Foundations of Object-Oriented Languages

TR94-22
Giuseppe Castagna and Gary T. Leavens

November 23, 1994

Iowa State University of Science and Technology
Department of Computer Science
226 Atanasoff
Ames, IA 50011
Foundations of
Object-Oriented Languages
2nd Workshop report
Giuseppe Castagna and Gary T. Leavens

TR #94-22
November 1994

**Keywords:** object-oriented programming, type checking, verification, modules, binary method, matching, subtyping, record update, state, semantics, covariance, contravariance, flow analysis, self, method schemas.


This report is to appear in ACM SIGPLAN Notices. © Giuseppe Castagna and Gary T. Leavens, 1994. Copies may be made for any purpose whatever, provided this copyright notice appears on the copy.

Department of Computer Science
226 Atanasoff Hall
Iowa State University
Ames, Iowa 50011-1040, USA
Abstract


1 Introduction

In Paris on July 1–2 1994, researchers gathered at the second workshop of a series of NSF and ESPRIT-sponsored workshops to discuss the foundations of object-oriented (OO) programming languages. This series of workshops is organized by Kim Bruce and Giuseppe Longo. The previous edition was held in Stanford and has been reported in [7].

This year’s workshop was held in the Conservatoire National des Arts et Métiers and organized by the Laboratoire d’Informatique de l’Ecole Normale Supérieure; local arrangements were handled by Roberto Bellucci and Giuseppe Castagna. Participation was by invitation only (see the appendix).

The purpose of this workshop was to exchange ideas about the latest research in the semantics of OO programming languages. The workshop consisted of several presentations and some discussion. A lively email discussion also developed shortly after the workshop about binary methods.

This report is organized as follows. Section 2 summarizes the talks given at the workshop. For each talk, one or two references to the work are given. (The bibliography and document pointers are available on the World-Wide Web via the URL: http://www.cs.iastate.edu/~leavens/FOOL94.html.)

Section 3 summarizes the discussions that took place at the workshop, and afterwards.

2 Reviews of Presentations

Ten of the participants made presentations. The following are brief summary reviews of their talks. For each talk, we try to describe the problem being solved, and then give an overview of the solution and its significance. The summary follows the order in which the presentations were given.

Kim Bruce: Matching is better than constraining for bounded polymorphism in OOLs.

Kim Bruce discussed his work on providing a flexible, yet type-safe, polymorphic and imperative OO programming language [9]. In an imperative OO language, there are fewer subtypes than in previously-studied theoretical OO languages that do not have mutation. Hence bounded polymorphism in which the bounding relationship is a subtype relationship is not as useful as might be desired. This problem underlies the problem users experience with contravariance rules, and is one reason Eiffel uses an unsafe covariant rule.

PolyTOIL, is a polymorphic, imperative OO language that is type-safe. PolyTOIL has expressions that denote classes, and thus a class has a type, written ClassType(σ, τ). This type is distinguished from the type of its objects, ObjectType τ, because the object type suppresses information about the instance variables (σ).

An important property of PolyTOIL is that methods do not have to be type-checked each time that they are inherited. This is accomplished by giving self the type MyType, and only assuming that MyType “matches” ObjectType τ. The definition of matches implies that if class C′ inherits from C, and the corresponding object types are ObjectType τ and ObjectType τ, then ObjectType τ′ matches ObjectType τ. Using matching permits PolyTOIL to have a type system that is nearly as expressive as Eiffel (which uses unsafe covariant type-checking), but which is provably type safe, and does not have Eiffel’s link-time global check. This expressiveness is especially important for methods that take or return
an argument of type MyType.

Bruce contrasted matching with subtyping, and noted that in PolyTOIL these orderings on types are distinct. He noted that matching is more useful than subtyping as the ordering in bounded polymorphism. For example, consider a polymorphic function that needs a comparable type argument, $T$, where comparable is a type with a binary method $\text{leq}$ of type $\text{MyType} \rightarrow \text{Bool}$. This works better if the bound is “$T$ matches $\text{comparable}$” rather than requiring that “$T$ is a subtype of comparable.”

