantics of method overriding and evaluation in virtual objects and objects with roles.

4.1 Virtual Object Types

A virtual object role, which for the sake of simplicity we will sometimes call virtual object or virtual role, can be seen as a pair formed by the base object role and a mapping which may hide, rename, or even add some fields (even state components) with respect to the original object role (as described in the storage model in Section 4.3). More precisely, a virtual object role can generally be based on a set of base objects, with a mapping which manipulates their components and gives the external impression of a unique virtual entity.

Virtual object roles are typed with the `view` type constructor:

`<`$\mathcal{T}_1$`,...,`$\mathcal{T}_m$`> view [`$\mathtt{A}_1$`:`$\mathcal{S}_1$`;...;`$\mathtt{A}_n$`:`$\mathcal{S}_n$`]`

where $\mathcal{T}_1, \ldots, \mathcal{T}_m$ are role types; $\mathtt{A}_1, \ldots, \mathtt{A}_n$ are labels; $\mathcal{S}_i$ are types. As a syntactic abbreviation, the type $\mathcal{S}_i$ of a label $\mathtt{A}_i$ can be omitted and it is assumed that it is the type of label $\mathtt{A}_i$ in the first type, among $\mathcal{T}_1, \ldots, \mathcal{T}_m$, where $\mathtt{A}_i$ appears. Intuitively, the statement:

`O: <`$\mathcal{T}_1$`,...,`$\mathcal{T}_m$`> view [`$\mathtt{A}_1$`:`$\mathcal{S}_1$`;...;`$\mathtt{A}_n$`:`$\mathcal{S}_n$`]`

means that `O` is a virtual object based on m object roles with types $\mathcal{T}_1, \ldots, \mathcal{T}_m$, and with signature `[`$\mathtt{A}_1$`:`$\mathcal{S}_1$`;...;`$\mathtt{A}_n$`:`$\mathcal{S}_n$`]`. For any object role type $\mathcal{T}$ with signature $\Downarrow\mathcal{T}$, the following type equivalence holds:

`<`$\mathcal{T}$`> view [`$\Downarrow\mathcal{T}$`]` $\sim \mathcal{T}$

where $\mathcal{T} \sim \mathcal{S}$ means that $\mathcal{T} \leq \mathcal{S}$ and $\mathcal{S} \leq \mathcal{T}$, i.e. that values of each type can be considered as if they were values of the other one. Similarly, for every record type `[`A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]` the following type equivalence holds:

`<> view [`A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]` $\sim$ `[`A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]`

The subtype relationship among real and virtual object types is defined in the next section.

Type hierarchies with `view` types

A subtype relationship $\mathcal{T}' \leq \mathcal{T}$ means that any operation which can be applied to any value of type $\mathcal{T}$ can also be applied to any value of type $\mathcal{T}'$. A virtual object `O` with type:

$\mathcal{T}$`= <`$\mathcal{T}_1$`,...,`$\mathcal{T}_m$`> view [ `A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]`

accepts two kinds of operations: message passing (`O.`A_i, `O!`A_i), and object extraction (`O As` $\mathcal{T}_j$). Hence, a type $\mathcal{T}'$ is a subtype of $\mathcal{T}$ if it contains enough object identities and components to be able to deal with every object extraction and message passing operation which is supported by $\mathcal{T}$. More precisely:

`<`$\mathcal{T}'_1$`,...,`$\mathcal{T}'_m$`> view [ `A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]`
$\leq$ `<`$\mathcal{T}_1$`,...,`$\mathcal{T}_r$`> view [ `B_1`:`$\mathcal{R}_1$`;...;`B_s`:`$\mathcal{R}_s$`]`

if (a) for each $\mathcal{T}_i$ there exists a $\mathcal{T}'_j$ such that $\mathcal{T}'_j \leq \mathcal{T}_i$ and (b) for each pair `B`$_i$`:`$\mathcal{R}_i$ there is a pair `A`$_j$`:`$\mathcal{S}_j$ such that `A`$_j$ = `B`$_i$ and $\mathcal{S}_j \leq \mathcal{R}_i$.

The principle that an object view type specifies the object extraction and message passing operations which can be applied to a type, implies that the following two equivalences also hold:

$\mathcal{T}$ $\sim$ `<`$\mathcal{T}$`> view [` $\Downarrow\mathcal{T}$`]`
`[` A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]` $\sim$ `<> view [` A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]`

4.2 Virtual Object Constructors

The operators to build virtual objects are: `project`, `rename`, `extend` and `times`. These operators can be applied to sequences of objects using the notation `project*`, `rename*`, `extend*` and `times*`.

Project

`project` is used to hide properties from an object role.

`O project [`A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]`

returns the object `O` with components $A_1,\ldots,A_n$ only. More precisely, if `O` is an object role with type

`<`$\mathcal{T}_1$`,...,`$\mathcal{T}_m$`> view [ `A_1`:`$\mathcal{S}'_1$`;...;`A_n`:`$\mathcal{S}'_n$`; `B_1`:`$\mathcal{R}_1$`;...;`B_l`:`$\mathcal{R}_l$`]`,

and if each $\mathcal{S}'_i$ is a subtype of $\mathcal{S}_i$, then `O project [` A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]` returns the same object role `O` seen through type

`<`$\mathcal{T}_1$`,...,`$\mathcal{T}_m$`> view [` A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]`.

The type $\mathcal{S}_i$ of a label A_i can be omitted, and it is then assumed to be the type $\mathcal{T}_i$ of A_i in `O`.

Note that although `project` is formally defined on virtual objects alone, it can also be applied to real objects too, thanks to the type equivalence

<$\mathcal{T}$> view [$\Downarrow\mathcal{T}$] $\sim \mathcal{T}$.

The same observation holds for all the other virtual object operators that will be presented, which can all be applied in the same way to real and virtual objects to produce virtual objects.

Example 1 *Let us consider the following definitions:*

```
let type AnAddress :=
    [Street: var string;
     City: var string;
     Country: var string ];

let rec Persons class Person <->
    [Name: string;
     Income: var int;
     Address: AnAddress;
     BirthYear: int;
     Parents: [ Father: Person;
                Mother: Person ];
     WhoAreYou:= meth(): string is "My name is  " & self.Name & "." ];

let rec Employees subset of Persons class
Employee <-> is Person and
    [Salary: var int;
     Company: Company;
     WhoAreYou:= meth(): string is
                 super.WhoAreYou & " I work with company " & self.Company.Name ]
and
Companies class Company <->
    [Name: string;
     Location: string;
     Revenue: int ];

let type PersonView :=
         <Person> view [ Name;
                         Address: [ Street: var string;
                                    City: var string ];
                         WhoAreYou ];
```

Let `John` *be an object role of type* `Person`, `JohnView` *a virtual role of type* `PersonView`, `Foo` *and* `Goo` *two functions defined as follows:*

```
let JohnView :=
      John project [ Name;
                     Address: [ Street: var string;
                                City: var string ] ;
                     WhoAreYou] ;
let Foo := fun(x: Person) :Person is x;
let Goo := fun(x: PersonView):int is
              if x isalso Employee
                 then at (x As Employee).Salary
                 else at (x As Person).Income;
```

The following considerations apply:

