

Definitional interpreters for higher-order programming languages

Krzysztof Wróbel

Institute of Computer Science
University of Wrocław

strzkrzysiek@gmail.com

11 March, 2015

- In 1972, that is at the time of Reynolds' research, there was no or very few methods to express the semantics of a given language.
- Most of the languages were usually defined by interpreters written in a programming language based on lambda calculus, that was hopefully better understood.
- One can see a problem that by writing such an interpreter some of the features of the defining language could be implicitly incorporated by the defined language (e.g. strategy of evaluation).
- The idea is to make somehow the defined language independent of the nature of the defining language.

The roadmap

- A simple applicative language
- Description of the defined language
- First and simple meta-cyclic interpreter
- Introducing CPS and defunctionalization
- A try to get rid of higher-order functions
- Independence from the strategy of evaluation – continuations
- A glimpse at some imperative features

The defining language

The language our interpreters are written in.

The defined language

The language defined by those interpreters.

The defining language – variables and constants

Variables

Set of symbols that is evaluated to some value specified in the given environment – a mapping between variables and values.

Constants

We will not specify the set of constants precisely, but it should contain at least integers and Boolean true and false. Their evaluation gives the same value regardless of the environment.

The defining language – syntax

Lambda abstraction

$$\lambda(r_1, \dots, r_n).r_{body}$$

Application

$$r_f(r_1, \dots, r_n)$$

Simple conditional expression

$$\text{if } r_p \text{ then } r_c \text{ else } r_a$$

Multiple conditional expression

$(r_{p1} \rightarrow r_{c1}, \dots, r_{pn} \rightarrow r_{cn})$ is equivalent to
if r_{p1} **then** r_{c1} **else** ... **if** r_{pn} **then** r_{cn} **else error**

The defining language – syntax

Let expression

let $x_1 = r_1$ **and** ... **and** $x_n = r_n$ **in** r_b

Recursive let expression

letrec $x_1 = r_1$ **and** ... **and** $x_n = r_n$ **in** r_b

The defined language

- Functions will be limited to a single argument. Thus all applicative expressions will have a single operand, and all lambda expressions will have a single formal parameter.
- Only call by value will be used.
- Only simple conditional expressions will be used.
- Nonrecursive let expressions will be excluded.
- All recursive let expressions will contain a single declaration and their declaring expressions can be only in a form of a lambda expression.
- Values will be integers, boolean, and functions (actually closures).
- Basic operations will be *succ* (returns the successor of an integer n) and *equal* (tests integer equality).

Abstract syntax of the defined language

Since this is beyond the scope of this talk, we won't be bothering about lexing and parsing the program. Instead, we will consider a program to be already in a form of a *abstract syntax tree*.

The nodes in that tree will be represented as records together with some adequate accessors, constructors, and classifiers. Consider a set S_0 of all records of the same "type". We will write:

$$S_0 = [a_1 : S_1, \dots, a_n : S_n]$$

where fields of these records are elements of respective set S_i and a_i denotes an accessor to the i th field. Moreover, implicitly we declare here a constructor $mk\text{-}s_0$ of n arguments and a classifier $s_0?$ that test whether its argument belongs to S_0 .

Abstract syntax of the defined language

Right now we are ready to define the data structures that will be used by the interpreter.

$$\text{EXP} = \text{CONST} \cup \text{VAR} \cup \text{APPL} \cup \text{LAMBDA} \cup \text{COND} \cup \text{LETREC}$$
$$\text{APPL} = [\text{opr}: \text{EXP}, \text{opnd}: \text{EXP}]$$
$$\text{LAMBDA} = [\text{fp}: \text{VAR}, \text{body}: \text{EXP}]$$
$$\text{COND} = [\text{prem}: \text{EXP}, \text{conc}: \text{EXP}, \text{altr}: \text{EXP}]$$
$$\text{LETREC} = [\text{dvar}: \text{VAR}, \text{dexp}: \text{LAMBDA}, \text{body}: \text{EXP}]$$
$$\text{VAL} = \text{INTEGER} \cup \text{BOOLEAN} \cup \text{FUNVAL}$$
$$\text{FUNVAL} = \text{VAL} \rightarrow \text{VAL}$$
$$\text{ENV} = \text{VAR} \rightarrow \text{VAL}$$

