Abstract. The paper considers the problems and the solutions that are concerned with the extension of Java with Higher Order mechanisms. We start from the motivations namely the code reusability and the code expressivity. Through the description of the JavaΩ project we discuss the design of constructs for Higher Order methods, methods as parameters and for closures that can be integrated in Java and whose implementation must preserve re-targetability and portability of the Java compilers and of the JVM. We give the formal definition of such constructs, including the formalization of a translation semantics which can be converted into a source-to-source translation and implemented as one-pass preprocessor. Examples of the use of the formalized constructs are included.

1 Introduction

The paper considers the problems and the solutions that are concerned with the extension of Java with Higher Order (HO) mechanisms. These mechanisms are typical of functional languages and include higher order abstraction, that makes a Java method parametric with respect to other methods that can be passed as parameters, and code as first class value, that can be assigned to variables, passed as parameter, returned by method invocations, and executed in different points of a Java program. The main motivations for the presence of HO mechanisms in a programming language are code reusability and expressivity. The code reusability degree of a language [BT96] is concerned with the structures that the language provides to re-adapt the code of a program to the modifications which are consequent to its use, for instance to eliminate program bugs, or to modify its context or interface of use, or to add new program functionalities. Java provides inheritance with this aim, while functional languages provide HO functions: Combining the two forms highly enhances the degree of code reusability of the resulting language. Code expressivity [Fel91] is concerned with the (data and control) structures that the language provides to express the expected behaviors of the programs, for instance, to represent the application data, or to control data visibility and access, or to model code iteration or code induction on different types of data. A good expressivity improves the use of the language and decreases programming errors. Java provides object, classes and inner classes with this aim, while functional languages provide functions as first class value. HO methods and closures (as anonymous functions as first class values are improperly called in Java) aims [Gaf07] are (a) simplifying many kinds of programmings that currently rely on anonymous inner classes; (b) enabling more powerful libraries with methods (Control APIs) in which the controlled statements are received as closures or as method parameters (mc_parameters); (c) improving concurrent programming and simplifying applications that rely on existing concurrent APIs; (d) supporting a programming style that enhances the use of aggregate operations, i.e. operations that apply an operation to collection members; (e) enabling future APIs design to replace language design for extending the Java platform. In fact, at the base of
all the above points is the verbosity and the awkwardness of anonymous inner classes in expressing functions and more in general, parameterized code. In addition, the Java scope rules limit, in inner classes, the use of non local variables (or class members) to those that are declared final in the method code. This increases the difficulty in expressing parameterized code that share variables with methods that are invoked during the computation and are no more active.

Describing the evolution of the JavaΩ project we discuss the problems that arise from the design of higher order constructs that can be added to Java and whose implementation must preserve, at the same time, the features of re-targetability and portability of the Java compilers and of the Java Virtual Machine, JVM. We give the formal definition of such constructs, including the extensions to the syntax of the Java Language Specification, JLS, and the formalization of a translation semantics. This semantics provides a formal way to map programs of the extended language onto behaviorally equivalent programs of the ordinary language. Hence, it describes the behavior of the new constructs in terms of compositions of well known constructs of the original language.

Several languages exist which integrate OO and HO features, they are: C# [WH03], F# [Sym03], J# [Lak04], Python [Com10], Scala [Ode10], Ruby [FM08]. All such languages allow to define closures and a few ones, C#, F#, J#, Scala, also allow to pass methods as parameters but the construct features are different according to the specific language structures, including type system, method overloading etc. and a presentation would require a detailed analysis which is beyond the scope of this paper.

In Section 2 a structured presentation of JavaΩ introduces to the faced problems, the proposed solutions, and the results that are obtained on extending Java with HO mechanisms. In Section 3 the principles and the formalization of the translation semantics, used in the subsequent sections, are illustrated. The mechanism of closures and that of HO methods with mc_parameters are designed in such a way that they result one another orthogonal and can be discussed separately. Hence, in Section 4 Java 1.5 is extended to allow closures. The syntax and semantics adopted are essentially those in [BO09a] except for minor changes in the translation semantics due to technical reasons concerned with the used formalism. In Section 5 Java 1.5 is extended to allow HO methods and mc_parameters, as defined in [BO08a,BO09b], but the translation semantics is modified in order to consider Java 1.5 (with generics) instead of Java 1.3. The Section 6 contains the most meaningful examples published in the above cited papers revisited because of the the integration of the two mechanisms. The Section 7 integrates the two extensions by putting the two translations together and addresses the implementation. The last section contains some final remarks.

2 The JavaΩ Project

This project originated in 2004 [BO04] where, through a case study, we compare the expressivity of five different programs, all based on the OO programming paradigm but using also higher order mechanisms (mainly in the form of methods with higher order functionalities), all solving the same problem. Three programs are written in Java, one is written in J++ and the last one is written in Pizza. As far as the the higher order mechanisms are concerned, the first two programs in Java use reflection, the third one uses anonymous classes, the J++ program uses delegates and the Pizza program uses function abstractions (closures). The result of the comparison led to identify three
possible mechanisms to add to Java to extend it with higher order: higher order methods, method extraction and closures.

In [BO05] the integration of such mechanisms in an OO paradigm is studied, through the extension of the $\zeta$-calculus [AC96]. The new $\zeta_x$-calculus provides, for such mechanisms, the reduction semantics and the typing rules and prove calculus confluence, subtyping, subtyping contravariance, subject reductions hence type soundness.

If the problem of the integration of HO mechanisms in an OO language can be considered positively solved, the problem of how this can be accomplished in Java is still open. As a matter of fact, Java has an implementation based on a mixture of compilation and interpretation. At the top level of the implementation process is a compiler that translates Java programs into Java bytecode for the Java virtual machine, JVM. The machine is a collection of structures for processing bytecode and has been designed to be defined once and to have as many different implementations as the different platforms that run Java programs.

This implementation ensures re-targetability of Java compilers and high portability of the Java object code. Extending Java with new mechanisms it is good practice to plan to preserve such appealing features of its implementation. In [BO07b] we began to study the design of HO mechanisms especially tailored for Java and its peculiar implementation, defining the structure of $m$-parameters as a form of higher order methods for Java. [BO07b,BO08c] provide the syntactic extensions to Java grammar to pass methods as parameters to other methods. A formal semantics is also defined and examples that show the use of the mechanism are shown. The expressivity is discussed by comparing programs that use such a mechanism with programs that compute in an almost equivalent way but are written in ordinary Java.

One of the most remarkable aspects of our study is the formal semantics which is defined in the form of a translation semantics. This kind of semantics maps programs of the extended language into programs of ordinary Java as discussed in Section 3. This approach in the design of HO extensions for Java is used in [BO08a] where the translation semantics of Java 1.4, extended with $m$-parameters, is defined and used to form the kernel [BO08b] of a pre-processor for Java compilers [BO07b]. The drawbacks of $m$-parameters are the impossibility to i) statically check type correctness and ii) pass overloaded methods. Such drawbacks are both overcome in [BO08a,BO09b] where $mc$-parameters are defined. $mc$-parameters have a type which specifies also the class hierarchy in which the passed method must be defined, allowing a static type checking and overloading static resolution. The translation semantics of $mc$-parameters is defined using anonymous inner classes, involved in the Java methodology known as callback [Hor07].

In 2006, the Java community started a debate about the opportunity to introduce closures in Java. The proposals resulting from such a debate are contained in [BGGvdA06,CS06b,BLB06] and more recently [BGGvdA08,CSC08,Rei09]. In such proposals Java is extended with different closure forms and other mechanisms with the aim to allow a better use of closures and integration in the language (closure conversion, shared variables, loop abstraction). As a matter of fact, our paper [BO05] about $\zeta_x$-calculus already contained the definition of a form of closures. In [BO09a] we apply the translation semantics approach to the extension of Java with a form of closures which adopts many of the features of the above cited papers. As for HO methods the approach guarantees the goals of portability and re-targetability but is not well suited to study different and alternative definitions of closures and related mechanisms. For this reason
we resorted to a minimal core calculus, more meaningful and suited to Java features, to extend it with closures. Such a calculus [BO10] is obtained extending Featherweight Java [ABW01].

3 Translation Semantics and Source to Source Translations: Principles

The translation semantics describes the meaning of the (data and control) structures of a language (called, the source language) in terms of the structures of another language (called the object or target language). In the JavaOmega project, the source language is a Java extension while the target language is ordinary Java. Hence, It has several benefits including:

- It defines the meanings of the new constructs in terms of the well known constructs of Java, without the introduction of the semantic structures otherwise needed.
- It helps in evaluating the expressivity power of the new constructs allowing an immediate comparison between a program that uses the new constructs and its translation.
- It allows an implementation of the execution support of the extended language completely re-targetable on all the compilers developed for Java.

The translation semantics is basic in the compiler design and implementation, where it is used in combination with other kinds of semantics (in order to guarantee the correctness of the compiler back-end) and is almost all expressed through attribute grammars. Other uses of translation semantics are in compiler bootstrapping, source-to-source translation [AGG+80,KLV03], reverse engineering [War94].

A translation semantics can be formally defined by means of a rule system containing one or more rules for each of the productions of an unambiguous grammar G of the source language. Let \( p \) be the production \( C_0 := C_1 \ldots C_n \) of the grammar, with, \( C_0, \ldots, C_n, n + 1 \) (possibly non different) grammatical categories. A grammatical category can be either a syntactic, e.g. Expression, or a terminal (lexical) category, e.g. InfixOp (cfr. the Java grammar of JLS [GJSB05]). Let \( G_S \) and \( G_L \) be the set of all the syntactic and terminal respectively, categories of a grammar \( G \). Moreover, a grammatical category can define either an infinite set of syntactic structures, e.g. all the expressions that can be written in the language, or a finite set, e.g. all the infix operators that can occur in the expressions of the language. Terminal categories never occur in the left part of a production and may consist of singletons called tokens. Let \( \mathcal{E} \) be a source to source language translation and \( \Sigma^E \) be the rule system that defines \( \mathcal{E} \) using grammar \( G \). Then, in correspondence to production \( p \), \( \Sigma^E \) contains \( k \geq 1 \) rules of the form:

\[
\mathcal{E}(A_0) = r_i(A_1, \mathcal{E}(A_1), \ldots, A_n, \mathcal{E}(A_n)) \text{ with } b_i(A_0, A_1, \ldots, A_n)
\]