- *By projecting* `Address` *to the indicated supertype of the original type,* `Country` *component is hidden in* `JohnView`*;*
- `John = JohnView` *returns true since* `project` *is an identity preserving operator and* `JohnView` *is defined starting from* `John`*;*
- `John.WhoAreYou = JohnView.WhoAreYou` *returns true since the method executed to answer the message* `WhoAreYou` *sent to* `JohnView` *is the method defined for* `John`*;*
- *the component* `Street` *of the* `Address` *of* `JohnView` *can be updated and this will also effect* `John`*. Likewise, an update of* `John` *will have the same effect on* `JohnView`*;*
- *as will be explained in Section 4.1,* `Person` *is a subtype of* `PersonView`*, but not vice versa, hence* `Goo(John)` *would be typed, while* `Foo(JohnView)` *would not;*
- *let us assume that* `John` *has been extended with the role type* `Employee`*. The following virtual object:*

```
let JohnEmplView :=
      (John As Employee)
      project [ Name;
                Company: [ Name: string; Location: string ] ;
                WhoAreYou] ;
```

has a type which is a supertype of `Employee`*, but it is not comparable with either* `Person` *or* `PersonView`*.*

- *The following example shows the definition of a derived sequence:*

```
let EmployeesView := derived
      Employees project* [ Name;
                           Company: [ Name: string; Location: string ] ;
                           WhoAreYou] ;
```

The elements of `EmployeesView` *are those of* `Employees` *with a supertype of* `Employee`*.*

□

Extend

`extend` is used to add or redefine methods and fields to an object.

`O extend [`A_1`:`$\mathcal{S}_1$ `:= Expr`$_1$`;...;`A_n`:`$\mathcal{S}_n$ `:= Expr`$_n$ `]`

returns the object `O` extended with new fields whose value is specified by the expressions $Expr_1,\ldots,Expr_n$. If a label A_i was already present in `O`, `extend` overrides A_i with a value of a possibly unrelated type. More precisely, if `O` has type

`<`$\mathcal{T}_1,\ldots,\mathcal{T}_m$`> view [ `$B_1$`:`$\mathcal{R}_1$`;...;`$B_l$`:`$\mathcal{R}_l$`;`$A_1$`:`$\mathcal{S}_1$`;...;`$A_k$`:`$\mathcal{S}_k$`]`

with $k \leq n$, then the extension expression above has the type:

`<`$\mathcal{T}_1,\ldots,\mathcal{T}_m$`> view [ `$B_1$`:`$\mathcal{R}_1$`;...;`$B_l$`:`$\mathcal{R}_l$`;`$A_1$`:`$\mathcal{S}_1$`;...;`$A_n$`:`$\mathcal{S}_n$`] .`

Note that this rule allows one to extend both real and virtual objects.

`extend` can be used to add both new fields and new methods, which belong to the virtual part of the object. To this end, the expression associated with a label may contain the pseudo-variable `me`, which denotes, recursively, the whole virtual object *after* it has been extended. This `me` variable can be used much in the same way as `self` in the real object, but there is a difference.

When a method defined in a role $\mathcal{R}$ but activated by inheritance by a message sent to a subrole $\mathcal{S}$, sends a message `msg` to `self`, then method lookup for `msg` starts from role $\mathcal{S}$. When any expression, message passing included, is applied to `me`, then `me` denotes the virtual object that has been created by the `extend` operation it is bound to, hence method lookup for `me.msg` starts from the virtual object where the method invoking `me.msg` is found. Technically, we say that `self` is *dynamically bound* to the role that receives the message, while `me` is *statically bound* to the virtual object that is created by the `extend` expression which `me` is bound to. The reason for this difference, which is discussed in Section 4.5, is that there is no guarantee that the virtual object which receives the message belongs to a subtype of the virtual object where the method is defined, and is connected with the requirement that the behavior of a real object must not be affected by the existence of a related virtual object.

Example 2 *The following example shows how to redefine the structure and behavior of a person object; note that, in the definition of fields* `Mother` *and* `Father`*, using* `me` *or* `John` *makes no difference, while* `me` *is essential to access field* `Age` *inside* `WhoAreYou`*.*

```
let rec AnotherJohnView :=
  (John extend [ Age := meth():int is
                        CurrentDate().Year - me.BirthYear;
                 Mother := meth():Person is me.Parents.Mother;
                 Father := meth():Person is me.Parents.Father;
                 WhoAreYou := meth(): string is
                              (me As Person)!WhoAreYou &
                               " I am " & stringofint(me.Age) & " years  old"]
   ) project [ Name; Age; Mother; Father; WhoAreYou] ;
```

□

Example 3 *The following example shows how to redefine the behavior of the person object* `John` *without changing its type so that it can be used as a parameter of the function* `Foo`*:*

```
let rec JohnForItalians := John extend
                              [WhoAreYou := meth(): string is
                                            "Mi chiamo  " & me.Name & "." ];
```

□

Rename

`rename` is used to change the name of the properties of an object. If `O` has the labels $\mathtt{A}_1,\ldots,\mathtt{A}_n,\mathtt{B}_1,\ldots,\mathtt{B}_l$,

`O rename (`$\mathtt{A}_1$ `=> A'`$_1$`; ...;` $\mathtt{A}_n$ `=> A'`$_n$`)`

returns `O` with the labels `A'`$_1$, ..., `A'`$_n$, $\mathtt{B}_1$, ..., $\mathtt{B}_l$, which must all be different.[3].

More precisely, if `O` has the type:

`<`$\mathcal{T}_1,\ldots,\mathcal{T}_m$`> view [` $\mathtt{B}_1$`:`$\mathcal{R}_1$`; ...;` $\mathtt{B}_l$`:`$\mathcal{R}_l$`;` $\mathtt{A}_1$`:`$\mathcal{S}_1$`; ...;` $\mathtt{A}_n$`:`$\mathcal{S}_n$`]`

[3]Every $\mathtt{A}_i$ may also be a path $\mathtt{A}_i^1.\mathtt{A}_i^2.\ldots.\mathtt{A}_i^{n_i}$ [3]

and A'$_1$, ..., A'$_n$, B$_1$, ..., B$_l$, contains no duplicate, then the renaming expression above has type:

<$\mathcal{T}_1$,...,$\mathcal{T}_m$> view [B$_1$:$\mathcal{R}_1$; ...; B$_l$:$\mathcal{R}_l$; A'$_1$:$\mathcal{S}_1$; ...; A'$_n$:$\mathcal{S}_n$] .

Example 4 *The following definition gives a view of the person object* John *for Italian users:*

```
let JohnViewForItalians :=
        John rename (Name => Nome)
             extend [ Presentati := meth(): string is
                                    "Mi chiamo " & me.Nome "."]
             project [ Nome; Presentati] ;
```

□

It may appear that **rename** is an operator that can be defined in terms of **extend** and **project**; for example; let O be an object with an attribute A$_i$:

```
let r := O rename ( Ai => Ai' );

let r' := (O extend [  Ai' := meth() :Ti is me.Ai ] )
              project [  <all attributes except Ai> ]
```

The two expressions above actually have a different meaning, since both r'.A'$_i$ and r'!A'$_i$ are equivalent to O.A$_i$, while r.A'$_i$ and r!A'$_i$ are generally different since r keeps upward and double lookup distinct.

Times

times is used to create a virtual object by starting from two objects whose component names are all different.

```
O times O'
```

returns an object which contains the identities of both O and O', and has the fields of O and O'. More precisely, if:

O: <$\mathcal{T}_1$,...,$\mathcal{T}_m$> view [A$_1$:$\mathcal{S}_1$;...;A$_n$:$\mathcal{S}_n$]
O': <$\mathcal{T}'_1$,...,$\mathcal{T}'_l$> view [B$_1$:$\mathcal{R}_1$;...;B$_k$:$\mathcal{R}_k$]

then O times O' has type:

<$\mathcal{T}_1$,...,$\mathcal{T}_m$,$\mathcal{T}'_1$,...,$\mathcal{T}'_l$> view [A$_1$:$\mathcal{S}_1$;...;A$_n$:$\mathcal{S}_n$;B$_1$:$\mathcal{R}_1$;...;B$_k$:$\mathcal{R}_k$] .