Finally... a meta-circular interpreter

```
eval =  $\lambda(r, e).$   
  (const?(r)  $\rightarrow$  evcon(r),  
   var?(r)  $\rightarrow$  e(r),  
   appl?(r)  $\rightarrow$  (eval(opr(r), e))(eval(opnd(r), e)),  
   lambda?(r)  $\rightarrow$  evlambda(r, e),  
   cond?(r)  $\rightarrow$  if eval(prem(r), e)  
     then eval(conc(r), e) else eval(altr(r), e),  
   letrec?(r)  $\rightarrow$  letrec e' =  
      $\lambda x.$  if x = dvar(r) then evlambda(dexp(r), e') else e(x)  
     -----  
   in eval(body(r), e'))
```

Finally... a meta-circular interpreter

$$evlambda = \lambda(l, e). \underline{\lambda a. eval(body(l), ext(fp(l), a, e))}$$
$$ext = \lambda(z, a, e). \lambda x. \text{if } x = z \text{ then } a \text{ else } e(x)$$

$$interpret = \lambda r. eval(r, initenv)$$
$$initenv = \lambda x. (x = \text{"succ"} \rightarrow \underline{\lambda a. succ(a)},$$

$$x = \text{"equal"} \rightarrow \underline{\lambda a. \lambda b. equal(a, b)})$$

Before we proceed... CPS

On the next slides we will use two techniques: converting to CPS and defunctionalization. It would be good to begin with some simple examples. We will start with converting a simple factorial function to CPS.

Direct style

$$fact = \lambda n. \text{ if } n = 0 \text{ then } 1 \text{ else } n * fact(n - 1)$$

Continuation-passing style

$$\begin{aligned} fact\text{-}c &= \lambda(n, c). \text{ if } n = 0 \text{ then } c(1) \text{ else } fact\text{-}c(n - 1, \lambda m. c(n * m)) \\ fact &= \lambda n. fact\text{-}c(n, \lambda x. x) \end{aligned}$$

Before we proceed... CPS

Another example of converting a program to CPS. We will write a function that multiplies all elements in a given list. In order to do that we have to extend our language with functions *empty?(l)*, *head(l)* and *tail(l)*.

Direct style

```
mult =  $\lambda l.$  if empty?(l) then 1 else head(l) * mult(tail(l))
```

Continuation-passing style

```
mult-c =  $\lambda(l, c).$  if empty?(l) then c(1)  
           else if equal(head(l), 0) then 0  
           else mult-c(tail(l),  $\lambda m.c(head(l) * m)$ )  
mult =  $\lambda n.mult-c(n, \lambda x.x)$ 
```

Before we proceed... Defunctionalization

The aim of defunctionalization is to get rid of use of higher-order features of our language. This means that we don't want any function either to be an argument of another function or to be returned by a function. Recall the example with factorial function.

$$\begin{aligned} fact-c &= \lambda(n, c). \text{ if } n = 0 \text{ then } c(1) \text{ else } fact-c(n - 1, \underline{\lambda m. c(n * m)}) \\ fact &= \lambda n. fact-c(n, \underline{\lambda x. x}) \end{aligned}$$

Two underlined lambdas are here passed as an argument of a function. Obviously, we initially we wanted them to be functions since they are continuations. But maybe there is a way to represent them. Indeed, there is.

Closure

Evaluation of a lambda expression which binds all occurrences of free variables to their values in a given environment.

We will represent that two lambda expressions as records which contain values of their global variables at the time of definition. Let's make use of already mentioned *record equations*.

