

# Computer Science 131, Spring 2001

## Assignment 5: Interpreters

Out: Wednesday, February 21

**Due: Wednesday, February 28, 11:00am**

To complete this assignment, retrieve the files `assign5.sml` and `assign5-tests.sml` from the Assignments web page. You must then run SML/NJ on turing using the command `/cs/cs131/bin/sml-cs131-a5`. This is a version of the compiler which has some important code predefined.

Your task is to complete the functions in the `assign5.sml` file and turn this in. As always, your file should contain no syntax or type errors. The submission process is the same as before (`cs131submit assign5.sml`).

The language NQSML (Not Quite SML) is very similar to the Mini-ML language you have already seen, but with the following changes:

- The arithmetic and comparison operators are slightly different.
- Function values are written with explicit type annotations; in the function value `fix f(x:t1):t2 is e`, the type  $t_1$  is the type of function's argument and  $t_2$  is the type that the function returns.
- Pairs and projections have been added, as in the previous assignment.
- To make the language a bit more interesting, lists have been added as well. As in SML, for any type  $t$  we have the type `t list` which is the type of lists whose elements all have type  $t$ . The expression `e1 :: e2` represents the list whose head (first element) is  $e_1$  and whose tail (the list of remaining elements) is  $e_2$ . Since this language contains no polymorphism, the empty list `nil[t]` is marked with the type of the elements that this list would contain if it weren't empty. For example, `nil[Int]` is the empty list of integers, while `nil[Bool->Bool]` is the empty list of functions which map booleans to booleans. (The type annotation is necessary in order to ensure that every expression in the language has a unique type, which makes typechecking much easier.) Finally we have a case-analysis for list values, `case e1 of nil => e2 | x::y => e3`. Note that unlike SML this case statement must have *exactly* two cases, and that the non-nil case *must* be a pattern of the form `x :: y` where  $x$  and  $y$  are variables. (More complicated patterns can be encoded using this construct, however.) The variables  $x$  and  $y$  are then bound variables with  $e_3$  as their scope.

---

```

v ::= n
   | tt
   | ff
   | fix f(x:t1):t2 is e2
   | ⟨v1, v2⟩
   | nil [t]
   | v1 :: v2

e ::= v
   | x
   | e1+e2
   | e1<e2
   | if e1 then e2 else e3
   | let x be e1 in e2
   | e1 e2
   | ⟨e1, e2⟩
   | fst e
   | snd e
   | e1 :: e2
   | case e1 of nil => e2 | x::y => e3

t ::= Int
   | Bool
   | t1->t2
   | t1 × t2
   | t list

```

Figure 1: Abstract Syntax for This Assignment

---

All of your programs will require using the following representation for NQSML programs. (These types have been pre-defined for you in the `sml-cs131-a3` binary. You will *not* need to type this in as part of your `assign5.sml` file.)

```

type varname = string
datatype ty   = Int_t
              | Bool_t
              | Arrow_t of ty * ty
              | Times_t of ty * ty
              | List_t of ty
datatype aop  = Plus | Minus | Times
datatype cop  = Less | Equal
datatype exp  = Num of int
              | Bool of bool
              | Var of varname
              | Arith of exp * aop * exp
              | Compare of exp * cop * exp
              | If of exp * exp * exp
              | Let of varname * exp * exp
              | Fix of varname * varname * ty * ty * exp
              | Apply of exp * exp
              | Pair of exp * exp
              | Fst of exp
              | Snd of exp
              | Nil of ty
              | Cons of exp * exp
              | Listcase of exp * exp * (varname * varname * exp)

```

This datatype should be mostly self-explanatory. You can also look at some of the pre-defined expressions in the `assign5-tests.sml` file to see their abstract syntax and the SML representations of this abstract syntax. Alternatively, Figure 2 gives a formal definition of the translation between abstract syntax into the SML representation. The notation  $[e]$  will be used to denote the SML representation corresponding to the abstract syntax for  $e$ , while  $[t]$  will be used to denote the SML representation of the type  $t$ .

The `sml-cs131-a3` binary also includes the type `'a env` (environments associating values of type `'a` with variable names) and the following items:

```

empty   : 'a env
extend  : 'a env * varname * 'a -> 'a env
lookup  : 'a env * varname -> 'a

```

These act exactly like the implementations shown in a previous class, except that instead of `lookup` returning either `NONE` or `SOME` it raises the `Error` exception when the variable is not found in the environment.