For example, if $\mathcal{T}$ and $\mathcal{T}$' are two real object role types with different attributes, then

O times O': <$\mathcal{T}$,$\mathcal{T}$'> view [$\Downarrow\mathcal{T}$;$\Downarrow\mathcal{T}$'] .

This type indicates that the virtual object (a) can answer all the $\Downarrow\mathcal{T}$, $\Downarrow\mathcal{T}$' messages, and (b) contains both a $\mathcal{T}$ and a $\mathcal{T}$' object role, which can be recovered with the obj As $\mathcal{T}$ operator previously defined.

The **times** operator has been introduced to complete the object algebra, and to give a semantics to the **from** clause of a query that includes several collections of objects. **times** may appear similar to the **extend** operator, and as a matter of fact in many cases **extend** is sufficient, but there are two differences: **extend** is used to add attributes to a base object, or to redefine some of them, and the result is a virtual object that preserves the identity of the base object; **times** combines two base objects with different attributes, and the result is a virtual object that preserves the identity of both the base objects.

Example 5 *The following view, built on the* `Employees` *and* `Companies` *classes in Example 1, defines a class of employees who are combined with their company.*

```
let EmployeesAndCompanies := derived
  select Emp times (Emp.Company rename (Name => CompName))
  from   Emp In Employees;
```

The result has the type: `seq <Employee,Company> view [Name; Income; Address; BirthYear; Parents; WhoAreYou; Salary; Company; CompName; Location; Revenue]`. *If* `John` *is an element of* `Employees-AndCompanies`, *then* `John As Company` *returns the company where* `John As Employee` *works.*

□

4.3 A Storage Model for Virtual Objects

The behavior of virtual objects can be explained in terms of the simple storage model in Figure 4. In this model, a virtual object is an interface adaptor for a virtual or real object. When a virtual object is created from an object `O`, a reference is returned to a structure whose shape depends on the operator which has been used as follows :

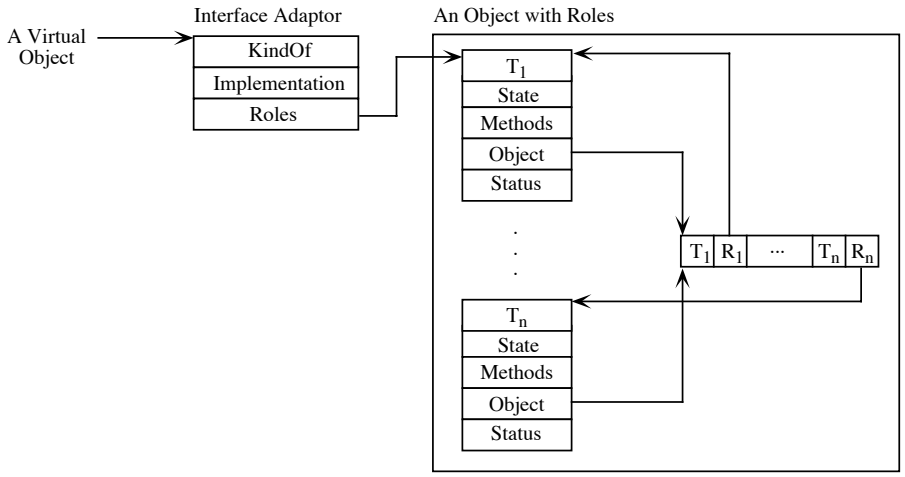

Figure 4: The structure of a virtual object.

- `O project [A`$_1$`:`$\mathcal{S}_1$`; ...; A`$_n$`:`$\mathcal{S}_n$`]` has no run-time effects and returns a reference to `O` with a new type which hides the components different from `[A`$_1$`; ...; A`$_n$`]` ;
- `O extend [A`$_1$`:`$\mathcal{S}_1$ `:= Expr`$_1$`; ...; A`$_n$`:`$\mathcal{S}_n$ `:= Expr`$_n$`]` returns a reference to an interface adaptor of kind **`extend`** which contains the new methods and fields `A`$_1$`; ...; A`$_n$ defined in the virtual object, and a reference to the object `O` (Figure 5); the special identifier **`me`** in a method is recursively bound to the interface adaptor itself;
- `O rename (A`$_1$ `=> A'`$_1$`; ...; A`$_n$ `=> A'`$_n$`)` returns a reference to an interface adaptor of kind **`rename`** which contains a table that maps the attribute and method names `A'`$_i$ in the virtual object to the original names `A`$_i$ in `O`, and a reference to the object `O` (Figure 6);
- `O times O'` returns a reference to an interface adaptor of kind **`times`** which contains a table that associates every attribute with the object where the attribute should be looked for, and two references to the objects `O` and `O'` used to define the virtual object (Figure 7).

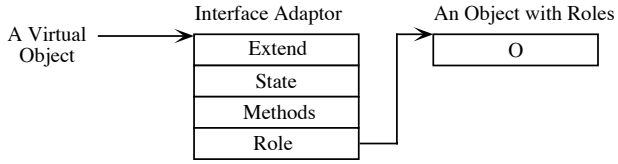

Figure 5: The structure of a virtual object of kind `extend`.

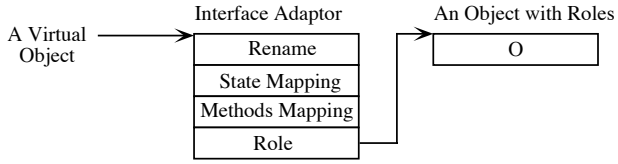

Figure 6: The structure of a virtual object of kind `rename`.

When a virtual object receives a message `m`, it tries to resolve it using its interface adaptor, otherwise it sends the message to the object used to define the virtual object.

Figure 8 shows an example of a virtual object `VirtualJohn` defined from the object role `John` as follows:

```
let VirtualJohn := John project [Name]
                       rename (Name => Surname)
                       extend[NewAttribute := ...;
                              NewMethod := meth() ... ]
```

In a more realistic storage model, chains of interface adaptors would be avoided, and every interface adaptor of a virtual object would be a combination of those used to explain the effect of each operator, as shown in Figure 9.

4.4 The Semantics of `As`

The `As` operator allows one to navigate among the roles of an object and to recover the real object from a virtual one. An expression `O As` $\mathcal{T}$, where `O: <`$\mathcal{T}_1, \ldots, \mathcal{T}_n$`> view [...]`, is well typed iff (a) $\mathcal{T}$ is a role type and (b) if at least one of the $\mathcal{T}_i$ has a common object supertype with $\mathcal{T}$. In this case, the `As` operator searches to see whether one of the real objects on which the virtual object is based has the $\mathcal{T}$ role. If this is case, the $\mathcal{T}$ role of that object is returned; otherwise, a failure arises. The `As` operator raises two issues:

- protection: in some situations, recovering the real object inside the virtual object should be disallowed;

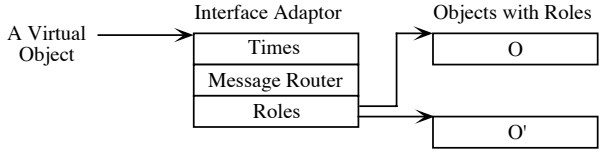

Figure 7: The structure of a virtual object of kind `times`.