**Benjamin Pierce:** Positive Subtyping.

Benjamin Pierce described joint work with Martin Hoffman [15]. The problem he addressed is the record update problem: how to type and reason about update in a $\lambda$-calculus model of an OO language without mutation. An update function is one that returns an object of the same type. For example, consider a type Point. A function $\text{bump}$ of type $\text{Point} \rightarrow \text{Point}$ is an update function. The famous subclass of Point, ColoredPoint would also have a $\text{bump}$ function, but this one of type $\text{ColoredPoint} \rightarrow \text{ColoredPoint}$. The problem is how to model $\text{bump}$, with a polymorphic function; as is well known, the usual bounded polymorphic definition does not work.

Several complex solutions for this problem have been offered by others. Pierce offered a much simpler solution to a smaller problem: that is a solution in which subtyping for function types is only applied in positive positions, such as function result types. With this restriction, the semantics of a subtype relationship, $S \leq T$, is not just a coercion function from $S$ to $T$, but a pair of functions. The first element of the pair is the coercion from $S$ to $T$, while the second function, of type $S \rightarrow T \rightarrow S$, is able to update the $T$ part of an $S$ object.

In the second part of his talk, Pierce discussed equational laws for reasoning about the calculus, and an example proof. The program that is the subject of the proof contains two class definitions, the second inheriting some of its behavior from the first. Although both classes involve recursive self-reference through the pseudo-variable $\text{self}$, the proof is structured so that properties of the second class are established without looking at the implementation of the first class. Thus the proof is modular in the same sense that the object-oriented program is modular.

**Didier Rémy:** Programming objects in ML-A.RT.

Didier Rémy described a solution to the problem of supporting OO programming idioms in ML [20]. The idea is to use the OO model proposed by Pierce and Turner [18] in a type inference framework. To that end he enriched the ML+references language by some features already present in the type inference literature. More precisely ML+references is extended by extensible records (to obtain polymorphic extension for inheritance), projective types (to describe the type structure of fields independently of their presence), existential types (to have state encapsulation) and recursive types.

The difficult problem in ML is to get the objects to share the vector without exposing the representation to clients of the object. To add inheritance, one must support $\text{self}$. The usual encoding of object records with $\text{self}$ is recursive, so one would have to create objects with an unsafe fixpoint, which would not work with the ML type system. In order to avoid this, Rémy modified the Pierce and Turner model by passing both the with exposed state and $\text{self}$ to each method, rather than just the exposed state. State encapsulation is still realized though the use of existential types. Recursive types are now required since sending messages create fix points. Inheritance is then coded using wrappers.

The final result is the encoding of an object-oriented style in the extended ML, which allows to write type safe object-oriented programs in a type inference framework. Yet some problems persist in this approach: coercions are difficult to write and must be explicit, message sending cannot be polymorphic (so there is no subtype polymorphism), and creation of recursive values is too limited.

**Scott Smith:** State in Object-Oriented Programming Languages.

Scott Smith discussed joint work with Jonathan Eifrig, Valery Trifonov, and Amy Zwarico [14]. This work addresses the problem of how to provide a type system for an OO language with mutable state. This work also shares many of the goals of Kim Bruce’s, in that it also deals with a polymorphic, imperative OO programming language.

Smith discussed an approach to this problem in which the semantics of the imperative OO language are given by translation to a non-OO language. For its static type system heavy use is made of $\text{F}$-bounded polymorphism. He noted that a goal is to find a translation into a typed calculus such that one can refer to $\text{self}$ in expressions that initialize instance variables. (This allows object-overridable methods, as in Modula-3, and circular structures.)

Smith started by discussing an untyped language, LOOP. In LOOP one can create objects, send messages, and read and write instance variables. To allow one to refer to $\text{self}$ in expressions that initialize instance variables, class and object constructions have
the following form.

\begin{verbatim}
c = class(self)
instance variables x = ... self ...
methods m = ... self ...
end
o = new c
\end{verbatim}

These are translated into the following typed calculus, called SOOP.

\begin{verbatim}
c = \lambda s. let y = \{x = ref ... self(...), m = ... self(...)\}
in \lambda()y
o = let r = ref \lambda().⊥
in r := c(\lambda().r()); r
\end{verbatim}

Since there are no reference types in an object, one gets subtyping, but this construction does not catch circular definitions. In discussions it was agreed that Smith’s use of F-bounded subtyping is able to do what Kim Bruce’s did with matching, but without an explicit matching relation.