Symbols \( A_0, \ldots, A_n \) are variables and are ranging on the structures of the set defined by the corresponding categories \( C_0, \ldots, C_n \). \( b_i \) is a predication and expresses a condition that constrains the structures that can be bound to each \( A_j \) in order to apply the \( i^{th} \) rule of production \( p \). Eventually, \( r_i \) is a syntax constructor and expresses the translation produced by the rule on the structure bound to \( A_0 \). Hence, the rule can be read: if the structure that we are translating is a structure of category \( C_0 \) and is exactly composed of a structure \( A_1 \) of category \( C_1 \) followed by...followed by a structure \( A_n \) of category \( C_n \)
and moreover, on the structures $A_0, A_1, \ldots, A_n$ condition $b_i$ holds, then the structure $A_0$ is translated in the structure obtained by applying $r_i$ to the structure $A_1$, and possibly, to the translated structure of $A_1, \ldots$ to the structure $A_n$, and possibly, to the translated structure of $A_n$. The one described above, is the general structure of a rule system for source to source translation. In fact, 1) predication $b_i$ rarely requires more that one or two of the $n + 1$ arguments indicated, as well as the constructor $r_i$ rarely requires the translation of all the component structures $E(A_1), \ldots, E(A_n)$ and the use of both $A_i$ and of translation $E(A_i)$; 2) More often, the rule system contains only one rule for production. In this case condition $b_i$ is obviously absent. 3) Eventually, only a small number of rules have a different $r_i$. These rules are the one really significant for characterizing the translation defined by rules system. The most part of rules have the following form:

$$E(X_0) = f(X_1, E(X_1)), \ldots, f(X_n, E(X_n)) \text{ with } b_i(X_0, X_1, \ldots, X_n)$$

In the rule, the constructor $r_i$ is the identity constructor that leaves unchanged the structure $X_0$ but replaces each substructure $X_j$ ($j > 0$) with the corresponding translation if any: Hence $f(X_j, E(X_j))$ stands for $X_j$ if $X_j \in G_L$, for $E(X_j)$ if $X_j \in G_S$. A Rule as the above one is really, a metarule when each symbol $X_j$ is a metavariable ranging on variables of many different categories. A metarule stands for all the rules that can be obtained, starting from the productions of the grammar, by instantiating symbols $X_j$ with variables of the right categories. Looking at the two rule systems given in next sections, we note that they have 12 and 8 respectively, significant rules and only one metarule for a Java grammar that has many dozen of productions. The metarule is for both the systems,

$$E(X_0) = f(X_1, E(X_1)), \ldots, f(X_n, E(X_n)) \text{ otherwise}$$

It is called otherwise metarule, and applies to the each production of the grammar for which none of the other rules of the system apply. Eventually we note that, the use of language grammar guarantees completeness and termination of source to source translation, namely each program has its translation and it is computed in as many steps as the productions used to parse the program. Moreover, the non ambiguity of grammar guarantees soundness, namely the program translation is unique.

### 4 Closures: Anonymous Functions in Java

Closures enclose a piece of code, namely a statement list and/or an expression, possibly parameterized, and make such code available for possibly, many and different invocations in the program. More important, closures are values (function objects [Gaf08]) that can be computed by and passed to methods: They allow the definition of methods whose code, hence whose behavior, is parametric with respect to the closures with which are invoked. A closure may contain free variables that are bound in the lexical scope of the blocks enclosing the closure definition.

#### 4.1 Closure basic structure

A basic common structure for Java closures emerges from [BGGvdA08,CSC08,Rei09]. This structure can be described by the following characteristics: Closures
1. encapsulate an arbitrary piece of Java code and generalize it through parameters (as anonymous functions [Lan66], written in Java)

2. have a type which depends on the argument types, the type of the possibly computed value, and the exception types. In Java closure types are generic.

3. are values hence can be assigned, passed as parameters, returned as value result (of the correct type)

4. can be invoked in a uniform way. In particular if a closure is bound to a formal parameter \(x\) of a method, or another closure, the invocation of \(x\) does not depend on (selectors defined in the) the closure type.

5. can contain non-local variables, in this case the non local variable:
   - is the one bound in the lexical scope of the nearest block enclosing the closure, according to Java static scope rule.
   - such variable remains accessible even though the frame related to the execution of the block has been deallocated.
   - it can be accessed and modified as local variables.

Syntax apart, such a structure leads to express the same behaviors (i.e. intended semantics) and programming uses (i.e. methodologies) for closures [Gaf08,CSC08], and includes those in [BLB06]. As a matter of fact, such closures basic structure is adopted in [BO10] and also here.

### 4.2 Syntax of closures.

A closure is syntactically an expression. The following example shows a closure definition: it has a formal parameter of name \(y\) and type int and returns a value of type int, the body contains a non local variable \(x\).

\[
\{(\text{int} \ y) \ : \text{int} \Rightarrow x += y; \ \text{return} \ ++x;\}
\]

The declaration of the return type follows the syntactic structure of method definition in Java and does not force the implementation to infer the return type as in [BGGvdA08]. However, we defer to type checking, of ordinary Java compilers, the burden of checking the correctness of the types declared in the translated programs. As a matter of fact, our translation maps closures with incorrect types into programs with incorrect types too. A closure has a type and it may be assigned to a variable, passed as a parameter of the right type, returned as a value, invoked as a method using the reserved selector invoke.

All the extensions needed to Java grammar are reported in Appendix A. The main extensions for closures are:

\[
\begin{align*}
\text{Primary} ::= & \ldots | \text{Closure} | \ldots \\
\text{Closure} ::= & \{ \text{FParameters} : (\text{Type} | \text{void}) \ \text{ThrowsOpt} \Rightarrow \text{Block} \} \\
\text{Type} ::= & \text{ParameterizedType} | \text{BasicType} | \text{ClosType} \\
\text{ClosType} ::= & \{ ([\text{ExtendedTypeList}] : (\text{Type} | \text{void}) \ \text{ThrowOpt}) \\
\text{ExtendedTypeList} ::= & \text{ExtendedType} | (\text{ExtendedType})^* \\
\text{NonInvocationSelector} ::= & \ldots | .\text{invoke} [\text{Arguments}] | \ldots
\end{align*}
\]

Closures are anonymous functions but unlike functional abstractions can (access and) modify (free, i.e. non local) variables bound in the lexical scope of the block in which the closure is defined. In particular a closure can use variables that are bound in a scope
that is no more active [LY96] at the time the closure is invoked, see [Tro08] and the use of the hashtable in the example in Fig. 4. Since local variables are allocated in a Java frame [LY96] when the scope is active and deallocated when it is not active, then the handling of non local variables in closures needs a specific treatment. A solution of this problem leads to a new, additional, notion for those local variables that can be accessed and modified as non local in a closure. Such variables must be allocated in the heap. In [Gaf08] this kind of variables are annotated @shared to mark the difference with ordinary local variables. In the structure adopted in this paper, shared is an additional modifier for local variables to mark that the difference between this new kind of variables and the variables of ordinary Java is mainly semantics and it involves memory allocation (heap vs. frame [LY96]), access and modification of the variable. Hence the syntax of both the local variable declarations and the formal parameters [GJSB00] is extended adding:

LocalVariableDeclarationStatement::= /final | shared / Type VariableDeclarators
FormalParameter::= /final | shared/ [Annotations] ExtendedType VariableDeclaratorId

4.3 Semantics of closures.

The intended meaning of closures associates to a:

(i) closure type \{(ET_1, ..., ET_n) : ET [throws QI_1, ..., QI_k] \}, through a one-to-one correspondence \$\exists\$, a reference type \$\exists(ES_1, ..., ES_n, [QI_1, ..., QI_k])$\hspace{1mm}ES \). Where, for each \$i, ES_i\$ (resp. \$ES\$) is the type \$ET_i\$ (resp. \$ET\$), if \$ET_i\$ (resp. \$ET\$) is not a closure type, otherwise it is the reference type associated to \$ET_i\$ (resp. \$ET\$).

(ii) closure literal \{(FS_1 \hspace{1mm}At_1 \hspace{1mm}ET_1 \hspace{1mm}V_1, ..., \hspace{1mm}FS_n \hspace{1mm}At_n \hspace{1mm}ET_n \hspace{1mm}V_n) : ET [throws QI_1, ..., QI_k] \Rightarrow BB\}, a (function) object, of type \$\exists(ES_1, ..., ES_n, [QI_1, ..., QI_k, ES])\$, that wraps a function \$\mathcal{H}_{BB}\$. Type \$\exists(ES_1, ..., ES_n, [QI_1, ..., QI_k, ES])\$ is obtained according to (i) above from type \{(ET_1, ..., ET_n) : ET [throws QI_1, ..., QI_k] \} that is extracted from the closure literal in the obvious way and is the type of the closure literal. Function \$\mathcal{H}_{BB}\$, applied to a n-tuple of values \$V_1, ..., V_n\$ of type \$ES_1, ..., ES_n\$ respectively, either computes a value \$V\$ of type \$ES\$ or fails depending on the execution of \$BB_r\$ in a frame where \$V_1\$ is bound to value \$V_1\$ and, ..., and \$V_n\$ is bound to value \$V_n\$. Where \$BB_r\$ is the ordinary Java code resulting from the sequence of statements (and/or declarations) of closure body \$BB\$ in which each free (i.e. non local) variable, if any, is bound to the corresponding \$\text{shared}\$ (or \$\text{final}\$) variable of the lexical scope in which the closure occurs. If the computation fails then an exception (of a type possibly in \$QI_1, ..., QI_k\$) is thrown.

(iii) closure invocation \$E\_\text{invoke}(E_1, ..., E_n)$, where \$E\$ is an expression computing a (function) object \$c\$ of type \$\exists(ES_1, ..., ES_n, QI_1, ..., QI_k, ES)$, the application of the function, that \$c\$ is wrapping, to the n-tuple of values \$V_1, ..., V_n\$ that results from the evaluation of the argument list \$E_1, ..., E_n\$. Arguments \$E_1, ..., E_n\$ must be of the right types \$ES_1, ..., ES_n\$.