Finally, the `sml-cs131-a3` binary supplies the functions

---

<code>[Int]</code>	<code>= Int_t</code>
<code>[Bool]</code>	<code>= Bool_t</code>
<code>[t<sub>1</sub>-&gt;t<sub>2</sub>]</code>	<code>= Arrow_t([t<sub>1</sub>],[t<sub>2</sub>])</code>
<code>[t<sub>1</sub> × t<sub>2</sub>]</code>	<code>= Times_t([t<sub>1</sub>],[t<sub>2</sub>])</code>
<code>[t list]</code>	<code>= List_t([t])</code>
<code>[m]</code>	<code>= Num(m)</code>
<code>[tt]</code>	<code>= Bool(true)</code>
<code>[ff]</code>	<code>= Bool(false)</code>
<code>[x]</code>	<code>= Var("x")</code>
<code>[e<sub>1</sub>+e<sub>2</sub>]</code>	<code>= Arith([e<sub>1</sub>],Plus,[e<sub>2</sub>])</code>
<code>[e<sub>1</sub>-e<sub>2</sub>]</code>	<code>= Arith([e<sub>1</sub>],Minus,[e<sub>2</sub>])</code>
<code>[e<sub>1</sub>*e<sub>2</sub>]</code>	<code>= Arith([e<sub>1</sub>],Times,[e<sub>2</sub>])</code>
<code>[e<sub>1</sub>&lt;e<sub>2</sub>]</code>	<code>= Compare([e<sub>1</sub>],Less,[e<sub>2</sub>])</code>
<code>[e<sub>1</sub>=e<sub>2</sub>]</code>	<code>= Compare([e<sub>1</sub>],Equal,[e<sub>2</sub>])</code>
<code>[if e<sub>1</sub> then e<sub>2</sub> else e<sub>3</sub>]</code>	<code>= If([e<sub>1</sub>],[e<sub>2</sub>],[e<sub>3</sub>])</code>
<code>[let x be e<sub>1</sub> in e<sub>2</sub>]</code>	<code>= Let("x",[e<sub>1</sub>],[e<sub>2</sub>])</code>
<code>[fix f(x:t<sub>1</sub>):t<sub>2</sub> is e<sub>2</sub>]</code>	<code>= Fix("f","x",[t<sub>1</sub>],[t<sub>2</sub>],[e<sub>2</sub>])</code>
<code>[e<sub>1</sub> e<sub>2</sub>]</code>	<code>= Apply([e<sub>1</sub>],[e<sub>2</sub>])</code>
<code>[⟨e<sub>1</sub>,e<sub>2</sub>⟩]</code>	<code>= Pair([e<sub>1</sub>],[e<sub>2</sub>])</code>
<code>[fst e]</code>	<code>= Fst([e])</code>
<code>[snd e]</code>	<code>= Snd([e])</code>
<code>[nil [t]]</code>	<code>= Nil([t])</code>
<code>[e<sub>1</sub> :: e<sub>2</sub>]</code>	<code>= Cons([e<sub>1</sub>],[e<sub>2</sub>])</code>
<code>[case e<sub>1</sub> of nil =&gt; e<sub>2</sub>   x::y =&gt; e<sub>3</sub>]</code>	<code>= Listcase([e<sub>1</sub>],[e<sub>2</sub>],("x","y",[e<sub>3</sub>]))</code>

Figure 2: SML Representations of Abstract Syntax

---

```

ppty  : ty -> unit
ppexp : exp -> unit
pptyenv : ty env -> unit
ppexpenv : exp env -> unit

```

which print types, expressions, type environments, and value environments respectively. You may find these useful for debugging. (For example, while debugging my solution to #4 I temporarily changed the code

```

fun evalDyn (env,e) =
  (case e of
    Num _ => e

```

to

```

fun evalDyn (env,e) =
  (ppexp e;
   case e of
     Num _ => e

```

so that it would print a trace of every call to `evalDyn`.)

## 1 Typechecking

The typing rules for the extended language are as in the NQSML handout, plus the typing rules for pairs and projections given in Assignment 2, plus new rules for the list-related expressions. The new rules are as follows:

$$\frac{}{\Gamma \vdash \text{nil}[t] : t \text{ list}} \quad (1)$$

$$\frac{\Gamma \vdash e_1 : t \quad \Gamma \vdash e_2 : t \text{ list}}{\Gamma \vdash (e_1 :: e_2) : t \text{ list}} \quad (2)$$

$$\frac{\Gamma \vdash e_1 : u \text{ list} \quad \Gamma \vdash e_2 : t \quad \Gamma, x:u, y:(u \text{ list}) \vdash e_3 : t}{\Gamma \vdash (\text{case } e_1 \text{ of nil } \Rightarrow e_2 \mid x::y \Rightarrow e_3) : t} \quad (3)$$

Define the function

```
typeof : (ty env) * exp -> ty
```

which given an environment associating variable names with types and an expression, returns the type of the expression if it has one and otherwise raises the `Error` exception. (This exception has been predefined for you.) Remember that for expressions like  $e_1+e_2$  even though you immediately know if this has a type it must be `Int`, you still have to check that  $e_1$  and  $e_2$  are well-typed with type `Int`.

Several of the cases for this function have been given to you in the file `assign5.sml`. Fix the function so that it returns the correct answer for the remaining cases instead of raising the `Unimplemented` exception.