Smith gave some rules for reasoning about LOOP. He also discussed the proof of the soundness of the type system of SOOP, which is done by subject-reduction.

**Martín Abadi:** An imperative object calculus.

Martín Abadi presented joint work with Luca Cardelli [3] [2]. The problem he addressed is the definition of a primitive calculus to model object-oriented programming. The necessity of such a calculus is supported by the problems met in performing typing and subtyping. In particular, in this work, the modeling and typing of imperative features is studied.

Abadi started by describing an untyped imperative calculus formed by variables and objects with the primitives of method invocation, method override and object cloning. He showed how these constructs suffice to encode both fields and the “imperative \( \lambda \)-calculus” (pure \( \lambda \)-calculus plus assignment). After having described the operational semantics, Abadi defined a first-order typing for the calculus and proved its soundness. Namely, he proved that, starting from a “sound” store, well-typed programs do not go wrong. Abadi concluded this part of his talk by arguing that in an imperative setting some of the difficulties that resulted in the use of sophisticated functional type theories could be avoided; the example he used is the one with movable points, i.e., points with a method that moves the object: while in a functional language such a method must return a new object with the same type as \textit{self}, in the imperative setting it is not necessary to specify any type, since the method works by side-effects.

Starting from the previous type system, Abadi then presented a partial correctness logic. The idea is to write specifications of objects that generalize the previous types and that serve as invariants. The final product is a Hoare logic on programs written in (a modified version of) the untyped imperative calculus described above. Abadi concluded his talk with an example, showing how to use this logic to prove that a given program computed a given value without modifying the store.

**Giuseppe Castagna:** Covariance and contravariance: conflict without a cause.

Castagna presented work that addresses the long-standing problem of covariance vs. contravariance [10]. The so-called \textit{contravariant} rule, used to subtype function types, while assuring type safety, seemed to prevent satisfactory typing of some special cases, notably when binary methods were involved. For this reason in some systems this rule had been replaced by a \textit{covariant} one, but type safety was lost.

Castagna argued that covariance and contravariance appropriately characterize two distinct and independent mechanisms. The so-called contravariance rule correctly captures the \textit{substitutivity}, or subtyping relation (that establishes which sets of codes can replace in every context another given set). A covariant relation, instead, characterizes the \textit{specialization} (in OO jargon, the \textit{overriding}) of code (i.e., the definition of new code that replaces the old one in some particular cases). Therefore, covariance and contravariance are not opposing views, but distinct concepts that each have their place in object-oriented systems; both can (and must) be type safely integrated in an object-oriented language. He also showed that the independence of the two mechanisms is not characteristic of a particular model but is valid in general, since covariant specialization is present also in record-based models but it is hidden by the incapability of the existing calculi to model multiple dispatching. Thus he showed how it is possible to model multiple dispatch in the record-based models and argued that this improvement naturally allows one to (sub)type in a straightforward (and type safe) way binary or more generally \( n \)-ary methods even in these models. The example he gave was the one of \texttt{Point} and \texttt{ColoredPoint} with an \textit{equal} method (presented also in other talks of this workshop). He demonstrated that if in the \textit{equal} field of a \texttt{ColoredPoint} one specifies both the code for an argument of class \texttt{Point} and the code for an argument of class \texttt{ColoredPoint} and if the code is dynamically selected by multiple dispatching, it is then type safe to have \texttt{ColoredPoint<Point}. 

3
Gary Leavens: Modules for Multi-methods: Information Hiding and Program Composition.

Leavens discussed joint work with Craig Chambers [12]. The problem he addressed was one pointed out by William Cook at previous workshops [13]: how to achieve encapsulation and composability for systems of multi-methods. In comparison to abstract data type languages (such as CLU or Ada), and single dispatching object-oriented languages such as Smalltalk, multi-method systems such as CLOS do not have good encapsulation. A method may be written in any part of the program that specializes on objects of a certain class, and thereby that method may gain access to the representation of such objects. When combining independently developed systems of multi-methods, type errors may occur, for which no part of the program is uniquely responsible.