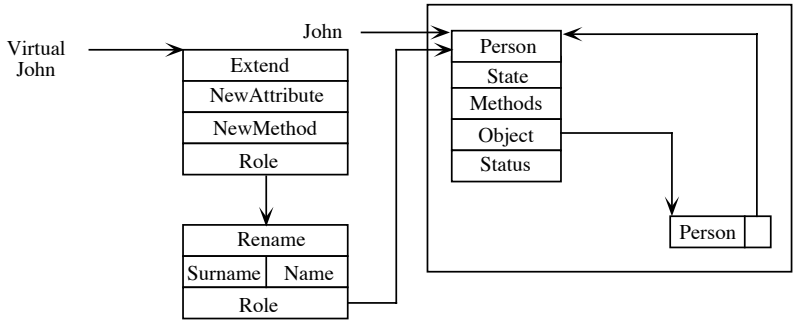

Figure 8: An example of a virtual object.

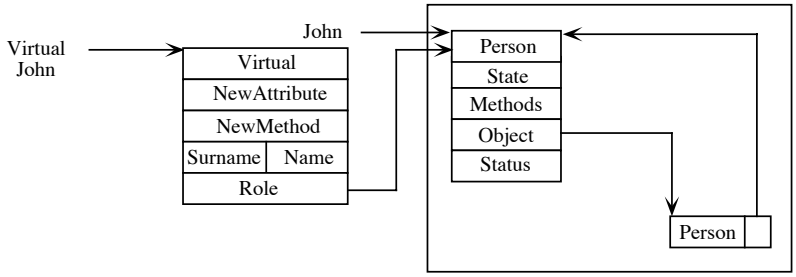

Figure 9: Another example of a virtual object.

- semantics: when a virtual object contains more than one real object with a $\mathcal{T}$ role, the `As` $\mathcal{T}$ operator must choose just one of them.

The protection issue can be dealt with by hiding a role type $\mathcal{T}$ in some sections of the code. Consequently, in those sections the system will not accept an `As` $\mathcal{T}$ operation.

As for the semantics, consider the following code:

```
let John := mkPerson...;
let Peter := mkPerson... rename ...;
let JP := John times Peter;
```

It is not clear whether `JP As Person` should be equal to `John` or to `Peter`. In Galileo97 semantics, it is equal to `John`. In fact, in Galileo97, the real object roles hidden in a virtual object are ordered. Specifically, in the virtual object `(O1 times O2)`, all real object roles in `O1` come before those in `O2`; the `As` operator respects this order while looking for the result. A reasonable alternative would be a non-deterministic semantics, which does not specify in which order the real objects are considered. This semantics would make the `times` operator commutative, and would leave more freedom to Galileo97 implementors. In any case, this kind of semantics ambiguity is not very likely to arise in a real application.

4.5 The Semantics of `self` and `me`

Suppose that a message `m` is sent to a role of type `T`, and that the lookup mechanism activates the method `M` defined for `m` in a supertype `T'` of `T`. Now, if `M` sends a message `p` to `self`, should the search for the corresponding method start from the originally receiving role `T`, or from the role `T'` where the method `M` has been found? In every object oriented language, in this situation `self` stands for the most specialized role `T` which first received the message (dynamic binding of `self`).

The dynamic binding of `self` is an essential component of object-oriented languages, but it has a price: it constrains inheritance to be only used to produce subtypes.

The situation is different when `extend` with `me` is considered. `extend` is essentially an inheritance operator, since it allows an object (the virtual one) to be built by starting from another one. The fact that here inheritance is exploited at instance level, rather than at the type level, is irrelevant. However, dynamic binding is not used for `me` for the following reasons:

- we cannot constrain `extend` to only produce values belonging to a subtype; one of the main reasons for introducing the viewing operator is to overcome the same limitation in the role mechanism;
- dynamic binding of `me` is not very useful. Dynamic binding of `self` is essential in role inheritance, because we expect that the role type where `self` is used will be extended with new subroles, and that these subroles may be used to complete the meaning of the current definition. With virtual objects, this style of programming by adding inheritance over inheritance (extension over extension) is not common. Moreover, the idea of completing or modifying the meaning of a virtual object by building a new object over it, is not compatible with the basic assumption that adding a view does not affect the behavior of the object viewed.

For the above reasons, the `me` identifier used in a method definition inside an `extend` operation is not linked dynamically to the virtual object that receives the message, but is statically linked to the object which is defined by the `extend` operation.

4.6 Equality

In the context of object databases, three kinds of equality are usually considered:

- *Identity equality* (*identical*, "`==`"). This corresponds to the equality of references or pointers in conventional languages: two objects are identical if their identities are the same.
- *Shallow equality* (*shallow equal*, "`=`"). Two objects are shallow equal if they have the same run-time type and their states are identical. That is, it goes one level deep, and compares corresponding identities of the state components.
- *Deep equality* (*deep equal*). This is a purely value-based equality: two objects are deep equal if they have the same run-time type and their states are value-based deep equal.

In Galileo97, every type is associated with an equality function, and the equality of two values is determined using the equality function associated with their static type. More precisely, `a = a'` is well typed if the type of `a` is either a supertype or a subtype of the type of `a'`, and `a` and `a'` are compared using the equality function associated with the supertype. As a consequence, the same pair of values can be equal or different according to the type it is accessed through. For example, the following two expressions would evaluate to `false` and `true`, respectively, since the `b` field would be ignored in the second comparison:

```
[ a:=1; b:=2] = [ a:=1; b:=3]; is false
[ a:=1; b:=2] :[ a:int] = [ a:=1; b:=2]; is true
```

In Galileo97, equality is value-based for most concrete types, such as record and sequence types, while it is determined by identity for modifiable values (values of type `var T`), functions and roles. Equality by identity seems the most appropriate for roles and locations for two reasons:

- in database applications it is usually more important to know whether two objects represent the same real-world entity, rather than knowing whether they give the same answer to every message. Likewise, for two locations, containing the same value is generally less interesting than being the same location;

- when two objects need to be compared by structure, they can simply be looked at through their record supertype. Likewise, to compare two locations `l1` and `l2` by content, one need only write `(at l1 = at l2)`.

The situation is slightly more complex with `view` types, which are to some extent intermediate between object and record types. In this case, the rule is that two values `O1` and `O2` belonging to type

`<`$\mathcal{T}_1$`,...,`$\mathcal{T}_m$`> view [ `A_1`:`$\mathcal{S}_1$`;...;`A_n`:`$\mathcal{S}_n$`]`

are equal if:

`(O1 As `$\mathcal{T}_1$`) = (O2 As `$\mathcal{T}_1$`) And ... And (O1 As `$\mathcal{T}_m$`) = (O2 As `$\mathcal{T}_m$`) And`
`  O1.`A_1` = O2.`A_1` And ... And O1.`A_n` = O2.`A_n` And`
`  O1!`A_1` = O2!`A_1` And ... And O1!`A_n` = O2!`A_n`.`

Note that:

1. for each `A`$_i$ field the associated equality is used. Hence, methods are compared by identity (since they are functions), updatable fields are also compared by identity, and constant concrete fields are compared by value;
2. two records are equal in type `[ ...]` iff they are equal in type `<> view [ ...]` .