(iv) \$\text{shared}\$ variable (resp. parameter) \$V\$ of type \$ET\$, possibly initialized with expression \$E\$, a variable of a class of objects wrapping variables of type \$ET\$ that are allocated in the heap. The wrapped variable is possibly initialized to the value of \$E\$. The \$\text{shared}\$ variable \$V\$ is then, accessed and modified through such a reference.
\[\mathcal{F}[\mathcal{C}B]_v = \{\mathcal{F}[\mathcal{M}B_i]_0 \ldots \mathcal{F}[\mathcal{M}B_n]_0\} \text{ with } CB = \{MB_1 \ldots MB_n\}\]

\[
\mathcal{F}[P_y] = \left\{ \begin{array}{ll}
\tau(p_y) & \text{with } P_y \in \text{Identifier} \mid \text{this} \\
\text{new } & \langle \mathcal{F}[E_T]_v, \ldots, \mathcal{F}[E_T]_v, \mathcal{F}[E_T]_v, Q_1, \ldots, Q_I \rangle \}
& \text{with } P_y \equiv \{(P_1, \ldots, P_n) : \text{ET TR } \Rightarrow \text{BB}\} \\
\mathcal{F}[\text{public ET invoke}(P_1, \ldots, P_n) TR\{
& \text{BB} \} \tau((\text{this} / (\text{\$self, on})) \}
& \land P_i \equiv \text{FS}_i \text{ET}_i \text{Vi} \\
& \land \text{TR } \equiv \text{throws } Q_1, \ldots, Q_I \\
\text{new } & \langle \mathcal{F}[E_T]_v, \ldots, \mathcal{F}[E_T]_v, Q_1, \ldots, Q_I \rangle \}
& \text{with } P_y \equiv \{(P_1, \ldots, P_n) : \text{void TR } \Rightarrow \text{BB}\} \\
\mathcal{F}[\text{public void invoke}(P_1, \ldots, P_n) TR\{
& \text{BB} \} \tau((\text{this} / (\text{\$self, on})) \}
& \land P_i \equiv \text{FS}_i \text{AT}_i \text{ET}_i \text{Vi} \\
& \land \text{TR } \equiv \text{throws } Q_1, \ldots, Q_I \\
\end{array} \right. 
\]

\[
\mathcal{F}[E_T]_v = \left\{ \begin{array}{ll}
\text{IShk } \langle \mathcal{F}[E_T]_v, \ldots, \mathcal{F}[E_T]_v, \mathcal{F}[E_T]_v, Q_1, \ldots, Q_I \rangle \}
& \text{with } ET \equiv \{(ET_1, \ldots, ET_n) : ET \text{ TR}\} \\
& \land \text{TR } \equiv \text{throws } Q_1, \ldots, Q_I \\
\text{IShkvoid } \langle \mathcal{F}[E_T]_v, \ldots, \mathcal{F}[E_T]_v, \mathcal{F}[E_T]_v, Q_1, \ldots, Q_I \rangle \}
& \text{with } ET \equiv \{(ET_1, \ldots, ET_n) : \text{void TR}\} \\
& \land \text{TR } \equiv \text{throws } Q_1, \ldots, Q_I \\
\end{array} \right. 
\]

\[
\mathcal{F}[L_x]_v = \left\{ \begin{array}{ll}
\text{final } \mathcal{F}[E_T]_v I_1 = \mathcal{F}[E]\tau((\text{\$this, off})) \\
& \text{with } L \equiv \text{final ET I } \equiv E \land E \in \text{Closure} \\
\text{final } T = \text{new } T(\mathcal{F}[E_T]_v) \\
& \text{with } L \equiv \text{shared ET V } \equiv E \land E \notin \text{Closure} \\
& \text{with } T = \text{C$Shared } \mathcal{F}[E_T]_v \}
& \land \text{ET } \equiv \text{throws } Q_1, \ldots, Q_I \\
\text{final } T = \text{new } T(\mathcal{F}[E_T]_v) \\
& \text{with } L \equiv \text{shared ET I } \equiv E \land E \in \text{Closure} \\
& \text{with } T = \text{C$Shared } \mathcal{F}[E_T]_v \}
& \land \text{ET } \equiv \text{throws } Q_1, \ldots, Q_I \\
\end{array} \right. 
\]

\[
\mathcal{F}[MB]_v = \left\{ \begin{array}{ll}
\text{final } \mathcal{F}[ET]_v T = \text{new } T(V_i), ET_v = \mathcal{F}[ET]_v \\
& \land L_0 = \text{final Object } \text{\$self } = \text{this}; \\
\text{and } L_1 = \{ \lambda \} & \text{if } FS, \in \text{[final]} \\
& \text{if } FS, \equiv \text{shared } \land T = \text{C$Shared } \mathcal{F}[ET]_v \}
& \\end{array} \right. 
\]

\[
\mathcal{F}[BB]_v = \mathcal{F}[L_x]_v, \mathcal{F}[BB]_v \tau((I/v)\text{value}) \text{ with } BB \equiv S;BB_v \land L = \text{shared ET I } \text{,[], } = E 
\]

\[
\mathcal{F}[E_T]_v = \tau(I) = \mathcal{F}[E]\tau((I/v)\text{this}) \text{ with } E \equiv I = E_1 \land E_1 \in \text{Closure} 
\]

where

\[
\tau \uparrow (E_1/E_i, \ldots, I_i/E_n) = \left\{ \begin{array}{ll}
E_1 & \text{if } x = I_i \equiv E_i \notin \{(\text{\$self, on}), (\text{\$self, off})\} \\
\text{\$self } & \text{if } x = \text{this } \equiv I_i \equiv E_i = (\text{\$self, on}) \\
\text{this } & \text{if } x = \text{this } \equiv I_i \equiv E_i = (\text{\$self, off}) \\
\tau(x) & \text{if } x \notin \{I_1, \ldots, I_n\} \\
\end{array} \right. 
\]

**Legend**: For arbitrary index p, A \in Arguments, AR \in ArrayCreatorRest, At \in Annotations, BB, \text{p} \in Block, BB, \text{BB} \in BlockStatement, BS, BS, p \in BlockStatement, BT \in BasicType, CB \in ClassBody, CD, CD, p \in ClassOrInterfaceDeclaration, CN \in CreatedName, CR \in ClassCreatorRest, CU \in CompilationUnit, D, D, \text{p} \in ModifierOpt, E, E, p \in Expression, EL, EL, \text{p} \in ExtendedTypeList, ET, ET, \text{p} \in ExtendedType, F, F, \text{p} \in ETs, FP, FP, \text{p} \in FP, FS, FS, p \in [final | shared], G \in \text{ExplicitGenericInvocationSuffix}, I, I, p \in \text{Identifier, IM} \in \text{ImportDeclaration, IO} \in \text{InfixedOp, K} \in \text{Literal, L} \in \text{LocalVariableDeclarationStatement, MB, MB} \in \text{MemberDecl, N} \in \text{NonInvocationSelector, O, o} \in ["", P, P, p \in \text{FormalParameter, PO} \in \text{PostfixOp, PS, PS, p} \in \text{PrimarySelector, Py} \in \text{Primary, Q, q} \in \text{TPs, Q} \in \text{QualifiedIdentifier, RT} \in \text{RootClass, S, S, p} \in \text{Statement, Sp} \in \text{Selector, T, T, p} \in \text{Type, T, T, p} \in \text{TypeArgument, TA} \in \text{Parsopt, TL, TL, p} \in \text{TypeList, TR} \in \text{ThrowOpt, TV, TV, p} \in \text{Type} | \text{void} U, U, J \in \text{ST, V, V, p} \in \text{VariableDeclarationId}, \\
\]

**Fig. 1.** Closures: F[] Translation Semantics.


4.4 The Translation Semantics $\mathcal{F}[]$. 

The translation semantics is defined by translation $\mathcal{F}[]$, in Fig. 1, it is based on the structures of *interfaces, anonymous classes*, and *classes* of variables, in the sense that $\mathcal{F}[]$ translates closures into a composition of such structures. The one-to-one correspondence 3 of the closure intended meaning is implemented as a bijective function $\mathcal{ClosType} \rightarrow \text{Interface}$, between the closure type and the interface for objects wrapping the corresponding type. The implementation of 3 is based on a family of interfaces, indexed by a couple of naturals $n, k$, defined as follows:

\[
\text{interface I}$nk$<ET_1, \ldots, ET_n, ET, QI_1 \text{ extends Throwable,} \ldots, QI_k \text{ extends Throwable}>\{
\text{public ET invoke}(FS_1 At_1 ET_1 V_1, \ldots, FS_n At_n ET_n V_n) \text{ throws QI}_1, \ldots, QI_k;\}
\]

\[
\text{interface I}$nkvoid$<ET_1, \ldots, ET_n, QI_1 \text{ extends Throwable,} \ldots, QI_k \text{ extends Throwable}>\{
\text{public void invoke}(FS_1 At_1 ET_1 V_1, \ldots, FS_n At_n ET_n V_n) \text{ throws QI}_1, \ldots, T_k;\}
\]

Each interface represents the functions with $n$ arguments raising exceptions of $k$ different types. In the first case the represented functions return a value of type $ET$. In the second case they do not compute any value (i.e. compute the nullary, unit value). According to the semantics defined for closures, $\mathcal{F}[]$ translates:

(i) the closure type $\{<ET_1, \ldots, ET_n>: ET ([\text{throws QI}_1, \ldots, QI_k])\}$ into an instantiation of the interface I$nk$ in which the type variables $ET_i$ are instantiated to the corresponding type $ES_i$ obtained translating $ET_i$;

(ii) a closure literal $\{(FS_1 At_1 ET_1 V_1, \ldots, FS_n At_n ET_n V_n): ET ([\text{throws QI}_1, \ldots, QI_k] \Rightarrow BB)\}$ of type $\{<ET_1, \ldots, ET_n>: ET ([\text{throws QI}_1, \ldots, QI_k])\}$ into an instance of the class implementing the corresponding interface, namely:

\[
\text{new I}$nk$<ES_1, \ldots, ES_n, ES, QI_1, \ldots, QI_k>()\{
\text{public ES invoke}(FS_1 At_1 ES_1 V_1, \ldots, FS_n At_n ES_n V_n)
\text{ throws QI}_1, \ldots, QI_k \{BB_r\}\}
\]

where $BB_r$ results from the translation of $BB$ and $ES_i$ results from the translation of $ET_i$, according to $\mathcal{F}[]$.

(iii) closure invocation $E$.invoke($E_1, \ldots, E_n$) into $\mathcal{F}[E]$.invoke($\mathcal{F}[E_1]_r, \ldots, \mathcal{F}[E_n]_r$).

As a matter of fact, closure invocation is not considered in the rules defined in Fig. 1 since it is trivially dealt with by the otherwise metarule.

(iv) shared variables of type $ET$ into final variables of the class:

\[
\text{class C}$Shared$<ET> \{ET\ \text{value}; \ C$Shared$<ET \ n>\{\text{value=n;}\}\}
\]

Each access to a variable $V$, declared shared, is replaced by an access to $V$.value. Class C$Shared$<ET> is also used to translate shared parameters. If a shared formal parameter of type $ET$ is declared in a method with body $\{BB\}$, it is translated into a parameter without modifier final. Moreover a final variable with a new name, is declared in the translation of $BB$ being of type C$Shared$<ES>, where $ES$ is the type resulting by the translation of $ET$, and is initialized to contain the value of the parameter. All occurrences of the parameter in the translation of $BB$ are translated into accesses to the field value of the new variable. The function $G$: Identifier $\rightarrow$ Identifier is used to generate the new name that does not clash with
the other identifiers in \( BB \). \( G \) is an injective function that applied to a name of a parameter generates a new name. In this way the translation \( F[\tau] \) preserves the method headers.