MULT = [*arg*: INTEGER, *next*: CONT] will represent $\lambda m.n * c(m)$

INIT = [] will represent $\lambda x.x$

CONT = MULT \cup INIT

Before we proceed... Defunctionalization

Since right now our continuations are records, we cannot simply apply them to an integer. We will make the following transformation:

$$c(n) \rightarrow cont(n, c)$$

Now we can define the final transformed version of the factorial function.

```
fact-c =  $\lambda(n, c).$  if  $n = 0$  then  $cont(1, c)$   
           else  $fact-c(n - 1, mk-mult(n, c))$   
fact =  $\lambda n. fact-c(n, mk-init())$ 
```

```
cont =  $\lambda(a, c).$  ( $init? \rightarrow a,$   
                 $mult? \rightarrow cont(arg(c) * a, next(c))$ )
```

Defunctionalizing the meta-cyclic interpreter

Recall the meta-cyclic interpreter. Some of the structures used in it were represented with functions. These were:

- functional values / closures (FUNVAL),
- environment (ENV).

We will try to defunctionalize them to records.

- $(eval(opr(r), e))(eval(opnd(r), e)) \rightarrow apply(eval(opr(r), r), eval(opnd(r), e))$
- $e(r) \rightarrow get(e, r)$

Let's start with the set FUNVAL.

Defunctionalizing FUNVAL

In the code of the interpreter elements of the set FUNVAL were underlined with solid line. There were four of them and for each we define a separate record equation.

$\lambda a. eval(\underline{body(l)}, \underline{ext(fp(l)}, a, e))$	$CLOSR = [lam : LAMBDA, en : ENV]$
$\lambda a. succ(a)$	$SC = []$
$\lambda a. \lambda b. equal(a, b)$	$EQ1 = []$
$\lambda b. equal(a, b)$	$EQ2 = [arg1 : VAL]$

$$FUNVAL = CLOSR \cup SC \cup EQ1 \cup EQ2$$

Defunctionalizing FUNVAL

$evlambda = \lambda(l, e).mk-closr(l, e)$

$initenv = \lambda x. (x = "succ" \rightarrow mk-sc(),$

 $x = "equal" \rightarrow mk-eq1()))$

$apply = \lambda(f, a).$
 $(closr?(f) \rightarrow \text{let } l = lam(f) \text{ and } e = en(f)$
 $\text{in } eval(body(l), ext(fp(l), a, e)),$
 $sc?(f) \rightarrow succ(a),$
 $eq1?(f) \rightarrow mk-eq2(a),$
 $eq2?(f) \rightarrow \text{let } b = a \text{ and } a = arg1(f) \text{ in } equal(a, b))$

Defunctionalizing ENV

Similarly, in the interpreter elements of set ENV were underlined with dashed line. There were three of them and again, we have three record equations.

An initial environment:

$$\text{INIT} = []$$

A simple extension of an environment:

$$\text{SIMP} = [bvar : \text{VAR}, bval : \text{VAL}, old : \text{ENV}]$$

A letrec extension of an environment:

$$\text{REC} = [letx : \text{LETREC}, old : \text{ENV}, \text{new} : \text{ENV}]$$

$$\text{ENV} = \text{INIT} \cup \text{SIMP} \cup \text{REC}$$

Replacement of the three environment-producing lambda expression gives:

$$\text{letrec?}(r) \rightarrow \mathbf{letrec} \ e' = \text{mk-rec}(r, e) \dots$$

$$\text{ext} = \lambda(z, a, e). \text{mk-simpl}(z, a, e)$$

$$\text{initenv} = \text{mk-init}()$$

Defunctionalizing ENV

... and the environment producing function is:

$get = \lambda(e, x).$

$(init?(e) \rightarrow (x = "succ" \rightarrow mk-sc(), x = "equal" \rightarrow mk-eq1()),$

