# CS4215—Programming Language Implementation

Martin Henz

Thursday 15 March, 2012

# Chapter 11

# imPL: A Simple Imperative Language

## 11.1 Introduction

So far, we considered only languages, in which an identifier refers to a value. Once the value is computed, it does not change. Pass-by-need makes use of this property by avoiding repeated evaluation; the resulting value would be the same, anyway. A given identifier in a given environment always denotes the same value. A given expression is always evaluated to the same value in a given environment. Languages that have this property are called *referentially transparent*. This property makes it particularly easy to formally define language semantics and to reason about programs such as prove their correctness and termination. The languages ePL, simPL and rePL are referentially transparent.

However, it is often convenient to deviate from referential transparency. Many algorithms can be formulated more naturally in a language, in which identifiers refer to locations of a random-access memory. An operation called *assignment* allows to change the value stored in the memory location associated with an identifier. Languages with such a construct are called *imperative* languages.

In this chapter, we examine the semantics and implementation of imperative languages. We start with an extension of simPL by typical imperative constructs such as assignment and loops, resulting in imPL0. In order to investigate parameter passing techniques, we extend imPL0 by records and property assignment, resulting in imPL1. We will examine pass-by-value and pass-by-reference parameter passing, both for identifiers and for records, and pass-by-copy parameter passing for records.

## 11.2 imPL0: Imperative Programming

We extend the syntax of simPL by assignments, sequences and while-loops, resulting in imPL0. As in rePL, we abandon simPL's type declarations here.

**Assignment.** An assignment expression allows to change the value of an identifier to the result of evaluating an expression:

$$\frac{E}{x\ \texttt{:=}\ E}$$

**Sequence.** A sequence expression allows to evaluate first on expression to change the value of identifiers, and then evaluate another expression with the changes in effect:

$$\frac{E_1 \quad E_2}{E_1\ \texttt{;}\ E_2}$$

**While-loop.** A while loop allows to repeatedly evaluate an expression as long as a boolean expression evaluates to *true*:

$$\frac{E_1 \quad E_2}{\texttt{while}\ E_1\ \texttt{do}\ E_2\ \texttt{end}}$$

## 11.3 Examples

**Example 11.1** *The body of the following let expression repeatedly changes the value of the identifier* **x**. *The results of evaluation of intermediate values are ignored. They are only executed for their "side effect", namely the changing of the value of* **x**.

```
let x = 0 in
   x := 1;
   x := x + 2;
   x := x + 3;
   x
end
```

**Example 11.2** *We now have the possibility to write interesting programs without using recursion. An alternative function for computing the factorial function is:*

```
fun x ->
   let i = 1 in
      let f = 1 in
         while \ i > x do
            f := f * i;
            i := i + 1
         end;
         f
      end
   end
end
```

**Example 11.3** *The GCD program can now be written as a loop:*

```
let gcd = fun a b ->
             while \(a = b) do
                if a > b
                then a := a - b
                else b := b - a
                end
             end;
             a
          end
in
   (gcd 6 10)
end
```

## 11.4 Denotational Semantics of imPL0

Of course, the language that we use to describe our denotational semantics, namely mathematical notation, enjoys referential transparency. So the question arises how to describe assignment in our denotational semantics.

**Example 11.4** `let x = 0 in`

```
   x := 1;
   x := x + 2;
   x := x + 3;
   x
end
```

This program can only be understood in terms of its effect on the stored value of `x`. In order to make the intuitive notion of "stored value" explicit in our semantic framework, we introduce a new semantic domain, called the *store*. In that way we can say that the first assignment makes a binding of `x` to 1 in a given store $\Sigma$, resulting in a new store $\Sigma'$. This new store is then used by the second assignment, resulting in yet another store $\Sigma''$ and so on. By introducing the store, on which a program operates, we can describe the meaning of programs in a referentially transparent way.