### 4.5 The Formal Definition of \( F[\tau] \)

The translation semantics is defined using the syntax directed rule system described in Section 3. Hence, it has at least one rule for each production of the unambiguous grammar of Appendix A, and the translation computes applying to each program construct the rule that is associated to the production with which the construct is parsed. As a matter of fact, the rule system \( F[\tau] \) contains only 12 significant rules and one otherwise metarule: The metarule applies when the production with which the construct is parsed has no significant rules associated to it, and the translation leaves unchanged the construct structure but replaces each component with the result of its translation. In the sequel, we illustrate the translation discussing rule by rule (only for the significant ones) how they work. Firstly we introduce the environment \( \tau \) used as a parameter in the translation \( F[\tau] \), whose definition is contained at the end of Fig. 1. The environment \( \tau \) associates an identifier \( I \in Identifier \) to an expression \( E \in Expression \). The empty environment is denoted by \( \emptyset \) and does not contain bindings. Any other environment is obtained by \( \tau \uparrow (I_1/E_1, \ldots, I_k/E_k) \) that adds \( k \) new bindings to a given environment \( \tau \) (possibly, the empty environment). Each binding associates to the name \( I_i \) an expression \( E_i \), and we have that \( \tau(I_i) = E_i \). Eventually, \( \forall I \emptyset(I) = \perp \) where \( \perp \) is the undefined term. The environment \( \tau \) contains a binding for:

1. all the shared local variables or shared formal parameters that have the construct (to which \( F[\tau] \) applies) in their (lexical) scope. In this case the value of the binding is \( I_i.value \) as explained in the previous subsection 4.4.
2. a local variable whose value is a closure. When a closure is assigned to a local variable identifier, all occurrences of such identifier, in the closure body are recursive invocations of the closure. Hence they must be replaced by the self reference \( this \) to the function object that the translation produces.
3. \( this \): Since the translation maps closures into function objects and occurrences of \( this \) are recursive invocations, we need a different identifier to refer to the object on which the method that defines the closure is invoked. Hence we define a reserved identifier \( s$elf \) and add a declaration \texttt{final Object s$elf = this} at the beginning of each method body \( BB \) see the fifth rule. The environment \( \tau \) is consequently extended to support an on-off mechanism which says when \( F[\tau] \) has to replace the self reference \( this \) with the reference to the object bound to variable \( s$elf \). In this way, in the translated program, \( this \) in the method \texttt{Apply}, refers to the function object implementing the closure and allows recursive invocations, while the object on which the method that defines the closure is invoked is reached through the reserved identifier \( s$elf \).

**Theorem 1** (this transparency). Occurrences of \( this \) in a closure are references to the object on which the method that defines the closure is invoked.

The rules are placed in Fig. 1 in an order which is quite near to the order of their use in translating a program containing closures (if we ignore the otherwise metarule that is omitted in the figure and should be placed at the top since it will be the most frequently used).
ClassBody:CB. Rule 1 applies to constructs of the category ClassBody and resets to empty, $\emptyset$, the environment of the translation of each MemberDecl construct which occurs in the ClassBody, since no closure can have such constructs in its scope.

Primary:Py. The rule 2 replaces the identifier with its value in $\tau$, while rules 3 and 4 replace a closure with the function object obtained instantiating the interface $I$nk on the types obtained translating each formal parameter type hence defining and translating the method invoke in the environment $\tau$ extended with the binding $(\text{this}/(s$elf,on))$, needed to guarantee the this transparency property.

ExtendedType:ET. Rules 5 and 6 replace the closure type with the interface type $I$nk instantiated on types obtained applying $\mathcal{F}[]$ to the formal parameter types. Two different cases are considered, depending whether a return type or void is defined.

LocalVariableDeclarationStatement:L Rule 7 considers the case in which a local variable is declared and initialized to a closure. In this case the closure is to be translated using a $\tau$ in which this is bound to the declared identifier to allow recursive invocations of the closure, case 2 above for $\tau$. Rule 8 considers the declaration of a shared variable. Such declaration is replaced with the declaration of an object of type $C$shared whose variable type is instantiated on the translated type of the declared variable. Rule 9 combine first and second case defining a shared closure.

BlockStatements:BB. The extension of the environment $\tau$ to replace the name $I$ of the shared variable with $I$.value is dealt with in rule 11.

MemberDecl:MB. Rule 10 considers the case of a constructor or method declaration. In this case the translation must add a declaration for the local variable $s$elf which must contain the reference to this, deal with shared formal parameters which are treated analogously to shared local variables adding for each shared parameter a local variable declaration whose type is $C$share instantiated to the translated type parameter. The name of the new variable is obtained translating the parameter name through function $G$. The rule also updates the environment $\tau$ in which the method body is translated, adding the binding for each formal parameter $I_i/G(I_i).value$ for the shared parameters and the identity for the non shared ones. Eventually the binding $(\text{this}/s$elf, off) is added if $\tau(\text{this})=\perp$, that is this is not bound in $\tau$. In fact, if $\tau$ contains a binding for this such binding can only be $(s$elf,on) which means that we are translating an invoke method of a function object representing a closure hence the binding for this is not to be modified.

Expression:E. Rule 12 deals with an assignment expression in case the expression on the righthand side is a closure. In such a case the identifier $I$ on the lefthand side is replaced by its binding in $\tau$ and the closure expression, analogously to local variable declaration, is translated using a $\tau$ in which $I$ is replaced by this.

As anticipated in Section 2, in [BO10] we define a minimal core calculus extending Featherweight Java [ABW01] with a form of closures which differ from the one adopted in Java$\Omega$ only syntactically. In that paper we define a reduction semantics for the calculus and prove type safety. Other interesting properties, like the abstraction property will be investigated in future works with the final goal to prove that the two semantics (reduction and translation semantics) commute.

5 Higher Order Methods: MC parameters.

The mechanism here defined is a revised version of the one described in [BO08a,BO09b]. It introduces constructs to define HO methods for both class (static) and object methods.
HO methods have at least one `mc_parameter` that is a parameter that is bound, during the invocation of the HO method, to a method of an arbitrary class or object of the program. Moreover, the parameter can be used, inside the body of the HO method, as if it were any method: It is invoked on a class or an object and applies to arguments of the right types for the bound method.

### 5.1 Syntax and semantics for `mc_parameters`

The main extensions are concerned with formal and actual parameters to include methods as parameters in the definition and invocation of HO methods: Hence, new type expressions for formal `mc_parameters` and new (value) expressions for the actual ones. A complete definition of the extended Java grammar is in Appendix A, while the syntactic categories directly affected are listed below.

\[
\begin{align*}
\text{ExtendedType} & \::= \text{Type} \mid \text{FType} \\
\text{FType} & \::= \text{Fun RootClass}: ([\text{ExtendedTypeList}] \rightarrow (\text{void} | \text{Type})) \text{ ThrowOpt} \\
\text{RootClass} & \::= \text{Identifier} \mid \text{ParameterizedType} \\
\text{ParameterizedType} & \::= \text{Identifier ParsOpt (}. \text{Identifier ParsOpt} {) ^+} \\
\text{ParsOpt} & \::= {[<\text{TypeArguments}>]} \\
\text{ThrowOpt} & \::= \{ \text{throws } \text{QualifiedIdentifier}^+ \}
\end{align*}
\]

`FType` is the new category that we add to Java grammar to extend the type system with a type for `mc_parameters`. The new type begins with the keyword `Fun` and specifies a class type and a signature for the methods that can be bound to the `mc_parameter`. The class type can be either an `Identifier`, namely a bound type variable, or a `ParameterizedType`. It specifies the root class of the hierarchy to which the classes, containing the methods that can be passed as parameter, belong. Then, the signature specifies number and types of the arguments of such methods and the type of the computed value, and possibly, of the throwable exceptions. Note that `FType` is not a sub-category of `Type`, which contains the types of ordinary Java, but `FType` and `Type` are both sub-categories of `ExtendedType`. This new category allows to constraint, already at syntactic level, the use of methods as values. In fact, methods can only be used as values of `mc_parameters`, hence passed as parameters to HO methods and to closures (see the definition of `FormalParameter` in Appendix A, where `ExtendedType` is used instead of `Type`). However, methods cannot be assigned to variables (or fields), nor can be returned as values computed by other methods (or closures) (see for instance, definition of `MethodOrConstructorDecl` where category `Type` is still used for the return type).

To deal with actual `mc_parameters` we extend Java expressions adding the syntactic category `AExp` below:

\[
\begin{align*}
\text{AExp} & \::= \text{Abs MethodSpecifier from Type} \\
\text{MethodSpecifier} & \::= \text{Identifier } ([\text{ExtendedTypeList}] \rightarrow (\text{void} | \text{Type})) \text{ ThrowOpt}
\end{align*}
\]

Hence, an actual `mc_parameter` is an expression that begins with the keyword `Abs` and computes a possibly overloaded method having name, signature and declaration class as specified in the expression. Given a program, we denote by \(\mathcal{M}(I, EL, T, TL, T_c)\) the method, if any, that has name \(I\), is either defined (possibly overridden) or inherited in class \(T_c\), is applicable (by subtyping, conversion and variable arity, see Section 15.12 [GJSB05]) to invocation with argument type list \(EL\), has return type \(T\) (possibly \text{void}),
and exception type list TL. If method \( \mathcal{M}(I, EL, T, TL, T_e) \) exists then such a method is unique (see Section 8.2 [GJSB05]). Otherwise \( \mathcal{M}(I, EL, T, TL, T_e) \) is the undefined \( \perp \).

The intended meaning of mc-parameter associates to:

(i) Expression \( \text{Abs} \ I \ (EL) \rightarrow T \ [\text{throws} \ TL] \) from \( T_a \), the partial function below:

\[
\lambda f : C \rightarrow \lambda c : C \rightarrow \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_c) \text{ if } c \preceq f \preceq T_a \land \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_a) \neq \perp
\]

that, given \( f \) and \( c \), selects the method \( \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_c) \), i.e. the method named \( I \) in class \( T_c \), that can be invoked with argument type list \( EL_{T_a} \), return type \( T \) and has exceptions \( TL_{T_a} \), where \( TL_{T_a} \) is included (see Section 8.4.6 in [GJSB05]) in \( TL_{T_a} \). If \( I \) is not overloaded in \( T_a \) then \( EL_{T_a} \equiv EL \). Otherwise, \( EL_{T_a} \) is the list of types computed in class \( T_a \) by Java overloading resolution (see Section 15.2 in [GJSB05]). If \( \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_c) \neq \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_a) \) then \( \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_c) \) is an overriding of \( \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_a) \). If \( \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_a) \) does not exist then the program is not legal.