To solve these problems, Leavens described a module mechanism for the language Cecil [11]. This mechanism directly addresses the encapsulation problem, since one can limit access to fields by declaring them private. However, since Cecil has subtyping, it becomes possible for a method to be invoked on an object whose type is defined in a module that is not explicitly imported at the point of the method call. This problem is avoided by requiring subtypes to be defined in “extension modules.” The methods defined in modules that extend $M$ are implicitly available at run-time in any module that explicitly imports $M$.

While the module system solves much of the composability problem, there remains the problem of independent extensions of a given module. To solve this problem, Leavens observed that the person who puts both independently-developed extension modules in a program should be responsible for resolving any problems in a “most-extending module.” A unique most-extending module must exist for each module, and this condition can be checked quickly at link-time. This seems to provide multi-methods with the same degree of information hiding as standard OO languages, while retaining other advantages of multi-methods.

Jens Palsberg: A Type System Equivalent to Flow Analysis.

Jens Palsberg presented a joint work with Patrick O’Keefe [17]. He addressed the problem of finding a type system that accepts exactly the same programs as safety analysis. Safety analysis is a flow analysis of programs which collects type information and uses this information to accept only safe programs (i.e., programs that cannot go wrong). Such analyses have the advantage that they can be applied to untyped languages, where the more traditional abstract interpretation needs types for defining abstract domains.

More in detail he considered the untyped lambda calculus with zero and successor together with the type system defined by Amadio and Cardelli [6] and composed by recursive, arrow, top, and bottom types and Int, plus a subtyping relation with subsumption. Palsberg defined the safety analysis of a term as the resolution of a given set of constraints and showed a cubic time algorithm that computes it. Finally, he proved that a term is typable in Amadio-Cardelli’s type system if and only if there exists a solution to the safety analysis. As a result he obtained a type inference algorithm for the given type system, thereby solving an open problem.

Luca Cardelli: Primitive Object Types with Self.

Luca Cardelli described joint work with Martin Abadi [2], which is an improvement of the work in [1]. The problem he addressed was the definition of a primitive calculus to model object-oriented programming. The necessity of such a calculus is supported by the problems met in performing typing and subtyping. A focus of this work is the type of self.

The starting calculus is the untyped lambda calculus presented in [1], while the starting type theory is the second order type theory of [4] with some improvements: a new quantifier $Obj$ is used which, unlike the previous quantifier, can “move points”, override self-returning methods, and encode classes. Special annotations are used to identify read-write fields (typically instance variables), read-only fields (typically methods), and write-only fields. These annotations also allow to encode arrow types enjoying the usual contravariance/covariance properties. Cardelli also introduced the notion of pre-method (not to be confused with the CLOS concept of the same name), which is a function that is later used to construct a method. This definition is useful since class can then be seen as collections of pre-methods and inheritance relation between classes as pre-method reuse. Although very few primitives are used, the resulting calculus is powerful enough to express object types, class types, and method specialization, covering both class-based and delegation-based frameworks. While the delegation-based frameworks are essentially built-in, Cardelli demonstrated the expressibility of class-based frameworks by closely emulating TOOPLE [8].

Emmanuel Waller: Method Schemas.

Emmanuel Waller described joint work with Serge Abiteboul and Paris Kanellakis [5]. He addressed the problem of static verification of consistency of method schemas (decidability and complexity of various cases).
A method schema is a simple programming formalism for object-oriented databases with features such as classes, methods, inheritance, name overloading, and late binding. An important problem is to check whether a given method schema can lead to an inconsistency in some interpretation (a method called with arguments for which this method is undefined). When no restrictions are enforced on method schemas, dynamic errors are possible, and the consistency question is undecidable in general. (This is shown using program schemas.) Decidability is obtained for monadic and/or recursion-free method schemas. In particular, consistency of monadic method schemas is shown to be decidable in $O(n^3)$ time, where $n$ is the size of the method definitions and $c$ is the size of the class hierarchy; also, it is logspace-complete in PTIME, even for monadic, recursion-free schemas. Method signature covariance is shown to simplify the computational complexity of key decidable cases.