We are changing the semantic domains such that now, identifiers always denote *locations*. These locations are used to access a store, which holds storable values.

| Domain name | Definition | Explanation |
|---|---|---|
| **EV** | $\mathbf{Int} + \mathbf{Bool} + \mathbf{Fun} + \{\mathbf{error}\}$ | expressible values |
| **SV** | $\mathbf{Int} + \mathbf{Bool} + \mathbf{Fun}$ | storable values |
| **DV** | $\mathbf{Loc}$ | denotable values |
| **Fun** | $\mathbf{DV} * \cdots * \mathbf{DV} * \mathbf{Store} \leadsto (\mathbf{EV},\ \mathbf{Store})$ | function values |
| **Store** | $\mathbf{Loc} \leadsto \mathbf{SV}$ | stores |
| **Env** | $\mathbf{Id} \leadsto \mathbf{DV}$ | environments |

Environments now do not refer directly to values, but instead to locations. These locations are then passed to stores in order to access the current value of the identifier with respect to the store.

For stores $\Sigma$ we introduce an operation $\Sigma[l \leftarrow v]$, which denotes a store that works like $\Sigma$, except that $\Sigma(l) = v$ (similar to the corresponding operation on environments). The symbol $\emptyset_{\mathbf{Store}}$ stands for the empty store, and similarly $\emptyset_{\mathbf{Env}}$ stands for the empty environment.

**Example 11.5** *Let us say we have a store with the value 1 at location $l$*

$$\Sigma = \emptyset_{\mathbf{Store}}[l \leftarrow 1]$$

*and an environment that carries the location $l$ at identifier $x$*

$$\Delta = \emptyset_{\mathbf{Env}}[x \leftarrow l]$$

*Then we can access the value of $x$ in the store as follows:*

$$\Sigma(\Delta(x)) = 1$$

By introducing the store on which a program operates, we can describe the meaning of programs in a referentially transparent way.

The semantic function $\cdot \rightarrowtail \cdot$ now needs to be defined using a semantic function $\cdot \mid \cdot \Vdash \cdot \rightarrowtail \cdot$ that gets a store as additional argument.

$$\cdot \rightarrowtail \cdot : \mathbf{imPL0} \to \mathbf{EV}$$

$$\frac{\emptyset_{\mathbf{Store}} \mid \emptyset_{\mathbf{Env}} \Vdash E \rightarrowtail (v, \Sigma)}{E \rightarrowtail v}$$

The semantic function $\cdot \mid \cdot \Vdash \cdot \rightarrowtail \cdot$ is defined as a four-argument relation (ternary partial function):

$$\cdot \mid \cdot \Vdash \cdot \rightarrowtail \cdot : \textbf{Store} * \textbf{Env} * \textbf{imPL0} \to \textbf{EV} * \textbf{Store}$$

The evaluation of `let` expressions works as follows (for clarity, we limit ourselves to the case with only one definition):

$$\frac{\Sigma'[l_1 \leftarrow v_1] \mid \Delta[x_1 \leftarrow l_1] \Vdash E \rightarrowtail (v, \Sigma'')}{\Sigma \mid \Delta \Vdash \texttt{ let } \{\cdot\}\ x_1 = E_1 \texttt{ in } \{\cdot\} E \texttt{ end} \rightarrowtail (v, \Sigma'')}$$

if $\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (v_1, \Sigma')$, and $l_1$ is a new location, which means $\Sigma'(l_1)$ is undefined

Correspondingly, the evaluation of identifiers needs to access the store.

$$\frac{}{\Sigma \mid \Delta \Vdash x \rightarrowtail (\Sigma(\Delta(x)), \Sigma)}$$

The semantic functions for assignment returns an updated store.

$$\frac{\Sigma \mid \Delta \Vdash E \rightarrowtail (v, \Sigma')}{\Sigma \mid \Delta \Vdash x \texttt{ := } E \ \rightarrowtail (v, \Sigma'[\Delta(x) \leftarrow v])}$$