(ii) Parameter \( \text{Fun} \ T_f : (EL_{T_f}) \rightarrow T_{T_f} \ [\text{throws} \ TL_{T_f}] \ p \), together with the argument \( \text{Abs} \ I \ (EL) \rightarrow T \ [\text{throws} \ TL] \) from \( T_a \) supplied for it, the partial function below:

\[
\lambda c : C \rightarrow \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_c) \text{ if } c \preceq T_f \preceq T_a \land \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_a) \neq \perp
\]

that, given a class \( c \), subclass of \( T_f \), selects method \( \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_c) \), provided that both following conditions hold: (a) \( T_f \preceq T_a \); (b) \( \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_a) \) is a contravariant [LW93, AC96, BO05] of the expected method, i.e. \( T \preceq T_f \land TL_{T_a} \sqsubseteq TL_{T_f} \land EL_{T_a} \sqsupseteq EL_{T_f} \).

(iii) Occurrence \( E_p(E_1,...,E_n) \) inside the body of a HO method invoked with argument \( \text{Abs} \ I \ (EL) \rightarrow T \ [\text{throws} \ TL] \) from \( T_a \) supplied for parameter \( \text{Fun} \ T_f : (EL_{T_f}) \rightarrow T_{T_f} \ [\text{throws} \ TL_{T_f}] \) \( p \), the invocation of method \( \mathcal{M}(I, EL_{T_a}, T, TL_{T_a}, T_c) \) where \( T_c \) is the effective (i.e. run-time class) class of the object computed by \( E \) and \( EL_{T_a} \) are correct types for the argument list \( E_1,...,E_n \).

The intended meaning states the following relevant semantics properties:

- The method bound to a mc-parameter has: a) the effective signature that results resolving overloading in the root class \( T_a \), and b) the effective definition that is found in the class \( T_c \) of the object on which the mc-parameter is invoked (see Java run-time method dispatch in Section 15.12.2.5 in [GJSB05]);
- the class \( T_c \) is a subclass of \( T_f \) which must be a subclass of \( T_a \);
- the method effectively invoked in the invocation of an mc-parameter is a contravariant of the method expected. Then invocation either computes a value which is subtype of the value expected or throws one exception among those expected. Moreover, invocation applies to arguments that have types that are subtypes of the signature of the effectively invoked method. Hence, in all the cases, invocation behaves well with any program context in which the invocation of a mc-parameter occurs and which is correct with respect to Java ordinary type checking.

\( \preceq (\text{and} \geq) \) is the Java subtyping relation (see Section 4.10 [GJSB05]) possibly extended on lists of types in the obvious way, i.e. \( t_1 \ldots t_n \preceq s_1 \ldots s_n \) means \( t_1 \preceq s_1 \land \ldots \land t_n \preceq s_n \). Eventually, \( t_1 \ldots t_n \sqsubseteq s_1 \ldots s_k \) means \( (\exists t_1 \preceq u_1, \ldots, t_n \preceq u_n) \{u_1, \ldots, u_n\} \sqsubseteq \{s_1, \ldots, s_k\} \), where \( \sqsubseteq \) is set inclusion.
The translation semantics defined in this paper is a constructive formalization of this semantics using Java structures. It replaces code containing definitions and invocations of \texttt{mc} parameters with code that contains the invocations of the methods that are effectively bound to such \texttt{mc} parameters according to the semantics. Moreover, it is based on the computational structures of the callback methodology [GHJV05].

5.2 The callback methodology: A semi-formal description

Callback is a way to pass executable code to procedures in event driven programming. In OO languages callback is implemented through function objects [Mey04,Hor07] that wrap a code in order to treat it as an ordinary datum that can be passed anywhere in the program and unwrap it when the code must be executed. The methodology, here discussed for objects wrapping methods, \textit{method objects}, can be summarized in four points.

1. \textbf{Interfaces for method objects}. We introduce an interface representing classes of objects wrapping methods, that can be passed, in the program, as parameters of \texttt{HO} methods. In fact, we need a family of interfaces \texttt{ApplyClass$\langle n,k \rangle$}, where the suffices \texttt{n} and \texttt{k} stand for the number of arguments and for the number of different types of throwable exceptions of the wrapped methods for which the interface is defined. Each interface has only one method \texttt{Apply} whose first parameter is an object, namely the object on which the wrapped method must be invoked, while the following \texttt{n} parameters are the \texttt{n} parameters to which the wrapped method applies. Eventually, \texttt{Apply} throws the same \texttt{k} different classes of exceptions of the wrapped method. Interface \texttt{ApplyClass$\langle n,k \rangle$} has \texttt{n+k+2} type variables: In addition to the \texttt{n} type variables for the types of the arguments and the \texttt{k} for those of the throwable exceptions, it has one type variable for the class of objects on which the wrapped method may be invoked, and one type variable for the type of the result that the wrapped method computes. Hence, interface \texttt{ApplyClassV$\langle n,k \rangle$} has \texttt{n+k+1} type variables and is the analog of \texttt{ApplyClass$\langle n,k \rangle$} for \texttt{void} wrapped methods.

   public interface \texttt{ApplyClass$\langle n,k \rangle$}\texttt{<RT,ET$_1$,...,ET$_n$,T,}
   \texttt{QI$_1$ extends Throwable,..,QI$_k$ extends Throwable>}
   \texttt{public \texttt{T} \texttt{apply(RTo,ET$_1$x$_1$,...,ET$_n$x$_n$) throws QI$_1$,..,QI$_k$;}}

   public interface \texttt{ApplyClassV$\langle n,k \rangle$}\texttt{<RT,ET$_1$,...,ET$_n$,QI$_1$ extends Throwable,..,QI$_k$ extends Throwable>}
   \texttt{apply(RTo,ET$_1$x$_1$,...,ET$_n$x$_n$) throws QI$_1$,..,QI$_k$;}

2. \textbf{Classes of method objects}. For each use of the method, which is to be passed as parameter, a class, which implements \texttt{ApplyClass$\langle n,k \rangle$} (resp. \texttt{ApplyClassV$\langle n,k \rangle$}, if it is a \texttt{void} method) and consequently \texttt{Apply}, must be defined. \texttt{Apply} invokes the method on the object passed to it as first argument, with the arguments passed in the rest of its arguments list. Suppose we have \texttt{C}$\langle T_{a_1}$,$\ldots$,\texttt{T}_{a_p} \rangle$ (where \texttt{T}_{a_1}$,$\ldots$,\texttt{T}_{a_p} are the type arguments supplied to the generic class \texttt{C}$\langle Q_1$,$\ldots$,\texttt{Q}_p \rangle$, of the program, with type parameters \texttt{Q}_1$,$\ldots$,\texttt{Q}_p) having methods named \texttt{m}_1$,$\ldots$,\texttt{m}_q, which are to be passed as parameters in the program. Then \texttt{q} classes are defined. In particular, for each, possibly generic, \texttt{m}_i, let \texttt{ET}_{ami}_1$,$\ldots$,\texttt{ET}_{ami}_n be the types of the arguments to which the method applies in the invocation, once passed as parameter, let \texttt{T}_{ami}
be the invocation return type, \( T_{tmi_1}, \ldots, T_{tmi_{kmi}} \) be the list of exception classes that invocation can throws. Then, a class \( I[m_i] \) is defined as follows:

\[
\text{static class } I[m_i] \text{ implements } \text{ApplyClass}\$nk_{mi} \langle C < T_{a_1}, \ldots, T_{a_p} >, ET_{ami_1}, \ldots, ET_{ami_n}, T_{tm_{i_1}}, \ldots, T_{tm_{i_{kmi}}} \rangle \{
\text{public } T_{mi} \text{ Apply}(C<T_{a_1}, \ldots, T_{a_p}> o, ET_{ami_1}, \ldots, ET_{ami_n}, x_n) \text{ throws } T_{tm_{i_1}}, \ldots, T_{tm_{i_{kmi}}} \{
\text{return } (o.m_i(x_1, \ldots, x_n));}
\}
\]

3. **HO methods.** Every HO method \( hm \) is defined having one \( \text{ApplyClass}\$nk \) parameter \( o \) for each object wrapping one of the method \( m_i \) which will be bound to it during an invocation of the HO method. This is expressed as below

\[
\text{public } ET_{hm} \text{ hm } (\ldots \text{ApplyClass}\$nk < ET_E, ET_1, \ldots, ET_n, T, T_1, \ldots, T_k > o \ldots)
\{
\text{... } o.\text{Apply}(E, E_1, ..., E_n) \ldots \}
\]

where \( E \) is the expression on which one of the methods \( m_i \) must be invoked and \( E_j (j \in [1, n_i]) \) are the arguments of \( m_i \).

4. **Creation of method objects.** The invocation of \( hm \) requires the construction of one object of the class \( I[m_i] \). This is accomplished in the invocation below, under the assumption, in point 2, that class \( I[m_i] \) is defined as an inner class of class \( C \):

\[
\ldots E.hm(\ldots \text{new } C.I[m_i]() \ldots)
\]

Point 4 concludes the description of the callback methodology that has been presented in the case that the passed methods are void and non void object and class (fixed arity) methods. The latter ones are considered only if the methods are invoked on objects of the class. If the passed class method is invoked directly on the identifier of the class: point 1 requires a specific interface \( \text{ApplyClass}\$nk \) (resp. \( \text{ApplyClassSV}\$nk \)) without the first parameter; point 2 has method \( \text{Apply} \) without first argument and the occurrence of \( o \) in the body replaced by the name of the class that contains the passed static method; point 3 has invocation of \( \text{Apply} \) without first argument. Also, if the passed method is a variable arity method interfaces change and point 1-3 are modified to consider the use of the syntactic form ‘...’ for the occurrence of a variable arity parameter (see Section 8.4.1 [GJSB05]). Moreover, different solutions can be considered for the definition of the classes of objects wrapping a method, given in point 2. In the presentation given above we use inner classes but anonymous classes and stand alone classes offer alternative supports for this aim with different advantages as discussed in [BO09b]. In particular, the use of anonymous classes avoids the need of introducing names for the classes (as for \( I[m_i] \) in point 2 of the methodology.). Hence, unlike the one presented in [BO09b], the translation presented in next section uses anonymous classes and combines points 2 and 4.