The incremental consistency checking of method schemas is a formalization of the database schema evolution problem. A detailed study of some decidable cases is given in [21].

3 Summary of Discussions

At the end of the workshop there was a general discussion. The discussion started with Luca Cardelli offering his (tongue-in-cheek) list of the 10 most important ideas in OO semantics. The list is given below.

1. subsumption (when one object/method can replace another)
2. subsumption
3. update
4. self-value (methods with self as a parameter)
5. self types (Modula-3 doesn’t have this)
6. don’t forget subsumption (subsumption should still work well with self types)
7. method specialization (methods to allow override, pre-methods)
8. classes and inheritance
9. polymorphism
10. still subsumption (subsumption should still work well with polymorphism)

In reply to this, Kim Bruce said he would put “matching” as items 2 (and 10!) on the list. Other suggestions for this list were: implementation efficiency (Benjamin Pierce), compilability (Carl Gunter), separation of classes and types (Scott Smith).

Some suggested adding multi-methods to the list, and Cardelli noted that you need some sort of encapsulation for that (as in $O_2$ or the work of Chambers and Leavens). A discussion about modularity ensued, and it was noted that the work of Pierce and Turner [19] and Katiyar, Luckham, and Mitchell [16] allows one to operate on two sets of objects, with representation access, using bounded existential quantifiers.

Carl Gunter started a new line of discussion by saying that, as a semanticist, he would like to get the features of OO in a way orthogonal to other features of a language. One might imagine ML with objects, where one could use OO techniques, but still program in other paradigms. Carl also wanted to know what other ideas were subsumed by OO ideas.

Finally, there was a discussion of binary methods, which reflects their importance in several of the talks. This discussion also continued after the workshop, in email. Pierce asserted that binary methods, which cause so many theoretical problems for modularity and type inference, should be banished. He asked whether there were any useful examples where binary methods were really needed. Cardelli offered a “proof” that binary methods are both useless and worse than useless. The basic idea is that binary methods hurt subsumption, so if one defines what would be binary methods as functions (outside of classes), you win. Cardelli concluded that the only really useful binary methods are multi-methods.

In response, Kim Bruce offered a program that showed a situation in which binary methods are useful (and can be type-checked in PolyTOIL). The example constructed ordered lists of items, in which the items can be compared. Although subsumption was lost because of the binary methods, matching allowed the program to perform a useful function. There were various recodings of this example (by Pierce, Bruce, Rémy and Cardelli). Codings without using matching either needed a more complex type system, or required a change in the interface used by clients of the program.

The result of the extended email discussion was a shared understanding of the extra usefulness of matching and binary methods, of the extra complexity needed to model them, and of specific techniques for modeling them. We now understand well how subsumption and matching can coexist. The issue of whether to eliminate one or the other, or neither, is still unresolved. (Several of the participants in this
discussion are planning an extended report on the binary method problem.)

Acknowledgements

Thanks to Kim Bruce, Luca Cardelli, Giuseppe Longo, Jens Palsberg, Benjamin Pierce, Didier Rémy, Scott Smith, and Emmanuel Waller for comments and corrections to drafts of this report.

Appendix: Participants

The participants in the workshop are listed in Table 1.

References


Table 1: Participants in the workshop.