**Example 11.6** $\emptyset_{\textit{Store}}[l \leftarrow 1] \mid \emptyset_{\textit{Env}}[\texttt{a} \leftarrow l] \Vdash \texttt{a := 2} \rightarrowtail (2, \emptyset_{\textit{Store}}[l \leftarrow 1][l \leftarrow 2])$
*The resulting store $\emptyset_{\textit{Store}}[l \leftarrow 1][l \leftarrow 2])$ is of course the same as $\emptyset_{\textit{Store}}[l \leftarrow 2]$. The original binding of $l$ to 1 is overwritten by the new value 2.*

Note that assignment is defined to always evaluate to the value to which its right hand side is evaluated. This choice is somewhat arbitrary, since typically, assignments are carried out for their "side effect", only, which means for the effect they have on the store.

In an imperative language, the question arises what parameter passing means when identifiers occur in argument position. Is the denotable value passed directly to the function, or do we create a new location, and place the value of the argument in the store at this new location? The first possibility is called "pass-by-reference" and the second "pass-by-value". We decide to use pass-by-value for identifiers, following most modern imperative languages. The following equations describe pass-by-value parameter passing. For simplicity, we only treat single-argument functions here.

$$\frac{}{\Sigma \mid \Delta \Vdash \texttt{fun } x \texttt{ -> } E \texttt{ end } \ \rightarrowtail (f, \Sigma)}$$

where $f(l, \Sigma') = (v', \Sigma'')$,
where $\Sigma' \mid \Delta[x \leftarrow l] \Vdash E \rightarrowtail (v', \Sigma'')$

Correspondingly, function application is defined as follows:

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (f, \Sigma') \qquad \Sigma' \mid \Delta \Vdash E_2 \rightarrowtail (v_2, \Sigma'')}{\Sigma \mid \Delta \Vdash \texttt{(}E_1\ E_2\texttt{)} \rightarrowtail f(l, \Sigma''[l \leftarrow v_2])} \quad \text{where } l \text{ is a new location in } \Sigma''.$$

Note the different treatment of environments and stores in function definition and application. Function values keep the environment of the function definition, but drop the store of the function definition. Applications use the environment of the function value, but the store resulting from the previous expression.

The meaning of sequences is given by the following rule.

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (v_1, \Sigma') \qquad \Sigma' \mid \Delta \Vdash E_2 \rightarrowtail (v_2, \Sigma'')}{\Sigma \mid \Delta \Vdash E_1\texttt{;}E_2 \rightarrowtail (v_2, \Sigma'')}$$

Note that the result $v_1$ of evaluating the first component $E_1$ of a sequence is ignored. The result of the sequence is the result $v_2$ of evaluating the second component $E_2$.

With while loops, we face the same problem as with `recfun` in simPL, namely circularity in rules. We give the following specification for the meaning of loops. A proper definition is beyond the scope of this course.

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (v_1, \Sigma')}{\Sigma \mid \Delta \Vdash \texttt{while}\ E_1\ \texttt{do}\ E_2\ \texttt{end} \rightarrowtail (\mathit{true}, \Sigma')} \quad \text{if } v_1 = \mathit{false}$$

The choice of *true* as the result of the expression is arbitrary; while loops are executed for their effect on the store, and not for obtaining their "value".

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (v_1, \Sigma')}{\Sigma \mid \Delta \Vdash \texttt{while}\ E_1\ \texttt{do}\ E_2\ \texttt{end} \rightarrowtail (v, \Sigma''')} \quad \text{if } v_1 = \mathit{true}, \text{ where } \Sigma' \mid \Delta \Vdash E_2 \rightarrowtail (v_2, \Sigma'') \text{ and } \Sigma'' \mid \Delta \Vdash \texttt{while}\ E_1\ \texttt{do}\ E_2\ \texttt{end} \rightarrowtail (v, \Sigma''')$$

This rule is circular, since the condition on the right hand side assumes a semantics of `while`; as with recursive functions in simPL, a thorough discussion of how to interpret such a circular rule is beyond the scope of this investigation. None the less, the rule will serve us as a specification for implementing while loops.