---

2 \( I[m_i] \) is a class identifier which must be unique for each use of method \( m_i \) as parameter in the HO methods of the program. The term use does not correspond exactly to an invocation of the method since different invocations of \( m_i \) can have the same argument type list, return type, exception type list: hence can share the same class \( I[m_i] \). Then, the specific name chosen for class \( I[m_i] \) depends from the method identifier, list of types used in the invocation and class in which it is defined, and by the technique used for the definition of the class: inner, anonymous or stand alone.
5.3 $\mathcal{E}[]\rho$: A Callback based, Translation Semantics of mc_ parameters

The translation semantics is defined using the syntax directed rule system described in Section 3. Hence, it has at least one rule for each production of the unambiguous grammar of Appendix A, and the translation computes applying to each program construct the rule that is associated to the production with which the construct is parsed. However, the rule system $\mathcal{E}[]\rho$ contains only 8 significant rules and the otherwise metarule of Section 3. Hence we illustrate the translation discussing rule by rule (only for the significant ones) how the rules work. However, firstly we introduce the parameter $\rho$ used in the translation $\mathcal{E}[]\rho$. This parameter is an environment for the application of the translation to a construct: It contains one binding for each parameter (of constructor or method) that has the considered construct in its (lexical) scope. The empty environment is expressed by $\emptyset$ and does not contain bindings. Any other environment is obtained by $R[T I]_\rho$ that adds a new binding $[ T I]$ to a given environment $\rho$ (possibly, the empty environment): The new binding associates to the name $I$ a type $T$. Environments are used in the translation in order to discover which identifiers, in the considered program construct, are parameters that are bound to a FType, hence are mc_ parameters (see the two rules 7 and 8 for PrimarySelector in Fig. 2).

CompilationUnit:CU. Rule 1 applies to constructs of the category CompilationUnit and resets to empty, $\emptyset$, the environment of the translation of each ClassOrInterfaceDeclaration construct which occurs in the CompilationUnit, since no method parameter can have such constructs in the scope.

MethodOrConstructorDecl:MB. Rule 2 extends the environment of the translation of the block with the method parameters, since the block is the method body and such parameters have it in their scope.

ExtendedType:ET. Rules 3 and 4 apply to type expressions of the new category FType and replace the type with the type interface ApplyClass$nk$ and ApplyClassV$nk$ used to model method objects as described in Section 5.2, point 1.

Expression:E. Rules 5 and 6 apply to the value expressions AExp that compute an actual mc_ parameter. The rules replace the expression with an instance creation expression of an anonymous class. Similarly to what we did in Section 5.2 using inner classes, the two rules insert the code to create an anonymous class that implements the appropriate interface, (point 2), and, at the same time, to create a method object that instantiates such a class, (point 4). Hence, rule 5 considers each expression Abs I EL→T[throws TL] from $T_c$ computing a non void method. It replaces the expression with the code that creates, (point 4), a method object wrapping a method that has name $I$, belongs to a class in the hierarchy rooted at class $T_c$, applies to argument type list $TL$ and has return type $T$. Analogously the rule 6 creates a void method object.

PrimarySelector:PS. The last two rules, 7 and 8, apply to value expressions of the category PrimarySelector. This category reformulates, in an unambiguous way, the syntactic structures defined, in the grammar of JSL [GJSB05], combining the two categories Primary and Selector. The two rules consider each invocation $PS . I A$ where a formal mc_ parameter $I$ is invoked on expression $PS$ and applies to the arguments in $A$. Since, in the translated program, $I$ is bound to an object wrapping the method to be invoked, the rules replace the invocation with an invocation on $I$ of method Apply that applies to the argument list containing the translation of $PS$ followed by the translations of the arguments in $A$. The translation of the arguments
as well as that of the entire invocation, in rule 8, are cast to the expected types, in order to guarantee the contracovariance of the wrapped method (see Theorem 2).

The following properties hold for: (i) the translation semantics of mc\_parameters, (ii) the rule system with which the semantics is defined, (iii) the programs obtained applying the rule system. The properties assert that the wrapped methods are essentially the methods of the intended semantics of mc\_parameters. Existence and access of the methods passed using mc\_parameter in the source program can be checked, at compile time, checking existence and access of the wrapped methods in the translated program. The methods that are passed by mc\_parameters are checked to be contracovariant with the expected methods, hence their invocation is type safe. Eventually, the execution of the translations of programs extended with mc\_parameters never goes in unexpected, unrecoverable, computation states.

**Theorem 2 (Wrapped Methods for MC\_parameters).** Let $E \equiv \text{Abs } I \ E_{L_a} \rightarrow T \ [\text{throws } TL]$ from $T_a$ in the scope of an environment $\rho$, and $EL \equiv E[EL_{a}]_{\rho}$. Let $M$ be the wrapped method of the object, if any, created executing $E[EL_{a}]_{\rho}$. Then $M$ is (possibly an overriding of) $M(I,EL_{T_a},T,TL_{T_a})$, where $EL_{T_a}$ and $TL_{T_a}$ are the overloading resolution, in class $T_a$, of an invocation of a method with name $I$, argument type list $EL$, return type $T$ and exception type list $TL$.

**Lemma 1 (Existence and Access of MC\_parameters).** Let $E \equiv \text{Abs } I \ E_{L_a} \rightarrow T \ [\text{throws } TL]$ from $T_a$ in the scope of an environment $\rho$, and $EL \equiv E[EL_{a}]_{\rho}$. Let $M(I,EL_{T_a},T,TL_{T_a})$ be the method that results, for any given $T_c \preceq T_f \preceq T_a$, from the intended semantics of $E$. Then $M(I,EL_{T_a},T,TL_{T_a})$ (i) exists if and only if $E[EL_{a}]_{\rho}$ has no compile-time type errors; (ii) can be invoked only if the $E[EL_{a}]_{\rho}$ has no method access violations.

**Lemma 2 (MC\_parameters are Contracovariant).** Let $p$ be a mc\_parameter of a HO method (in a source program) and $\text{Fun } RT_p : EL_p \rightarrow T_p[\text{throws } TL]$ be its type with $EL_p \equiv ET_{p_1},...,ET_{p_n}$. Let $o$ be a binding for $p$ in an invocation of the HO method in the translated program obtained by $E[\bot]$. Then, $o$ is an object wrapping a method which is contracovariant with $EL_p \rightarrow T_p$, i.e. has return type which is a subtype of $T_p$ and type signature $ET_{o_1},...,ET_{o_n}$ such that $ET_{o_i}$ is a supertype of $ET_{p_i}$ ($\forall i \in [1..n]$).

**Theorem 3 (Type Safety).** Let $P$ be any Java program extended with mc\_parameters. Then $P$ is type safe only if its translation $E[P]_{\bot}$ is a type safe program (of Java), i.e. (i) it is well-typed and, (ii) its execution does not generate untrapped errors.