The above points highlight that equality on virtual object values generalizes both role and record equalities. One may also expect that, whenever $T \sim S$, then $a : T = b$ is the same as $a : S = b$. However, this is not always true. As a counterexample, consider a pair of role types $\mathcal{S} \leq \mathcal{T}$. In accordance with the view types subtyping rules, the following equivalence holds:

`<`$\mathcal{S}$`> view [ ]` $\sim$ `<`$\mathcal{T}$`,`$\mathcal{S}$`> view [ ]`

Now, let `s` be a role of type $\mathcal{S}$ and `t1,t2` be two different roles of type $\mathcal{T}$. Comparing `t1 times s` with `t2 times s` gives two different answers in the two types above, since only in the second case are the two virtual objects also compared with respect to the result of the operation `x As` $\mathcal{T}$, which gives two different results.

```
(t1 times s):<S> view [ ] = (t2 times s); is true
(t1 times s):<T,S> view [ ] = (t2 times s); is false
```

The fact that two types that are equivalent with respect to subtyping are not equivalent with respect to equality is not very satisfactory, but could be avoided by adopting a more complex notion of equality, where two objects in type `<`$\mathcal{T}_1$`,...,`$\mathcal{T}_m$`> view [...]` are also compared with respect to the result of `O As` $\mathcal{S}$ for every supertype $\mathcal{S}$ of every $\mathcal{T}_i$. Choosing the best approach is a matter for further research.

As an example, consider the roles `John`, `JohnAsStudent`, and `JohnAsAthlete` defined at the beginning of Section 3, and the role `NewAthlete` defined as follows.

```
let NewAthlete := mkAthlete([ Name := "John Smith";
                              BirthYear:= 1967;
                              Code := 2;
                              Sport := "Basket" ] );
```

The following expression

```
JohnAsStudent  = JohnAsAthlete;
```

is not well typed since `JohnAsStudent` and `JohnAsAthlete` are not subtypes of each other.

`John` and `JohnAsAthlete` can be compared if they are considered of type `Person` as follows, and the equality predicates returns true:

```
(JohnAsStudent:Person)  = JohnAsAthlete;
```

Two virtual objects can be compared by identity using an appropriate view type. For example:

```
let type IdPerson := <Person> view [] ;

(JohnAsStudent:IdPerson) = (JohnAsAthlete:IdPerson); is true
```

Two role values can be compared by ignoring their identity using an appropriate record type, which allows both the shallow and deep equality to be simulated. For example:

```
let type PersonRecord := [Name:string; BirthYear: int] ;

(JohnAsStudent:PersonRecord) = (JohnAsAthlete:PersonRecord); is true

(JohnAsStudent:Person) = NewAthlete; is false

(JohnAsStudent:PersonRecord) = (NewAthlete:PersonRecord); is true
```

5 Virtual Classes

The virtual objects operators described so far, apart from being interesting by themselves, constitute the building blocks that can be used to define a higher level virtual class mechanism along the lines, for example, of the one proposed in [16]. The translation of such a mechanism in terms of virtual object operations defines its semantics in a precise way.

We will consider the following virtual class mechanism (`where`, `store`, `compute`, and `import` clauses are optional).

```
let NameView classview as
        x In BaseClass where Cond
    EleType := TBaseClass
                store    [ S1:= D1; ...; Sn:= Dn ]
                compute  [ C1:= E1; ...; Cn:= En ]
                import   [ I1; ...; In ];
```

This mechanism specifies:

1. the name *NameView* and the extent of the virtual class, by a query over a base class *BaseClass* which may be either real or virtual;
2. the type *EleType* of the objects in the virtual class, whose signature includes the base object attributes I_i specified with the clause `import`, the computed attributes C_i specified with the clause `compute`, and the stored attributes S_i (D_i are the default values) specified with the clause `store`. The I_i, C_i, and S_i must be distinct names.

A virtual class definition introduces a new kind of relation between classes, called *based-on*: *NameView* is *based-on BaseClass*. This relation implies the subset relation between *NameView* and *BaseClass* (as far as the identity based equality is concerned), but not a subtype relation.

A virtual class definition is translated into the definition of a view type for the virtual class element type, and a **derived** expression which defines the virtual class extent. The **compute** and **import** clauses are directly translated in terms of **extend*** and **project***; for the sake of simplicity, we omit the treatment of stored attributes, discussed in [3]. The type *TRealClass* in the translation is taken from the <*TRealClass*> component of the view type of the elements of *BaseClass*.

```
let type EleType := <TRealClass> view
                    [ I1; ...; In; C1 :T1; ...; Cn :Tn ];

let NameView := derived
                  select   x
                  from     x In BaseClass
                  where    Cond
                  extend*  [ C1:= E1; ...; Cn:= En ]
                  project* [ I1; ...; In; C1 :T1; ...; Cn :Tn ];
```

A virtual class mechanism must also allow one to insert a virtual class into the class hierarchy, considering its three aspects:

- the subtype relationship between the subclass element type and the superclass element type;
- the subset relationship between the subclass extent and the superclass extent;
- the inheritance relationship between the subclass definition and the superclass definition.

This is accomplished by the following mechanism, which allows a virtual class to be defined by inheritance from a previous virtual class definition.[4]

```
let NameSubView subset of SuperView classview as
        x In BaseClass where Cond
    SubEleType := is SuperEleType and
                  TBaseClass
                     store    [ S1 := D1; ...; Sn:= Dn ]
                     compute  [ C1:= E1; ...; Cn:= En ]
                     import   [ I1; ...; In ];
```

In this case, *SuperView* must be a (virtual) class based on a superclass of *BaseClass*. The class *NameSubView* inherits the **where**, **store**, **compute** and **import** clauses from the *SuperView* definiton, and it can extend them or override the attribute definitions with the *strict inheritance* constraint: whenever an attribute is redefined, its type must be a subtype of its previous type (or the same type). Thanks to these constraints and to the inheritance of the **where** clause, the extent of the defined virtual class is included in that of the superclass, and its element type is a subtype of that of the superclass.

The subset relationship between virtual classes, and the *based-on* relationship between a virtual class and a base class, are two different relationships: the first implies the subset, subtype, and inheritance relationships, while the second implies the subset relationship alone (Figure 10).

The semantics of the virtual subclass definition can be defined by extending the previous translation with a treatment of the inherited clauses. Inherited clauses are copied inside the translation; the order in which they are inserted in the translation determines the relative priorities. Since **project*** requires all the projected attributes to be different, we combine the different sets of attribute names with a **+** operator, which, whenever an attribute is both on its left and on its right hand side, puts in the result only the one on the right hand side: $(A:T; B:U) + (A:V; C:W) = (B:U; A:V; C:W)$. In the translation, we use I', C', T', E' for the inherited attributes and their types and definitions, and $Cond'$ for the inherited condition (for the sake of simplicity, we will again omit the treatment of stored attribute).

[4] It is also possible to inherit from a real class, which is seen as a virtual class based on itself, without **where**, **store**, and **compute** clauses, and where every attribute is imported.