$simp?(e) \rightarrow \text{let } z = bvar(e) \text{ and } a = bval(e) \text{ and } e = old(e)$

$\text{in if } x = z \text{ then } a \text{ else } get(e, x),$

$rec?(e) \rightarrow \text{let } r = letx(e) \text{ and } e = old(e) \text{ and } e' = e$

$\text{in if } x = dvar(r) \text{ then } evlambda(dexp(r), e') \text{ else } get(e, x))$

Finally... the second interpreter

$interpret = \lambda r. eval(r, mk-init())$

$eval = \lambda(r, e).$

$(const?(r) \rightarrow evcon(r),$

$var?(r) \rightarrow get(e, r),$

$appl?(r) \rightarrow apply(eval(opr(r), r), eval(opnd(r), e)),$

$lambda?(r) \rightarrow mk-closr(r, e),$

$cond?(r) \rightarrow \text{if } eval(prem(r), e)$

$\text{then } eval(conc(r), e) \text{ else } eval(altr(r), e),$

$letrec?(r) \rightarrow eval(body(r), mk-rec(r, e)))$

Finally... the second interpreter

```
apply =  $\lambda(f, a).$   
  (closr?(f)  $\rightarrow$   
    eval(body(lam(f)), ext(fp(lam(f)), a, en(f))),  
  sc?(f)  $\rightarrow$  succ(a),  
  eq1?(f)  $\rightarrow$  mk-eq2(a),  
  eq2?(f)  $\rightarrow$  equal(arg1(f), a))  
  
get =  $\lambda(e, x).$   
  (init?(e)  $\rightarrow$  (x = "succ"  $\rightarrow$  mk-sc(), x = "equal"  $\rightarrow$  mk-eq1()),  
  simp?(e)  $\rightarrow$  if x = bvar(e) then bval(e) else get(old(e), x),  
  rec?(e)  $\rightarrow$  if x = dvar(letx(e))  
    then mk-closr(dexp(letx(e)), e) else get(old(e), x))
```

Non-terminating expressions and evaluation-strategy dependence

Consider an example where exp is non-termination and f terminates, and doesn't need the value of expression exp . Then the answer for a question whether the following expression terminates depends on the strategy of evaluation of the defining language.

$$apply(eval(opr(r), e), eval(opnd(r), e))$$

However, we wanted our defined language to incorporate call-by-value strategy and because of this, in our interpreter that expression should never terminate.

Continuations

To deal with this problem we will introduce continuations

$\text{CONT} = \text{VAL} \rightarrow \text{VAL}$,

and change functions *interpret*, *eval* and *apply* to have the following form:

$\text{interpret} = \lambda r. \text{eval}(r, \text{mk-init}(), \lambda a. a)$

$\text{eval} = \lambda(r, e, c). \dots$

$\text{apply} = \lambda(f, a, c). \dots$

We will think our further actions to perform will be embedded into those continuations. This will allow us to have the control of order of execution.

Continuations

For all of the trivial functions (i.e. definitely terminating) we will simply pass the result of their application to the current continuation.

$$\begin{aligned} eval &= \lambda(r, e, c). \\ &\quad (const?(r) \rightarrow c(evcon(r)), \\ &\quad var?(r) \rightarrow c(get(e, r)), \\ &\quad \vdots \\ &\quad lambda?(r) \rightarrow c(mk-closr(r, e)), \dots) \\ apply &= \lambda(f, a, c).(\dots, \\ &\quad sc?(f) \rightarrow c(succ(a)), \\ &\quad eq1?(f) \rightarrow c(mk-eq2(a)), \\ &\quad eq2?(f) \rightarrow c(equal(arg1(f), a))) \end{aligned}$$

In the following instructions we would like to pass the current continuation (i.e. actions we have to perform later) as an argument of *eval*.