### 6 Examples

A first example is a classical problem for higher order programming already discussed in [BO04], but here modified to use generics an classes in Java APIs such as `LinkedList` and `ListIterator`. The example defines an abstract class `Shape` for geometric shapes and several concrete classes `Rectangle`, `Circle`, `Triangle` etc. and an extension of `LinkedList` in which a HO method `map` is defined. The HO method is used to compute the list of areas of all the shape contained in a list, even though the the method to compute the area is a different one according to the type of the element in the list that is being processed. A second example, taken from [Goe07], defines, in Fig.4.a, a method `memoized` that maps closures of type `\{(T_i) : T\}` into `memoized` closures of the
\[
\begin{align*}
\mathcal{E}[\text{CU}]_\rho &= \text{package } QI_1; \ IM \ \mathcal{E}[\text{CD}_1]_\rho \ldots \mathcal{E}[\text{CD}_n]_\rho \\
\text{with } \text{CU} &= \text{package } QI_1; \ IM \ \text{CD}_1 \ldots \text{CD}_n \\
\mathcal{E}[\text{MB}]_\rho &= D \ [Q] \ [TV] \ I \ \mathcal{E}[\text{FP}]_\rho \ O_o \ TR \ \mathcal{E}[\text{B}]_\rho \text{FP} \text{with } \text{MB} \in \text{MethodOrConstructorDecl} \\
\mathcal{E}[\text{ET}]_\rho &= \begin{cases} 
\text{ApplyClass}\$nk <\mathcal{E}[\text{RT}]_\rho, \mathcal{E}[\text{ET}_1]_\rho, \ldots, \mathcal{E}[\text{ET}_n]_\rho, \mathcal{E}[\text{QI}_1]_\rho, \ldots, \mathcal{E}[\text{QI}_k]_\rho > \\
\text{with } \text{ET} = \text{Fun } \text{RT}: (\langle \text{ET}_1, \ldots, \text{ET}_n \rangle) \rightarrow T_o \ [\text{throws } QI_1, \ldots, QI_k] \\
\text{ApplyClassV}\$nk <\mathcal{E}[\text{RT}]_\rho, \mathcal{E}[\text{ET}_1]_\rho, \ldots, \mathcal{E}[\text{ET}_n]_\rho, \mathcal{E}[\text{QI}_1]_\rho, \ldots, \mathcal{E}[\text{QI}_k]_\rho > \\
\text{with } \text{ET} = \text{Fun } \text{RT}: (\langle \text{ET}_1, \ldots, \text{ET}_n \rangle) \rightarrow \text{void } [\text{throws } QI_1, \ldots, QI_k] \\
\end{cases} \\
\mathcal{E}[\text{ET}]_\rho &= \begin{cases} 
\text{new ApplyClass}\$nk <\text{TL}_o>() \text{with } E = \text{Abs } I(\mid E I) \rightarrow T \ [\text{throws } \text{TL}]/ \text{from } T_o \\
\text{public } T \ \text{Apply } \text{FP}_r/\text{throws } \text{TL}() \{ \\
\text{return } \text{o.I}(x_1, \ldots, x_n) \} \wedge \ \text{EL} = \text{ET}_1, \ldots, \text{ET}_n \\
\text{where } \\
\text{TL}_r = T_o.\mathcal{E}[\text{ET}_1], \ldots, \mathcal{E}[\text{ET}_n], T, QI_1, \ldots, QI_k \\
\text{FP}_r = (T_o.\mathcal{E}[\text{ET}_1] x_1, \ldots, \mathcal{E}[\text{ET}_n] x_n) \\
\end{cases} \\
\mathcal{E}[\text{ET}]_\rho &= \begin{cases} 
\text{new ApplyClassV}\$nk <\text{TL}_o>() \text{with } E = \text{Abs } I(\mid E I) \rightarrow \text{void } [\text{throws } \text{TL}]/ \text{from } T_o \\
\text{public } \text{void } \text{Apply } \text{FP}_r/\text{throws } \text{TL}() \{ \\
\text{return } \text{o.I}(x_1, \ldots, x_n) \} \wedge \ \text{TL} = QI_1, \ldots, QI_k \\
\text{where } \\
\text{TL}_r = T_o.\mathcal{E}[\text{ET}_1], \ldots, \mathcal{E}[\text{ET}_n], QI_1, \ldots, QI_k \\
\text{FP}_r = (T_o.\mathcal{E}[\text{ET}_1] x_1, \ldots, \mathcal{E}[\text{ET}_n] x_n) \\
\end{cases} \\
\mathcal{E}[\text{PS}]_\rho &= \begin{cases} 
(T)(\text{I.Apply(}\mathcal{E}[\text{PS}_1]_\rho, \mathcal{E}[\text{ET}_1]_\rho, \ldots, \mathcal{E}[\text{ET}_n]_\rho, \mathcal{E}[\text{En}_1]_\rho)) \\
\text{with } PS = PS_1. I(\langle E_1, \ldots, E_n \rangle) \wedge \rho(I) = \text{fun } R C : \text{ET}_1, \ldots, \text{ET}_n \rightarrow T \ TR \\
\text{I.Apply(}\mathcal{E}[\text{PS}_1]_\rho, \mathcal{E}[\text{ET}_1]_\rho, \ldots, \mathcal{E}[\text{ET}_n]_\rho, \mathcal{E}[\text{En}_1]_\rho) \\
\text{with } PS = PS_1. I(\langle E_1, \ldots, E_n \rangle) \wedge \rho(I) = \text{fun } R C : \text{ET}_1, \ldots, \text{ET}_n \rightarrow \text{void } \ TR \\
\end{cases} \\
\text{where: } R[T I]_\rho(x) = T \quad \text{if } I = x \wedge T \in \text{Ftype} \\
\text{R[T I]_\rho(x) = } \bot \quad \text{if } I = x \wedge T \notin \text{Ftype} \\
\text{R[T I]_\rho(x) = } \rho(x) \quad \text{if } I \neq x \\
\theta(x) = \bot \\
\rho \uparrow (\cdot) = \rho \\
\rho \uparrow (P_1 P_2 \ldots P_n) = R[T E_1 I_1]_\rho(P_2 \ldots P_n) \quad \text{with } P_i = [\text{final }] \ AT_i \ ET_i \ I_i([\cdot])^* \\
\text{Legend: For arbitrary index } p, \ \wedge \text{Arguments, } AR \in \text{ArrayCreatorRest}, \ AT \in \text{Annotations,} \\
\text{B,Bp} \in \text{Block}, \ \text{BB, BBp} \in \text{BlockStatements}, \ \text{BS, BS}_p \in \text{BlockStatement}, \ \text{BT} \in \text{BasicType,} \\
\text{CB} \in \text{ClassBody}, \ \text{CD, CDp} \in \text{ClassOrInterfaceDeclaration}, \ \text{CN} \in \text{CreatedName}, \ \text{CR} \in \text{ClassCreatorRest}, \ \text{CU} \in \text{CompilationUnit}, \ \text{D} \in \text{ModifierOpt}, \ \text{E} \in \text{Expression}, \ \text{EL} \in \text{ExtendedTypeList,} \\
\text{ET, ET}_p \in \text{ExtendedType}, \ \text{F} \in \text{ETs}, \ \text{FP, FP}_p \in \text{FParameters}, \ \text{FS, FS}_p \in [\text{final } | \text{shared}], \ \text{GS} \in \text{ExplicitGenericInvocationSuffix}, \ \text{I, Ip} \in \text{Identifier}, \ \text{IM} \in \text{ImportDeclaration}, \ \text{IO} \in \text{InfixOp}, \ \text{K} \in \text{Literal}, \ \text{L} \in \text{LocalVariableDeclarationStatement,} \\
\text{MB, MBp} \in \text{MemberDecl, NI} \in \text{NonInvocationSelector, O, O} \in [\cdot]^*, \ P, P_p \in \text{FormalParameter,} \\
\text{PO} \in \text{PostfixOp}, \ \text{PS, PS}_p \in \text{PrimarySelector}, \ \text{Py} \in \text{Primary}, \ \text{Q, Qp} \in \text{TPs, QI} \in \text{QualifiedIdentifier,} \\
\text{RT} \in \text{RootClass, S, Sp} \in \text{Statement}, \ \text{Sp} \in \text{Selector, T, Tp} \in \text{Type,} \\
\text{Tq, Tq} \in \text{TypeArgument, TA} \in \text{ParamsOpt, TL, TLp} \in \text{TypeList, TR} \in \text{ThrowOpt, TV, TV}_p \in [\text{Type } | \text{void}] \ U, U_p \in \text{ST}, \ \text{V, V}_p \in \text{VariableDeclarationId,} \\
\end{align*}
\]
public abstract class Shape {
    public abstract Double Area();
    public abstract Double Perimeter();
}

public class Rectangle extends Shape {
    private double base;
    private double height;
    public Rectangle(double b, double h){base=b;height=h;}
    public Double Perimeter() {return new Double(2*base+2*height);}
    public Double Area() {return new Double(base*height);}
}

public class Circle extends Shape {
    private double radius;
    public Circle(double r){radius=r;}
    public Double Perimeter() {return new Double(2*3.14*radius);}
    public Double Area() {return new Double(3.14*radius*radius);}
}

public class Triangle extends Shape {...}

public class FList <C> extends LinkedList <C> {
    public <T> FList<T> Map(Fun C:() -> T s){
        FList<T> L=new FList<T>();
        for (C g: this) L.add(g.s());
        return L;}
}

public class Compute {
    public static void main (String [] args){
        FList <Shape> L= new FList<Shape>();
        ... L.add(new Rectangle(3.0,4.2)); L.add(new Circle(1.5));...
        ... L.Map(Abs Area ()->Double from Shape); }}

Fig. 3. a. HO Programming with a Class of Geometric Shapes: P

// first 15 lines are unchanged

public class FList <C> extends LinkedList <C> {
    public <T> FList<T> Map(ApplyClass$00<C,T> s){
        FList<T> L=new FList<T>();
        for (C h: this) L.add((T)(s.Apply(h)));
        return L;}
}

public class Compute {
    public static void main (String [] args){
        FList <Shape> L= new FList<Shape>();
        ... L.add(new Rectangle(3.0,4.2)); L.add(new Circle(1.5));...
        ... L.Map(new ApplyClass$00<Shape,Double>(){
            public Double Apply (Shape o){
                return o.Area();}});
    }
}

public interface ApplyClass$00 <T,S> {
    public S Apply(T m);}

Fig. 3. b. HO Programming: The Translated Program – $\mathcal{E}[[P]]_\rho$
same type and same external behavior but of different performance since a memoized closure avoids repeating computation for previously processed inputs. To obtain this, memoized has a closure parameter $f$, declared final, of type $\{(T_1) : T\}$, bound to the closure to be memoized, and a shared variable $table$ initialized to an empty hash table. memoized returns a closure that behaves as follows. It has a parameter $x$ of type $T_1$ and asks table $table$ for a key equals to the value of $x$. If the key is found then it returns the value bound in table to $x$, otherwise it behaves like $f$, invoking $f.invoke(x)$, updates table, adding the key/value pair $x/f.invoke(x)$, returns the value $f.invoke(x)$.

Note that, in this way, each hash table table is a private resource of the memoized closure since each time memoized returns the closure and ends, thus making table no more accessible outside the returned closure. Eventually, note that such a solution works properly for closures without the recursion operator. For recursive closures the reader is invited to consider the code for the Fibonacci numbers in Fig. 5, whose translation and execution are left to the reader:

```java
public class Memoize<T1,T>{
    public {\{T1\}:T} memoized(final {\{T1\}:T} f){//maps closures into memoized closures
        shared Hashtable<T1,T> table = new Hashtable<T1,T>();
        {\{T1\}:T} memo_f = {\{(T1 x): T \Rightarrow T res = table.get(x);
            if (res == null){res = f.invoke(x); table.put(x,res);}
            return res;}};
        return memo_f;}}
```

**Fig. 4. a.** Class Memoize: A Method that Memoizes Closures

```java
public class Memoize<T1,T>{
    public I$1<T1,T> memoized(final I$1<T1,T> f){//maps closures into memoized ... final Shared<Hashtable<T1,T>> table =
        new Shared<Hashtable<T1,T>>(new Hashtable<T1,T>());
    I$1<T1,T> memo_f = new I$1<T1,T>(){
        public T invoke(T1 x){T res = table.get(x);
            if (res == null){res = f.invoke(x); table.put(x,res);}
            return res;};
        return memo_f;}}
```

**Fig. 4. b.** Class Memoize: The Translated Program – $F[Memoize]$

```java
shared {(Integer):Integer} Yfib;
{(Integer):Integer}fib = {((Integer x):Integer \Rightarrow if (x==0 || x==1) return 1;
                else Yfib.invoke(x-1)+ Yfib.invoke(x-2))}
Yfib = new Memoize<Integer,Integer>().Memoized(fib);
// Yfib = fib; without memoization
```

**Fig. 5.** Class Memoize: Memoizing Fibonacci number
7 Putting Translations Together

We have introduced and discussed, in a separate way, the extension of Java 1.5 with mc_parameters and the extension of Java 1.5 with closures. For each extension, we have followed the same methodology and used the same techniques, obtaining two separate translation semantics. One of the most interesting aspects of our project is how the two extensions can be integrated in order to obtain only one language which extends Java with both mc_parameters and closures. Before delving into this aspect, we consider parameters and the extension of Java 1.5 with closures. For each extension, we have introduced and discussed, in a separate way, the extension of Java 1.5 with mc_parameters and closures. Before delving into this aspect, we consider parameters, and vice versa, HO methods having closures.

Case 1. Only one, between $\rho$ and $\tau$, is defined on $p$. This is the case for rules 3, 4, 5, 6, 7, 8 of $\mathcal{E}[P]_\rho$, and rules 2, 3, 4, 5, 6, 7, 8, 9, 11, 12 of $\mathcal{F}[P]_\tau$. All these rules become rules of $\mathcal{E}[P]_{\rho,\tau}$ when each occurrence of $\mathcal{E}[U]_\rho$ and $\mathcal{F}[U]_\tau$ are replaced by $\mathcal{E}[U]_{\rho,\tau}$ for any argument $U$.

Case 2. $\mathcal{F}[A_0]_{\tau} = r_i(A_1, \mathcal{F}[A_1], \ldots, A_n, \mathcal{F}[A_n])_\tau$ with $b_i(A_0, A_1, \ldots, A_n)$
$\mathcal{E}[A_0]_\rho = f(A_1, \mathcal{E}[A_1]_\rho), \ldots, f(A_n, \mathcal{E}[A_n]_\rho)$ with $b_i(A_0, A_1, \ldots, A_n)$
(Symmetric case: changing $\mathcal{F}[P]_\tau$ with $\mathcal{E}[P]_\rho$.) This is the case of the rule 1 of $\mathcal{E}[P]_\rho$, and of the rule 1 of $\mathcal{F}[P]_\tau$. In fact, both rules would be otherwise rules if they were not for

Theorem 4 (Completeness I). Translation $\mathcal{E}[P]_\emptyset$ is complete, i.e. it maps each program of Java extended with closures into an equivalent program of ordinary Java 1.5.