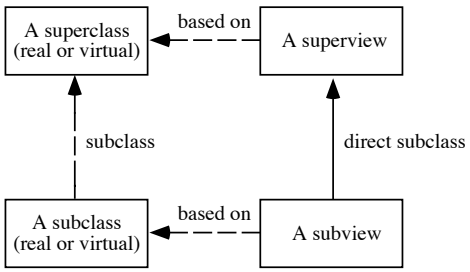

Figure 10: The subset and *based-on* relationships.

```
let type SubEleType := <TRealClass> view
                        [ (I'1; ...; I'n);
                        + (I1; ...; In);
                        + (C'1 :T'1; ...;C'n :T'n);
                        + (C1 :T1; ...;Cn :Tn) ];

let NameSubView := derived
                   select   x
                   from     x In BaseClass
                   where    Cond And Cond'
                   extend*  [ C'1:= E'1; ...; C'n:= E'n ]
                   extend*  [ C1:= E1; ...;Cn:= En ]
                   project* [ (I'1; ...;I'n);
                            + (I1; ...; In);
                            + (C'1 :T'1; ...;C'n :T'n);
                            + (C1 :T1; ...;Cn :Tn) ];
```

The value of such a translation is twofold:

- once the translation has been formally defined, the semantics of the virtual class construct is defined too; for example, the translation specifies what happens when an attribute with the same name is both "computed" in a superview and imported in a subview;
- the translation suggests which typing rules should be defined for the virtual class construct; for example, the translation suggests the strict inheritance constraint, and the fact that the expressions E_i should be type-checked in a context where the type of `me` is *TBaseClass* extended with the inherited computed attributes. Observe that, in a strongly typed context, defining the correct type rules is generally as difficult and as crucial as defining a non ambiguous semantics.

Example 6 *The following example shows two virtual classes* `Adults` *and* `EngineeringStudents` *defined from the classes* `Persons` *and* `Students`*:*

```
let Persons class Person <-> [Name: string; BirthYear: int];

let Students subset of Persons class
    Student <-> is Person and [SNumber: string; Faculty: string];


let rec
Adults classview as
```

```
        p In Persons where  CurrentYear() - p.BirthYear > 17
Adult := Person
          compute [WhoAreYou := meth(): string is
                                "My name is  " & me.Name & "." ]
          import  [Name];

let rec
EngineeringStudents subset of Adults classview as
    s. In Students where s.Faculty = "Engineering"
EngineeringStudent := is Adult and
                      Student
                      compute [Age := meth ():int is
                                      CurrentYear() - me.BirthYear]
                      import  [SNumber; Faculty];
```

The translation of these definitions using the virtual object operators is as follows:

```
let type Adult := <Person> view [Name;WhoAreYou:string];

let Adults := derived
   (select p
   from    p In Persons
   where   CurrentYear() - p.BirthYear > 17)
   extend*  [WhoAreYou := meth(): string is
                          "My name is  " & me.Name & "."]
   project* [Name];

let type EngineeringStudent :=
    <Student> view [Name;
                    SNumber; Faculty;
                    WhoAreYou:string; Age:int ];

let EngineeringStudents := derived
   (select s
    from   s In Students
    where  s.Faculty = "Engineering" And  CurrentYear() - s.BirthYear > 17 )
    extend*  [WhoAreYou    := meth(): string is
                              "My name is  " & me.Name & "."]
    extend*  [Age          := meth():int is CurrentYear() - me.BirthYear]
    project* [Name; SNumber; Faculty; WhoAreYou; Age];
```

□

6 Related Works

The idea of extending object databases with a view mechanism has been discussed by several authors (e.g., [12], [8], [9], [2], [27], [16], [18], [22], [24], [26], [30], [28], [17], [35],[19]). Most of them follow a "collection based" approach, which is quite different from our "object based" approach. The collection based approach is based on a "virtual class" mechanism. The object based approach, which is adopted in Galileo 97, first defines a set of virtual object operations, which allow a virtual object to be built starting from a set of other objects, and then combines these operators with a binding-by-name operation, such as our `let ViewName:= derived QueryExpr`, to define a view mechanism. The main issues which are studied in this approach are: (a) the semantics of virtual object operations, (b) method resolution and the semantics of `self`, and (c) a type system for virtual objects.

The essential differences between the two approaches are:

- Semantics: The object based approach is characterized by a simpler semantics. The notion of "a class whose objects are defined by a query" breaks some basic assumptions of the object-oriented data model, especially when classes are identified with object type extents and when virtual classes are inserted into the class hierarchy. This raises some problems, for example with respect to: (a) method resolution, (b) semantics of insertion/removal for virtual classes, (c) placement of virtual classes in the class hierarchy, and (d) assignment of an "object identity" to the elements of virtual classes. The object based approach, on the other hand, divides virtual class definition into two atomic notions whose semantics is easier to define.

- Expressive power: In the object based approach, object operations and binding-by-name are first class operators which can be freely used inside programs, while in the class based approach the virtual class operation can only be used to define schemas; this makes the object based approach very flexible. The ability to build virtual objects from more than one base object is also a feature which is not usually found in other approaches.

- Virtual class and virtual object updating: In the object based approach it is not possible to explicitly insert or remove elements from a virtual class, while this is possible, under certain conditions, in some class based approaches. This is an important limitation of the object based approach, which can however be overcome. At the object level, in both approaches every modifiable field of the base object which is directly accessible through the virtual object can be modified from the virtual object as well.

- Class hierarchies: Most class based approaches merge the subtype, inclusion and method lookup hierarchies. This fact creates some problems with virtual classes. For example, if a virtual class `V` is defined by both restricting and projecting a base class `B`, then `V` extension is included in `B`, but `V` type is a supertype of `B` type. A solution to these problems is in [16].

  In our object based approach we distinguish the subtyping, subset inclusion, and method lookup hierarchies (inheritance).[5] Virtual objects are assigned a first-class type, which is automatically placed in the subtype hierarchy, similarly to any other first class type of the language. Method resolution is not class based but is object based, hence virtual objects do not participate to the method lookup hierarchy.

More specifically, the different approaches can be analyzed with respect to the following questions:

1. Is the view mechanism based on collections of objects or on individual objects?
2. How do virtual objects and message resolution interact?
3. Does the language provide two orthogonal mechanisms to extend dynamically objects with new types and to define virtual objects?
4. Do the operators for defining virtual objects preserve object identity?
5. Is there a mechanism to group different objects into a single virtual object?
6. Is a virtual object a first class value, and, consequently, does it have a type and can it be used like any other value of the language? Similarly, does a virtual class behave exactly like a base class?

   Moreover, if the language is typed, is a virtual object type included in the language subtype hierarchy? Are rules provided to establish when a virtual object type is a subtype of an object type or of another virtual object type?

[5] By method lookup hierarchy we mean the order relation between object types which is used to perform method lookup.

7. Is it possible to insert objects into a virtual class?

8. Can a virtual object have its own state and methods?

9. Is there an operator to go from a virtual object to the object from which it has been defined?

Let us compare the solution offered by Galileo97 with three other proposals in the framework of a typed database programming language, AQUA [19], COCOON [28], and O_2 [2], and with the proposals in [18] (UniSQL/X), in [22], and in Chimera [16].

1. The Galileo97 and the Ohori-Tajima view mechanisms are based on the operation of building a virtual object, while the other approaches work by defining "virtual classes". While in Galileo97 and in the Ohori-Tajima approach virtual objects are built by value-level operations which can be exploited by any program, in the other approaches the creation of a virtual class is a schema-level operation, which is available during the schema definition phase only. This is the fundamental design choice, which has many consequences: the Galileo97 and Ohori-Tajima approach, based on a first class virtual object creation operation, gives rise to a more flexible mechanism, while the alternative approach, which confines virtual class creation to schema definition time, makes it easier to exploit traditional (i.e. relational) implementation techniques.