$$\text{letrec?}(r) \rightarrow (\text{eval}(\text{body}(r), \text{mk-rec}(r, e), c))$$

$$\vdots$$

$$\begin{aligned} &(\text{closr?}(f) \rightarrow \\ &\quad \text{eval}(\text{body}(\text{lam}(f)), \text{mk-simp}(\text{fp}(\text{lam}(f)), a, \text{en}(f)), c) \end{aligned}$$

We are left with two statements where we would like to force the order and strategy of evaluation (left-to-right and call-by-value).

$$appl?(r) \rightarrow eval(opr(r), e, \lambda f. eval(opnd(r), e, \lambda a. apply(f, a, c)))$$

$$cond?(r) \rightarrow eval(prem(r), e, \\ \lambda b. \textbf{if } b \textbf{ then } eval(conc(r), e, c) \textbf{ else } eval(altr(r), e, c))$$

The almost ready third interpreter

$interpret = \lambda r. eval(r, mk-init(), \underline{\lambda a. a})$

$eval = \lambda(r, e, c).$

$(const?(r) \rightarrow c(evcon(r))),$

$var?(r) \rightarrow c(get(e, r)),$

$appl?(r) \rightarrow eval(opr(r), e, \underline{\lambda f. eval(opnd(r), e, \underline{\lambda a. apply(f, a, c))})},$

$lambda?(r) \rightarrow c(mk-closr(r, e)),$

$cond?(r) \rightarrow eval(prem(r), e,$

$\underline{\lambda b. \text{ if } b \text{ then } eval(conc(r), e, c) \text{ else } eval(altr(r), e, c)})$

$letrec?(r) \rightarrow eval(body(r), mk-rec(r, e), c))$

The almost ready third interpreter

apply = $\lambda(f, a, c).$

$(\text{closr?}(f) \rightarrow$

$\text{eval}(\text{body}(\text{lam}(f)), \text{mk-simp}(\text{fp}(\text{lam}(f)), a, \text{en}(f)), c),$

$\text{sc?}(f) \rightarrow c(\text{succ}(a)),$

$\text{eq1?}(f) \rightarrow c(\text{mk-eq2}(a)),$

$\text{eq2?}(f) \rightarrow c(\text{equal}(\text{arg1}(f), a)))$

get = $\lambda(e, x).$

$(\text{init?}(e) \rightarrow (x = \text{"succ"} \rightarrow \text{mk-sc}(), x = \text{"equal"} \rightarrow \text{mk-eq1}()),$

$\text{simp?}(e) \rightarrow \text{if } x = \text{bvar}(e) \text{ then } \text{bval}(e) \text{ else } \text{get}(\text{old}(e), x),$

$\text{rec?}(e) \rightarrow \text{if } x = \text{dvar}(\text{letx}(e))$

$\text{then } \text{mk-closr}(\text{dexp}(\text{letx}(e)), e) \text{ else } \text{get}(\text{old}(e), x))$

Why almost?

By converting our interpreter to CPS, we have once again introduced higher-order functions. Since the conversion to first-order language is pretty mechanical, we will just write down the new record equations and then the fully ready, defunctionalized, strategy-of-evaluation independent interpreter.

The initial continuation – identity

$FIN = []$

Evaluate-operand continuation

$EVOPN = [ap : APPL, en : ENV, next : CONT]$

Apply-function continuation

$APFUN = [fun : VAL, next : CONT]$

Branch continuation

$BRANCH = [cn : COND, en : ENV, next : CONT]$

$CONT = FIN \cup EVOPN \cup APFUN \cup BRANCH$

Finally... the third interpreter

$interpret = \lambda r. eval(r, mk-init(), mk-fin())$

$eval = \lambda(r, e, c).$

$(const?(r) \rightarrow cont(c, evcon(r))),$

$var?(r) \rightarrow cont(c, get(e, r)),$

$appl?(r) \rightarrow eval(opr(r), e, mk-evopn(r, e, c)),$

$lambda?(r) \rightarrow cont(c, mk-closr(r, e)),$

$cond?(r) \rightarrow eval(prem(r), e, mk-branch(r, e, c)),$

$letrec?(r) \rightarrow eval(body(r), mk-rec(r, e), c))$

Finally... the third interpreter

apply = $\lambda(f, a, c).$

$(\text{closr?}(f) \rightarrow$

$\text{eval}(\text{body}(\text{lam}(f)), \text{mk-simp}(\text{fp}(\text{lam}(f)), a, \text{en}(f)), c),$