Theorem 5 (Completeness II). Translation $\mathcal{F}[P]_\emptyset$ is complete.

The following property is concerned with the two translations together and says that they are one another orthogonal: Hence one definition does not affect the other one. In particular, modifications on one translation could be treated without any consideration on the other one.

Theorem 6 (Orthogonality). The composition of the two translations is commutative, i.e.: $\mathcal{F}[\mathcal{E}[P]_\emptyset]_\emptyset = \mathcal{E}[\mathcal{F}[P]_\emptyset]_\emptyset$
Theorem 7 (Completeness). Translation $\mathcal{EF}_{\varnothing, 0}$ is complete: It maps each program of the extended language onto an equivalent program of ordinary Java 1.5.

8 Conclusions

In this work (chapter) we have described two HO mechanisms: mc_parameters and closures, to extend Java, defining a translation semantics and proving some fundamental properties including completeness and orthogonality. The implementation of the extended language, along the lines already experimented in JavaΩ [BO07a,BO08b], is immediate: The translation semantics $\mathcal{EF}_{\rho, \tau}$ can be formally converted into a source-to-source translation and implemented as a one-pass preprocessor [ALSU07] that is developed using Lex & Yacc [LM95] and GNU Bison [CS06a]. This implementation allows to quickly develop a prototype that can be used to test the programming features of the extended language and runs in combination with any Java compiler, including Java of SUN, GCJ of GNU, ECJ of Eclipse. Moreover, for competitive compilations of the extended language, the compiler modifications can be obtained, in a straightforward way, from the translation semantics: The abstract syntax generation phase of the compiler parser creates the abstract tree of $\mathcal{EF}_u$ in correspondence of the parsing of any language structure $u$, for suitable contexts (producing the environments) $\rho$ and $\tau$. We are currently investigating the possibility to apply software engineering techniques and Java annotations to support error localization in non-native constructs. The aim is that errors in a non-native construct of a program, for instance a closure type, found by the Java compiler during the analysis of the code, obtained by program preprocessing, are recognized as errors of the construct and localized (for possibly error recovery) on the source program, i.e. the program before preprocessing. Moreover, as mentioned in section 4, we are still investigating the definition of core Java sub-languages with reduction semantics to obtain different modelizations of a same construct, for instance closures, to experiment the adequacy to the initial aims and to prove properties of each modelization and eventually prove that all such properties are preserved in the full language.
extended with the new construct, when the reduction semantics of the sub-language and the translation semantics of the extended language commute.

A Appendix - An Unambiguous Grammar for Java

The syntactic categories that are used but not defined are those in chapter 18 of the Java JSL [GJSB05]. We use the same syntactic conventions of JSL, in particular for the alternation operator ($a|b$ means one occurrence of either $a$ or $b$) and for option operator ($[a]$ means zero or one occurrence of $a$).

ClassDeclaration ::= public class Identifier [TPs] [ST] [ITs] ClassBody
TPs ::= <TypeParameter (, TypeParameter)*>  ST ::= extends Type
ITs ::= implements TypeList
ClassBody ::= { (MemberDecl)* }
MemberDecl ::= ;
| ModsOpt FieldDeclarator
| MethodOrConstructorDecl
| ModsOpt [TPs] ClassOrInterfaceDeclaration
| [static] Block
MethodOrConstructorDecl ::= ModsOpt [TPs] [(Type | void) Identifier FParameters []* ThrowOpt Block]
ModsOpt ::= Modifier*
ThrowOpt ::= [throws QualifiedIdentifier+]
ExtendedType ::= Type | FType
Type ::= ParameterizedType | BasicType | ClosType
ExtendedTypeList ::= [ExtendedType [( , ExtendedType)]]
FParameters ::= ([Final | shared] [Annotations] ExtendedType VariableDeclaratorId
ParameterizedType ::= Identifier ParsOpt (. Identifier ParsOpt)[]* [*]
FType ::= Fun RootClass([ExtendedTypeList]) -> (void | Type) ThrowOpt
RootClass ::= Identifier | ParameterizedType
ParsOpt ::= [<TypeArguments+>]*
Expression ::= AExp | Expression1 [AssignmentOperator Expression1]
Expression3 ::= PrefixOp Expression3
| PrimarySelector [PostfixOperator]
PrimarySelector ::= PrimarySelector NonInvocationSelector
| PrimarySelector . [TypeList] Identifier Arguments
| Primary
Primary ::= ParExpression
TypeList ( ExplicitGenericInvocationSuffix | this Arguments )
| QualifiedIdentifier . this Arguments
| QualifiedIdentifier . super (Arguments | . Identifier [ Arguments ])
Literal
| QualifiedIdentifier . new TypeList CreatedName ( Arguments | ClassCreatorRest )
QualifiedIdentifier [ ([]) [] .class | [E] | Arguments]
BasicType [] .class
void.class
Closure
Closure::={|Parameters:(Type | void) Throwing| Block
AExp::= Abs MethodSpecifer from Type
MethodSpecifer::= Identifier ([ExtendedTypeList]) | (void | Type) Throwing
LocalVariableDeclaration::= | Final | Shared | Type VariableDeclarators
NonInvocationSelector::= , this
  super (Arguments | , Identifier [ Arguments ])
  new typeList CreatedName ( Arguments | ClassCreatorRest )
  [Expression]
  invoke [Arguments]

B Appendix - Proofs

Proof. Theorem 1. Entering a method, τ is extended with a binding of the fictitious identifier this with the pair (s$self, off) where s$self is the name of a fresh variable introduced in the assignment above, and off is a flag specifying that the binding for this is not active. In fact, the binding is set active when $F[]$ traverses a closure updating τ to set the flag to on, thus making the binding for this active when $F[]$ traverses a closure updating τ to set the flag to on thus making the the binding for this active in the third and fourth rule.

Proof. Theorem 2. Assume that execution of $E[E]_\rho$ creates an object. By rule 5 (6) of $E[E]_\rho$, the created object wraps a method that can be invoked by expression $o.I(x_1,...,x_n)$. Hence, the method must be invoked on objects of the class of $o$, namely (any subclass of) $T_a$, has the type signature that results from the overloading solution for a method named $I$, in the class $T_a$, of the type list of $x_1,...,x_n$. This type list is $EL$: Let $EL_{T_a}$ be its overloading solution (possibly, $EL_{T_a} \equiv EL$). Again, the wrapped method must compute a value of type $T$, since the expression is the argument of a return statement. Eventually, the method may throw exceptions of types included in the types of the list $TL_a$, since such a return statement is the body of an Apply method. Hence, the wrapped method is a possibly overriding method of a class $T_c \leq T_a$, has name $I$, arguments type list $EL_{T_a}$ and throwable types $TL_{T_a}$, and return type $T$. This completes the proof.

Proof. Lemma 1. (i)-if part. In this case $E[E]_\rho$ has no type errors. Hence, $E[E]_\rho$ execution creates an object on which Theorem 2 holds for a wrapped method. Such a method is, possibly an overriding of, $M(I,EL_{T_a},T,TL_a,T_a)$ for any class $T_c$ such that $T_c \leq T_f \leq T_a$, namely $M(I,EL_{T_a},T,TL_c,T_c)$.

(i)-only if part. Assume that method $M(I,EL_{T_a},T,TL_c,T_c)$ exists. Then, $T_c, T_a, EL_{T_a}, T$ and $TL_c$ are correct types.

(ii) Assume that method $M(I,EL_{T_a},T,TL_a,T_a)$ exists for $T_c \leq T_f \leq T_a$. Then, the method can be invoked where $E$ occurs in the program. This completes the proof.

Proof. Lemma 2. By the rule 7 (8) of $E[E]_\rho$ any invocation $E,p(\ldots)$ of a mc parameter of the source program is replaced in the translated program, with an invocation of the Apply method on the object bound to $p$: $(T_p)(p.Apply(E[p][E] \ldots))$. This object must be of type ApplyClass$\$nk and is inserted, in the translated program, by rule 5. Moreover, according to such a rule and by definition of the Interface ApplyClass$\$nk, this object is wrapping a method $m$. Let $T_o$ be the return type of $m$ and $ET_{a_1},\ldots,ET_{o_n}$
be the argument type list with which \( m \) can be invoked. Then, when \texttt{Apply} invocation applies to arguments \((c, e_1, \ldots, e_n)\), \( m \) is invoked on object \( c \) and \( \texttt{Apply} \) applies to arguments \((e_1, \ldots, e_n)\). Since \texttt{Apply} invocation is cast to \( T_p \), \( m \) must compute a value of a subtype of \( T_p \). Otherwise, a compile-time type error occurs in the object program. This completes the proof as far as covariance of the return type. Moreover, since \texttt{Apply} invocation is cast to \( T_p \), \( m \) must compute a value of a subtype of \( T_p \). Otherwise, a compile-time type error occurs in the object program. This completes the proof as far as covariance of the return type. □

Proof. Theorem 3. (i) Immediate from Lemma 2. (ii) Immediate from Lemma 1 and Lemma 2. □

Proof. Theorem 4. It is enough proving that \( E []_\emptyset \) is idempotent, i.e. \( E [E[P]_\emptyset]_\emptyset = E[P]_\emptyset \) for each program \( P \) (namely, compilation unit, in Java). None of \( E []_\rho \) rules inserts types \texttt{fun} or constructs \texttt{Abs} in the translated program. Instead, rules 3 and 4 remove types \texttt{fun} from each extended type \( ET \) and each of the types occurring inside \( ET \). Similarly, rules 5 and 6 act on the construct \texttt{Abs}. Hence, if \( E[P]_\emptyset \) contains an occurrence of either \texttt{fun} or \texttt{Abs} then such an occurrence was already in \( P \). But it cannot be since rules 3, 4 or 5, 6 would have removed the occurrence. □

Proof. Theorem 5. It can be given following the arguments of the proof of Theorem 4 reformulated on \( F []_{\emptyset} \) about the removal of \texttt{shared}, closure type and the closure construct. □

Proof. Theorem 6. By induction on the structure of \( P \): Proving that, rule by rule, \( F[E[U]_\rho]_{\tau} = E[F[U]_{\tau}]_\rho \) holds for each component \( U \) to which the rule applies (and any correct pair of environment \( \rho \) and \( \tau \)). □

Proof. Theorem 7. Immediate from Theorems on completeness and on orthogonality of the component translations. □

References


E. Gamma, R. Helm, R. Johnson, and J.M. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, 2005.