<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
<th>E-mail address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abadi, Martin</td>
<td>Digital, SRC</td>
<td><a href="mailto:ma@src.dec.com">ma@src.dec.com</a></td>
</tr>
<tr>
<td>Abiteboul, Serge</td>
<td>INRIA</td>
<td><a href="mailto:Serge.Abiteboul@inria.fr">Serge.Abiteboul@inria.fr</a></td>
</tr>
<tr>
<td>Bellucci, Roberto</td>
<td>LIENS</td>
<td><a href="mailto:bellucci@dmi.ens.fr">bellucci@dmi.ens.fr</a></td>
</tr>
<tr>
<td>Benzaken, Véronique</td>
<td>Université de Paris I - Sorbonne</td>
<td><a href="mailto:Veronique.Benzaken@lri.fr">Veronique.Benzaken@lri.fr</a></td>
</tr>
<tr>
<td>Bouladoux, Francois</td>
<td>ENS</td>
<td><a href="mailto:Francois.Bouladoux@dmi.ens.fr">Francois.Bouladoux@dmi.ens.fr</a></td>
</tr>
<tr>
<td>Brue, Kim</td>
<td>Williams College</td>
<td><a href="mailto:kim@cs.williams.edu">kim@cs.williams.edu</a></td>
</tr>
<tr>
<td>Cardelli, Luca</td>
<td>Digital</td>
<td><a href="mailto:luca@src.dec.com">luca@src.dec.com</a></td>
</tr>
<tr>
<td>Castagna, Giuseppe</td>
<td>LIENS</td>
<td><a href="mailto:castagna@dmi.ens.fr">castagna@dmi.ens.fr</a></td>
</tr>
<tr>
<td>Davies, Rowan</td>
<td>INRIA &amp; CMU</td>
<td><a href="mailto:rowan@cs.cmu.edu">rowan@cs.cmu.edu</a></td>
</tr>
<tr>
<td>Delobel, Claude</td>
<td>LRI</td>
<td><a href="mailto:claude@02tech.o2tech.fr">claude@02tech.o2tech.fr</a></td>
</tr>
<tr>
<td>Dhara, Kishore</td>
<td>Iowa State University</td>
<td><a href="mailto:dhara@cs.iastate.edu">dhara@cs.iastate.edu</a></td>
</tr>
<tr>
<td>Eifrig, Jonathan</td>
<td>The Johns Hopkins University</td>
<td><a href="mailto:eifrig@os.jhu.edu">eifrig@os.jhu.edu</a></td>
</tr>
<tr>
<td>Ghelli, Giorgio</td>
<td>University of Pisa</td>
<td><a href="mailto:ghelli@di.unipi.it">ghelli@di.unipi.it</a></td>
</tr>
<tr>
<td>Gunter, Carl</td>
<td>University of Pennsylvania</td>
<td><a href="mailto:gunter@cis.upenn.edu">gunter@cis.upenn.edu</a></td>
</tr>
<tr>
<td>Leavens, Gary</td>
<td>Iowa State University</td>
<td><a href="mailto:leavens@cs.iastate.edu">leavens@cs.iastate.edu</a></td>
</tr>
<tr>
<td>Longo, Giuseppe</td>
<td>CNRS-LIENS</td>
<td><a href="mailto:longo@dmi.ens.fr">longo@dmi.ens.fr</a></td>
</tr>
<tr>
<td>Milsted, Kathleen</td>
<td>France Telecom CNET</td>
<td><a href="mailto:milsted@issy.cnet.fr">milsted@issy.cnet.fr</a></td>
</tr>
<tr>
<td>Moggi, Eugenio</td>
<td>DISI, Univ. di Genova</td>
<td><a href="mailto:moggi@disi.unige.it">moggi@disi.unige.it</a></td>
</tr>
<tr>
<td>Palsberg, Jens</td>
<td>Aarhus University</td>
<td><a href="mailto:palsberg@daimi.aau.dk">palsberg@daimi.aau.dk</a></td>
</tr>
<tr>
<td>Pierce, Benjamin</td>
<td>LFCS, Univ. of Edinburgh</td>
<td><a href="mailto:bcp@dos.ed.ac.uk">bcp@dos.ed.ac.uk</a></td>
</tr>
<tr>
<td>Rémy, Didier</td>
<td>INRIA-Rocquencourt</td>
<td><a href="mailto:Didier.Remy@inria.fr">Didier.Remy@inria.fr</a></td>
</tr>
<tr>
<td>Smith, Scott</td>
<td>The Johns Hopkins University</td>
<td><a href="mailto:scott@cs.jhu.edu">scott@cs.jhu.edu</a></td>
</tr>
<tr>
<td>Trifonov, Valery</td>
<td>The Johns Hopkins University</td>
<td><a href="mailto:trifonov@cs.jhu.edu">trifonov@cs.jhu.edu</a></td>
</tr>
<tr>
<td>Waller, Emmanuel</td>
<td>U. Paris Sud</td>
<td><a href="mailto:Emmanuel.Waller@lri.fr">Emmanuel.Waller@lri.fr</a></td>
</tr>
</tbody>
</table>