$\text{sc?}(f) \rightarrow \text{cont}(c, \text{succ}(a)),$

$\text{eq1?}(f) \rightarrow \text{cont}(c, \text{mk-eq2}(a)),$

$\text{eq2?}(f) \rightarrow \text{cont}(c, \text{equal}(\text{arg1}(f), a)))$

get = $\lambda(e, x).$

$(\text{init?}(e) \rightarrow (x = \text{"succ"} \rightarrow \text{mk-sc}(), x = \text{"equal"} \rightarrow \text{mk-eq1}()),$

$\text{simp?}(e) \rightarrow \text{if } x = \text{bvar}(e) \text{ then } \text{bval}(e) \text{ else } \text{get}(\text{old}(e), x),$

$\text{rec?}(e) \rightarrow \text{if } x = \text{dvar}(\text{letx}(e))$

$\text{then } \text{mk-closr}(\text{dexp}(\text{letx}(e)), e) \text{ else } \text{get}(\text{old}(e), x))$

Finally... the third interpreter

$cont = \lambda(c, a).$
 $(fin?(c) \rightarrow a,$
 $evopn?(c) \rightarrow \text{let } f = a \text{ and } r = ap(c) \text{ and}$
 $e = en(c) \text{ and } c = next(c)$
 $\text{in } eval(opnd(r), e, mk-apfun(f, c)),$
 $apfun?(c) \rightarrow \text{let } f = fun(c) \text{ and } c = next(c) \text{ in } apply(f, a, c),$
 $branch?(c) \rightarrow \text{let } b = a \text{ and } r = cn(c) \text{ and}$
 $e = en(c) \text{ and } c = next(c)$
 $\text{in if } b \text{ then } eval(conc(r), e, c) \text{ else } eval(altr(r), e, c))$

Escape expressions

We will introduce now an imperative control mechanism.

If (in the defined language) x is a variable and r is an expression, then

escape x in r

is an *escape expression*. The evaluation of it in an environment e proceeds as follows:

- The body r is evaluated in the environment that is the extension of e that binds x to a function called the *escape expression*.
- If the escape function is never applied during the evaluation of r , then the value of r becomes the value of the escape expression.
- If the escape function is applied to an argument a , then the evaluation of the body r is aborted, and a immediately becomes the value of the escape function.

Escape expressions

In order to extend our interpreters to handle escape expressions, we begin by extending the abstract syntax appropriately:

$$\text{EXP} = \dots \cup \text{ESCP}$$
$$\text{ESCP} = [\text{escv} : \text{VAR}, \text{body} : \text{EXP}]$$

Since the escape variable is bound to a function, we must add to the set FUNVAL a new kind of record that represents escape functions:

$$\text{FUNVAL} = \dots \cup \text{ESCF}$$
$$\text{ESCF} = [\text{cn} : \text{CONT}]$$

Escape expressions

These records are created in the new branch of *eval*:

$$\begin{aligned} eval = \lambda(r, e, c).(\dots, \\ escp?(r) \rightarrow eval(body(r), mk-simp(escv(r), mk-escf(c), e), c)) \end{aligned}$$

and are interpreted by a new branch of *apply*:

$$\begin{aligned} apply = \lambda(f, a, c).(\dots, \\ escf?(f) \rightarrow cont(cn(f), a)) \end{aligned}$$



John C. Reynolds (1998)

Definitional interpreters for higher-order programming languages.

Higher-Order and Symbolic Computation 11(4), 363 – 397.

The End