2. In Galileo97, a virtual object is essentially composed by a virtual interface applied to a base object. When the virtual object receives a message, the corresponding method is first looked up in the interface and then in the base object. This lookup process resembles standard message resolution, with two important differences. Firstly, the semantics of `self` has to be modified, as explained in this paper. Secondly, a method defined for a virtual object, which overrides a method defined for the real object, does not affect how the real object itself answers the message, even when double lookup is used; the behavior of a real object can be changed only by extending it with a new type.

   In the Ohori-Tajima proposal, method definition and message resolution are not studied.

   The O_2 approach is very different. While in our approach each virtual object is represented by a specific data structure which contains a reference to the corresponding real object, in the O_2 approach only real objects have a physical representation (at least in the abstract model, which does not necessarily coincide with the actual implementation). Virtual objects, or rather virtual *classes*, come into play since all applications access data through a specific schema, which may be either the real schema or a virtual one, and the behavior of objects depends on the chosen schema. A virtual schema specifies a set of virtual classes, along with which objects belong to these virtual classes, and how methods are implemented for each virtual class. When a virtual schema exists, the interfaces and methods of every object are those specified by that schema. Hence, when a message is sent to an object, it is necessary to determine to which classes the object belongs according to the current schema in order to execute the corresponding method; determining these classes is generall not easy. Moreover, if the object belongs to different unrelated virtual classes which implement the message, the ambiguity needs to be broken in some way. It is interesting to note that the current version of O_2 deals with these problems by adopting an approach which is quite similar to our approach, i.e. by "materializing" virtual objects [13, 14]. It is also interesting to note, though this point is not explicitly discussed in the O_2 papers, that the semantics they adopt for `self` seems to be the same as the one we adopt, despite the major differences between the two approaches.

   The UniSQL/X approach is intermediate. Each element in a virtual class has an OID which is different from the OID of the corresponding base object. The OID identifies the virtual class to which the object belongs, and is used to perform method dispatching. Hence, method resolution is class based as in O_2, but every virtual object contains (in its OID) enough information to perform an efficient method dispatch, as in our approach.

In Chimera each virtual object has the same OID as the corresponding real object. Method resolution is based on both the object OID and the static type of the object. The type hierarchies of real and virtual objects are kept separate, hence there is no interference between real and virtual methods.

In COCOON overriding is not allowed, hence no message resolution is needed.

3. Only the Galileo97 type system supports two orthogonal mechanisms to extend dynamically objects with new types and to define virtual objects. The same features have been studied in Chimera.

4. Galileo97, O_2, the Ohori-Tajima approach, and COCOON provide operators for defining object views with object identity preserving semantics. In AQUA this property only holds for views that do not change object structures: the project and join operator return a tuple. UniSQL/X objects do not have the same OID as the corresponding base object, but the OID of the base object can be extracted from the virtual object. In Chimera, when a virtual class is defined, the programmer can choose whether the original OID will be preserved, or a new OID generated, or no OID provided.

5. Galileo97 provides an operator to construct a virtual object by starting from two (or more) objects; the resulting virtual object contains the identities of both the starting objects. AQUA provides an operator to construct a pair which has two objects as its components; this is also possible in the Ohori-Tajima approach. O_2 provides an operator to construct a record with the attributes of two objects, and a mechanism to construct a new object from this record. In UniSQL/X a virtual class whose objects are built starting from more than one object can be created without any particular problem since, in this approach, the identity of the virtual object is different from the identity of its base object(s). The same happens in Chimera by generating a new OID.

6. In all of these systems a virtual object is a first class value. Consequently, it has a type, and can be used like any other value in the language. A virtual object type is included in the language subtype hierarchy, and rules are provided to establish when a virtual object type is a subtype of an object type, a virtual object type, or a record type. In UniSQL/X two different hierarchies are defined for base object types and for virtual object types. In Galileo97 and in the Ohori-Tajima approach the subtype relationship is inferred automatically by the type checker, exploiting rules such as those in Section 4.1. In O_2 and in UniSQL/X, on the other hand, the subtype relationship must be defined explicitly.

7. Only the UniSQL/X approach allows one to insert objects into virtual classes, under some conditions. This is achieved by translating this operation into the insertion of an object into the base class. The same is true for class removal.

8. Only Galileo97 and Chimera enable one to add new state components and new methods to a virtual object. In the Ohori-Tajima approach new state components can be added, but methods are not dealt with. In O_2 and in UniSQL/X a virtual object can be extended with new computed attributes and methods, but can have no new modifiable state components on its own. In COCOON a virtual object can be extended only with new computed attributes. In AQUA a virtual object cannot be extended at all.

9. Galileo97, UniSQL/X and the implemented version of the O_2 approach support an operator to go from a virtual object to the object from which the virtual object has been defined. This operator may be added without any problems to the Ohori-Tajima approach.

The Galileo97 approach also has similarities with the one adopted in [11], where a set of operators on records is defined. In fact both approaches work in the context of a statically and strongly typed functional programming language. However, our work also deals with the object-oriented aspects, namely identity

preservation and semantics of *self*, while Cardelli and Mitchell's study extends to polymorphic functions, which are not addressed here.

7 Conclusions

Two mechanisms have been presented to add flexibility to a programming language for object databases: roles and virtual objects. Roles allow objects to be dynamically extended to model entities which change their behavior, and the class they belong to, over time. Virtual objects are a mechanism to give a different interface of objects, redefining their structure and behavior, without affecting the behavior of the original object.

Roles and virtual objects are similar in many respects, since they both allow objects to be defined whose behavior depends on the observer, and they both allow an object to be extended. There are, however, some essential differences:

- the set of the roles of an object is part of the object itself, and the object can be tested with the predicate `isalso` to find out which roles it has; while a view is conceptually external to the object (for this reason, an operation is defined to remove a role from an object, while no similar operation is needed for virtual objects);
- adding a new role to an object transforms its type into a subtype, while the corresponding view operation `extend` produces an object whose type may not be related to the original one;
- the behavior of an object changes when it gains a new role, while it is not affected by the creation of a new virtual object;
- the set of roles which an object can belong to, is constrained by the role type hierarchy which has been statically defined, while the set of view types which can be used to access an object is open ended;
- both roles and virtual objects can send messages to themselves, but in the first case the message is received by the role that received the "original message" (see Section 4.5), while in the second case the message is received by the virtual object where the method is defined.

The main contributions of this work are (a) the clarification of the relationship between the two mechanisms, (b) the study of a set of statically and strongly typed operators which allow the virtual object mechanism to be included in a typed language, and (c) the use of these operators to give the semantics of a higher level mechanism to define virtual classes. The semantics of these operators has been defined through a simplified storage model, which reflects some aspects of the current implementation. A formal account of the essential mechanisms of the language is presented in [3].

Both the role and the virtual object mechanisms have been implemented in a main memory implementation of the Galileo97 language for personal computers.

Acknowledgments

This article benefitted from discussions with S. Abiteboul, E. Bertino, G. Guerrini, G. Mainetto, and M. Scholl. Thanks also to the anonymous referees for their constructive comments, which helped us reorganize the paper and add important new material.

References

[1] Martín Abadi and Luca Cardelli. *A Theory of Objects*. Springer-Verlag, Berlin, 1996.

[2] S. Abiteboul and A. Bonner. Objects and Views. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 238–247, 1991.

[3] A. Albano, G. Antognoni, and G. Ghelli. View operations on objects with roles for a statically typed database language: The type rules and the operational semantics. Technical report, Dipartimento di Informatica, Università di Pisa, 1998.

[4] A. Albano, R. Bergamini, G. Ghelli, and R. Orsini. An Object Data Model with Roles. In *Proceedings of the International Conference on Very Large Data Bases (VLDB)*, pages 39–51, Dublin, Ireland, 1993.