## 11.5 imPL1: Mutable Records

Aggregate values such as rePL's records provide the opportunity to further increase the expressive power of an imperative language. The idea is to add the possibility of changing the value of an individual property in a record.

To this aim, we add record property assignment as a variant of assignment to imPL.

$$\frac{E_1 \qquad E_2}{E_1.q \texttt{:=} E_2}$$

In order to capture the semantics of record property assignment, we need to extend the semantic domains as follows:

| Domain name | Definition |
|---|---|
| **EV** | **Int** + **Bool** + **Fun** + **Rec** + $\{\bot\}$ |
| **SV** | **Int** + **Bool** + **Fun** + **Rec** |
| **DV** | **Loc** |
| **Rec** | **Id** $\rightsquigarrow$ **Loc** |

Note that records are represented as partial functions mapping identifiers representing properties to locations. The semantic function $\cdot \mid \cdot \Vdash \cdot \rightarrowtail \cdot$ is extended accordingly as follows.

$$\frac{\Sigma^{(0)} \mid \Delta \Vdash E_1 \rightarrowtail (v_1, \Sigma^{(1)}) \quad \Sigma^{(n-1)} \mid \Delta \Vdash E_n \rightarrowtail (v_n, \Sigma^{(n)})}{\Sigma^{(0)} \mid \Delta \Vdash \texttt{[}q_1 \texttt{:} E_1 \texttt{,} \ldots \texttt{,} q_n \texttt{:} E_n \texttt{]} \rightarrowtail (f, \Sigma')}$$

where $l_1, \ldots, l_n$ are new locations in $\Sigma^{(n)}$, $\Sigma' = \Sigma^{(n)}[l_1 \leftarrow v_1] \quad \cdots [l_n \leftarrow v_n]$, and $f(q_i) = l_i$ for $1 \leq i \leq n$.

$$\frac{\Sigma \mid \Delta \Vdash E \rightarrowtail (f, \Sigma')}{\Sigma \mid \Delta \Vdash E.q \rightarrowtail (\Sigma'(f(q)), \Sigma')}$$

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (f, \Sigma') \qquad \Sigma' \mid \Delta \Vdash E_2 \rightarrowtail (v, \Sigma'')}{\Sigma \mid \Delta \Vdash E_1.q \texttt{:=} E_2 \rightarrowtail (v, \Sigma''[f(q) \leftarrow v])}$$

Similar to passing of identifiers, the question arises here how records are passed to functions. Let us see what happens if a record is passed as argument to a function. According to the definition of function application, we allocate a new location in the store for the record. However, the locations *in* the record still refer to the original locations. Thus, assignments to record properties remain visible after returning from the function (pass-by-reference), whereas assignments to the record itself are not visible after returning from the function (pass-by-value).

**Example 11.7** *We have a similar behavior in Java, when passing objects to functions:*

```
void f(MyClass myobject) {
   myobject.myfield = 1;
   myobject = new MyClass();
   myobject.myfield = 2;
}
```

*After returning from a call* `f(obj)`*, the identifier* `obj` *still refers to the same object (in the sense of object identity), whereas the field* `Myfield` *has changed to 1.*

**Example 11.8** *Records in imPL are passed similar to objects in Java. The expression*

```
let a = [Myfield: 0]
in
    (fun b ->
        b.Myfield := 1;
        b := [Myfield: 2];
        b.Myfield := 3
     end
     a);
    a.Myfield
end
```

*evaluates to the integer 1.*

## 11.6 Pass-by-Reference Parameter Passing

In Section 11.4, we decided to pass arguments "by-value" to functions. In an application `(f x)`, the body of function `f` cannot change the value in the store to which `x` refers. Instead, the function call creates a new location into which the value is copied.

In this section, we explore an alternative to this scheme, called pass-by-reference, similar to the treatment of records as arguments. When variables appear as function arguments, we pass their location directly to the function, instead of copying their value to a new location. Thus, we pass the reference to the function, not the value. The semantic rule for function definition remains unchanged. The semantic rules for function application are as follows.

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (f, \Sigma')}{\Sigma \mid \Delta \Vdash \texttt{(}E_1\ x\texttt{)} \rightarrowtail f(\Delta(x), \Sigma')}$$

In the case where the argument is not an indentifier, the old rule for application applies.

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (f, \Sigma') \qquad \Sigma' \mid \Delta \Vdash E_2 \rightarrowtail (v_2, \Sigma'')}{\Sigma \mid \Delta \Vdash \texttt{(}E_1\ E_2\texttt{)} \rightarrowtail f(l, \Sigma''[l \leftarrow v_2])} \quad \begin{array}{l}\text{if } E_2 \text{ is not an identifier,}\\ \text{where } l \text{ is a new location in } \Sigma''.\end{array}$$

## 11.7 Pass-by-Copy Parameter Passing

In Section 11.5 we saw that a strange mix of pass-by-value and pass-by-reference occurs when records are passed as arguments to functions. The argument record itself is passed by-value, in a sense that assignments to the record parameter in the function body does not affect the argument record. On the other hand, the components of the record are passed by-reference; they can be changed in the body and these changes are visible in the record after the function has returned.

In an alternative meaning of passing records, we could pass the record components by-value. Since this semantics involves copying the record components during function application, this parameter passing technique is called pass-by-copy.

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (f, \Sigma') \qquad \Sigma' \mid \Delta \Vdash E_2 \rightarrowtail (v_2, \Sigma'')}{\Sigma \mid \Delta \Vdash \texttt{(}E_1\ E_2\texttt{)} \rightarrowtail f(l, \Sigma''')}$$

if $v_2 \in \mathbf{Rec}$, where $l, l'_1, \ldots, l'_n$ are new locations in $\Sigma''$, $\{q_1, \ldots, q_n\} = dom(v_2)$, $\Sigma''' = \Sigma''[l \leftarrow \{(q_1, l'_1), \ldots, (q_n, l'_n)\}]\ [l'_1 \leftarrow \Sigma''(v_2(q_1))]\ \ldots [l'_n \leftarrow \Sigma''(v_2(q_n))]$

## 11.8 Imperative Programming and Exception Handling

The usual implementation of error handling in imperative programming is that as soon as an error occurs, the current store is returned along with the error value. For example, division by zero gets the following meaning.

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (v_1, \Sigma') \qquad \Sigma' \mid \Delta \Vdash E_2 \rightarrowtail (0, \Sigma'')}{\Delta \Vdash E_1/E_2 \rightarrowtail (e, \Sigma'')} \quad \text{if } v_1 \notin \textbf{Exc} \text{ and where } e = \texttt{[divisionByZero:true]}, \text{ and } e \in \textbf{Exc}$$

In the rules for the `try...catch...with...end` expression, the store returned by the try part is used in the `with` part.

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (v, \Sigma')}{\Sigma \mid \Delta \Vdash \texttt{try } E_1 \texttt{ catch } x \texttt{ with } E_2 \texttt{ end} \rightarrowtail (v, \Sigma')} \quad \text{if } v \notin \textbf{Exc}$$

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (v_1, \Sigma') \quad \Sigma'[l \leftarrow v_1] \mid \Delta[x \leftarrow l] \Vdash E_2 \rightarrowtail (v_2, \Sigma'')}{\Sigma \mid \Delta \Vdash \texttt{try } E_1 \texttt{ catch } x \texttt{ with } E_2 \texttt{ end} \rightarrowtail (v_2, \Sigma'')} \quad \text{if } v_1 \in \textbf{Exc},\ l \text{ new loc.}$$

Thus, the changes to the store made in the `try` part are visible in the `with` part although an exception has occurred. The reason for this design choice is efficiency of implementation.