[5] A. Albano, L. Cardelli, and R. Orsini. Galileo: A Strongly Typed, Interactive Conceptual Language. *ACM Transactions on Database Systems*, 10(2):230–260, 1985. Also in S.B. Zdonik and D. Maier, editors, *Readings in Object-Oriented Database Systems*, Morgan Kaufmann Publishers, Inc., San Mateo, California, 1990.

[6] A. Albano, M. Diotallevi, and G. Ghelli. Extensible Objects for Database Evolution: Language Features and Implementation Issues. In *Proc. of the Fifth Intl. Workshop on Data Base Programming Languages (DBPL), Gubbio, Italy*, 1995.

[7] A. Albano, G. Ghelli, and R. Orsini. Fibonacci: A Programming Language for Object Databases. *The VLDB Journal*, 4(3):403–439, 1995.

[8] F. Bancilhon, S. Cluet, and C. Delobel. A Query Language for the O_2 Object-Oriented Database System. In R. Hull, R. Morrison, and D. Stemple, editors, *Proceedings of the Second International Workshop on Database Programming Languages*, pages 122–138, San Mateo, California, 1990. Morgan Kaufmann Publishers.

[9] E. Bertino. A View Mechanism for Object-Oriented Databases. In A. Pirotte, C. Delobel, and G. Gottlob, editors, *Advances in Database Technology – EDBT '92*, volume 580 of *Lectures Notes on Computer Science*, pages 136–151, Berlin, 1992. Springer-Verlag.

[10] E. Bertino and G. Guerrini. Objects with Multiple Most Specific Classes. In *Proc. Ninth European Conference on Object-Oriented Programming*, LNCS N. 952, pages 102–126, Berlin, 1995. Springer-Verlag.

[11] L. Cardelli and J.C. Mitchell. Operations on Records. *Mathematical Structures in Computer Sciences*, 1(1):3–48, 1991.

[12] U. Dayal. Queries and Views in an Object-Oriented Data Model. In R. Hull, R. Morrison, and D. Stemple, editors, *Proc. of the Second Intl. Workshop on Data Base Programming Languages (DBPL), Salishan Lodge, Oregon*, pages 80–102, San Mateo, California, 1989. Morgan Kaufmann Publishers.

[13] C. Souza dos Santos. Design and Implementation of an Object-Oriented View Mechanism. In *10èmes Journées Bases de Données Avancées*, Clermont-Ferrand, France, 1994.

[14] C. Souza dos Santos and E. Waller. The O_2 Views User Manual - Version 2. O_2 Technology, Versailles, France, 1995.

[15] G. Gottlob, M. Schrefl, and B. Röck. Extending Object-Oriented Systems with Roles. *ACM Transactions on Information Systems*, 14(3):268–296, 1996.

[16] G. Guerrini, E. Bertino, B. Catania, and J. Garcia-Molina. A Formal Model of Views for Object-Oriented Database Systems. *Theory and Practice of Object Systems (TAPOS)*, 3(2):103–125, 1997.

[17] S. Heiler and S.B. Zdonik. Object Views: Extending the Vision. In *IEEE Data Engineering*, pages 86–93, Los Angeles, 1990.

[18] W. Kim and W. Kelley. On View Support in Object-Oriented Database Systems. In W. Kim, editor, *Modern Database Systems*, pages 108–129. Addison-Wesley, Reading, Massachusetts, 1995.

[19] T. Leung, G. Mitchell, B. Subramanian, B. Vance, S. Vandenberg, and S. Zdonik. The AQUA Data Model and Algebra. In *Fourth Intl. Workshop on Database Programming Languages*, pages 157–175, New York, 1993.

[20] R. Motschnig-Pitrik. Requirements and Comparison of View Mechanisms for Object-Oriented Databases. *Information Systems*, 21(3):229–252, 1996.

[21] E. Odberg. Category Class: Flexible Classification and Evolution in Object-Oriented Databases. In G. Wijers, S. Brinkkemper, and T. Wasserman, editors, *Proc. of the 6th Conference on Advanced Information Systems Engineering (CAISE '94), Utrecht, The Netherlands*, number 811 in LNCS, pages 406–420, Berlin, 1994. Springer-Verlag.

[22] A. Ohori and K. Tajima. A Polymorphic Calculus for Views and Object Sharing. In *Proceedings of the ACM SIGACT-SIGMOD Symposium on Principles of Database Systems (PODS)*, pages 255–266, Minneapolis, Minnesota, 1994.

[23] M.P. Papazoglou and B.J. Krämer. A Database Model for Object Dynamics. *The VLDB Journal*, (6):73–96, 1997.

[24] C. Parent and S. Spaccapietra. An Algebra for a General Entity-Relationship Model. *IEEE Transactions on Software Engineering*, 11(7):634–643, 1985.

[25] J. Richardson and P. Schwartz. Aspects: Extending Objects to Support Multiple, Indipendent Roles. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages 298–307, Denver, CO, 1991.

[26] E.A. Rundensteiner. MultiView: A Methodology for Supporting Multiple Views in Object-Oriented Databases. In *Proc. of the Eighteenth Intl. Conf. on Very Large Data Bases (VLDB), Vancouver, British Columbia, Canada*, pages 187–198, San Mateo, California, 1992. Morgan Kaufmann Publishers.

[27] C.S. Dos Santos, S. Abiteboul, and C. Delobel. Virtual classes and bases. In *Advances in Database Technology – EDBT '94*, 1994.

[28] M.H. Scholl, C. Laasch, C. Rich, H.J. Scheck, and M. Tresch. The COOCON Object Model. Technical Report 211, Departement Informatik, ETH Zurich, 1994.

[29] M.H. Scholl, C. Laasch, and M. Tresch. Views in Object-Oriented Databases. In *Proc. Second Intl. Workshop on Foundations of Models and Languages for Data and Objects*, pages 37–58, 1990.

[30] M.H. Scholl and H.J. Scheck. A Relational Object Model. In S. Abiteboul and P.C. Kanellakis, editors, *Proc. of the Third Intl. Conf. on Database Theory (ICDT), Paris, France*, number 470 in LNCS, pages 89–105, Berlin, 1990. Springer-Verlag.

[31] M.H. Scholl and H.-J. Schek. Supporting Views in Object-Oriented Databases. *IEEE Data Engineering Bulletin*, 14(2), 1991.

[32] G.M. Shaw and S.B. Zdonik. An Object-Oriented Query Algebra. In R. Hull, R. Morrison, and D. Stemple, editors, *Proc. of the Second Intl. Workshop on Data Base Programming Languages (DBPL), Salishan Lodge, Oregon*, pages 103–112, San Mateo, California, 1989. Morgan Kaufmann Publishers.

[33] J.J. Shilling and P.F. Sweeney. Three Steps to View: Extending the Object-Oriented Paradigm. In *Proceedings of the International Conference on Object-Oriented Programming Systems, Languages and Applications (OOPSLA)*, volume 10 of *ACM SIGPLAN Notices*, pages 353–361, 1989.

[34] R. Wieringa, W.de Jonge, and P. Spruit. Using Dynamic Classes and Role Classes to Model Object Migration. *Theory and Practice of Object Systems (TAPOS)*, 1(1):61–83, 1995.

[35] S. Zdonik. Object-Oriented Type Evolution. In F. Bancilhon and P. Buneman, editors, *Advances in Database Programming Languages*, pages 277–288. Addison-Wesley, Reading, Massachusetts, 1990.