Arguably semantically more sound and intuitive would be a semantics that uses the incoming store of the `try...catch...with...end` expression for evaluating the `with` part as shown in the following rule.

$$\frac{\Sigma \mid \Delta \Vdash E_1 \rightarrowtail (v_1, \Sigma') \quad \Sigma[l \leftarrow v_1] \mid \Delta[x \leftarrow l] \Vdash E_2 \rightarrowtail (v_2, \Sigma'')}{\Sigma \mid \Delta \Vdash \texttt{try } E_1 \texttt{ catch } x \texttt{ with } E_2 \texttt{ end} \rightarrowtail (v_2, \Sigma'')} \quad \text{if } v_1 \in \textbf{Exc},\ l \text{ new loc.}$$

This would mean that a copy of the store would have to be saved for every `try...catch...with...end` expression. In programming language practice, this is infeasible. However, in databases, such a semantics is sometimes desirable. The corresponding technique is called "roll-back" and allows for recovery from database inconsistencies.

## 11.9 A Virtual Machine for imPL

The semantics of imperative constructs is designed to allow for an efficient implementation. The store is threaded through the entire run of the program. We can prove for the semantics presented in this chapter that after a rule has constructed a new store, the old store will never be used again.[1] In an actual implementation, there is no need for constructing a new store in each rule. We only need one copy of the store, and the operations in the store can be destructive.

[1] Note that this property would be violated by a "roll-back" semantics for `try...catch...with...` expressions.

In Chapter 9, we saw already a technique for realistically representing the data structures used in the machine. In the presented framework a heap allowed us to explicitly manipulate the objects that are created at runtime. It turns out that we can reuse the heap to provide a realistic implementation of imPL. We can view the targets of edges in the heap as locations. Following this view, the assignment expression simply changes the target node in the environment. Chapter 9 already introduced an *update* operation in order to efficiently manipulate stacks and other data structures on the heap. To implement assignment in the heap, we shall reuse this *update* operation.

Let us first translate assignment expressions using a new instruction `ASSIGNS` (Assign Symbolic) as follows.

$$\frac{E \hookrightarrow s}{x \texttt{ := } E \hookrightarrow s.\texttt{ASSIGNS } x}$$

Thus the instruction `ASSIGNS` carries with it the identifier, whose location gets a new value in the store, and finds that new value on the operand stack. Thus its execution needs to update the heap such that the identifier $x$ refers to the new value.

$$\frac{s(pc) = \texttt{ASSIGNS } x}{(os, pc, e, rs, h) \rightrightarrows_s (os, pc+1, e, rs, update(e, x, v, h'))} \text{where } (v, h') = pop(os, h)$$

The compilation of sequences introduces the new instruction `POP`.

$$\frac{E_1 \hookrightarrow s_1 \qquad E_2 \hookrightarrow s_2}{E_1 \,\texttt{;}\, E_2 \hookrightarrow s_1.\texttt{POP}.s_2}$$

The instruction `POP` simply pops the top entry from the operand stack and ignores it. This corresponds to the denotational semantics of sequences, which ignores the value of the first component.

$$\frac{s(pc) = \texttt{POP}}{(os, pc, e, rs, h) \rightrightarrows_s (os, pc+1, e, rs, h')} \text{where } (v, h') = pop(os, h)$$

The compilation of `while` expressions reuses the `JOFR`, `POP`, `GOTOR` and `LDCB` instructions and thus does not require any new instructions.

$$\frac{E_1 \hookrightarrow s_1 \qquad E_2 \hookrightarrow s_2}{\begin{array}{c}\texttt{while}\ E_1\ \texttt{do}\ E_2 \\ \hookrightarrow \\ s_1.(\texttt{JOFR}\ |s_2|+3).s_2.\texttt{POP}. \\ (\texttt{GOTOR}\ -(|s_1|+2+|s_2|)).\texttt{LDCB}\ \mathit{true}\end{array}}$$

Note that the `GOTOR` instruction jumps to the beginning of the code for $E_1$, which means that the condition will be re-evaluated in each iteration through the loop. The result of the body of the `while` expression is ignored using `POP`, and after the execution of the loop, the boolean value *true* is pushed as required by the denotational semantics of `while`.