---

$n[x \mapsto v]$	$= n$
$\mathbf{tt}[x \mapsto v]$	$= \mathbf{tt}$
$\mathbf{ff}[x \mapsto v]$	$= \mathbf{ff}$
$(e_1 + e_2)[x \mapsto v]$	$= (e_1[x \mapsto v]) + (e_2[x \mapsto v])$
$(e_1 - e_2)[x \mapsto v]$	$= (e_1[x \mapsto v]) - (e_2[x \mapsto v])$
$(e_1 * e_2)[x \mapsto v]$	$= (e_1[x \mapsto v]) * (e_2[x \mapsto v])$
$(e_1 < e_2)[x \mapsto v]$	$= (e_1[x \mapsto v]) < (e_2[x \mapsto v])$
$(e_1 = e_2)[x \mapsto v]$	$= (e_1[x \mapsto v]) = (e_2[x \mapsto v])$
$(\mathbf{if} \ e_1 \ \mathbf{then} \ e_2 \ \mathbf{else} \ e_3)[x \mapsto v]$	$= \mathbf{if} \ (e_1[x \mapsto v]) \ \mathbf{then} \ (e_2[x \mapsto v]) \ \mathbf{else} \ (e_3[x \mapsto v])$
$x[x \mapsto v]$	$= v$
$y[x \mapsto v]$	$= y \quad \text{if } x \neq y$
$(\mathbf{let} \ y \ \mathbf{be} \ e_1 \ \mathbf{in} \ e_2)[x \mapsto v]$	$= \mathbf{let} \ y \ \mathbf{be} \ e_1[x \mapsto v] \ \mathbf{in} \ e_2[x \mapsto v] \quad \text{if } x \neq y$
	$= \mathbf{let} \ y \ \mathbf{be} \ e_1 \ \mathbf{in} \ e_2 \quad \text{otherwise}$
$(\mathbf{fix} \ f(y:t_1):t_2 \ \mathbf{is} \ e_2)[x \mapsto v]$	$= \mathbf{fix} \ f(y:t_1):t_2 \ \mathbf{is} \ (e_2[x \mapsto v]) \quad \text{if } x \neq f \ \text{and } x \neq y$
	$= \mathbf{fix} \ f(y:t_1):t_2 \ \mathbf{is} \ e_2 \quad \text{otherwise}$
$(e_1 \ e_2)[x \mapsto v]$	$= (e_1[x \mapsto v]) \ (e_2[x \mapsto v])$
$\langle e_1, e_2 \rangle[x \mapsto v]$	$= \langle e_1[x \mapsto v], e_2[x \mapsto v] \rangle$
$(\mathbf{fst} \ e)[x \mapsto v]$	$= \mathbf{fst} \ (e[x \mapsto v])$
$(\mathbf{snd} \ e)[x \mapsto v]$	$= \mathbf{snd} \ (e[x \mapsto v])$
$\mathbf{nil}[t][x \mapsto v]$	$= \mathbf{nil}[t]$
$(e_1 :: e_2)[x \mapsto v]$	$= (e_1[x \mapsto v]) :: (e_2[x \mapsto v])$
$(\mathbf{case} \ e_1 \ \mathbf{of} \ \mathbf{nil} \ \Rightarrow \ e_2 \   \ y::z \ \Rightarrow \ e_3)[x \mapsto v]$	$= \mathbf{case} \ (e_1[x \mapsto v]) \ \mathbf{of} \ \mathbf{nil} \ \Rightarrow \ (e_2[x \mapsto v]) \   \ y::z \ \Rightarrow \ (e_3[x \mapsto v]) \quad \text{if } x \neq y \ \text{and } x \neq z$
	$= \mathbf{case} \ (e_1[x \mapsto v]) \ \mathbf{of} \ \mathbf{nil} \ \Rightarrow \ (e_2[x \mapsto v]) \   \ y::z \ \Rightarrow \ e_3 \quad \text{otherwise}$

Figure 3: Substitution of Closed Values

---

## 2 Substitution

Define the function

```
val substClosed : exp * varname * exp -> exp
```

by recursion on the first expression such that `substClosed([e], "x", [v])` returns the abstract syntax for the term  $e[x \mapsto v]$ . You may assume that the argument  $v$  is closed (has no free variables). This makes the substitution code simpler because there is no possibility of variable capture. Figure 3 contains the definition of substitution in the extended syntax, simplified by the assumption that the value  $v$  contains no free variables.

The file `assign5.sml` contains the code for several cases of this function. Fix this code so that it returns the correct answer for the remaining cases instead of raising the `Unimplemented` exception.

## 3 A Simple Evaluator

Here is a big-step (or “natural”) semantics for the NQSMML language.

$$\frac{}{v \Downarrow v} \quad (4)$$

$$\frac{e_1 \Downarrow m_1 \quad e_2 \Downarrow m_2}{e_1 + e_2 \Downarrow m_1 + m_2} \quad (5)$$

$$\frac{e_1 \Downarrow m_1 \quad e_2 \Downarrow m_2}{e_1 - e_2 \Downarrow m_1 - m_2} \quad (6)$$

$$\frac{e_1 \Downarrow m_1 \quad e_2 \Downarrow m_2}{e_1 * e_2 \Downarrow m_1 m_2} \quad (7)$$

$$\frac{e_1 \Downarrow m_1 \quad e_2 \Downarrow m_2}{e_1 < e_2 \Downarrow m_1 < m_2} \quad (8)$$

$$\frac{e_1 \Downarrow m_1 \quad e_2 \Downarrow m_2}{e_1 = e_2 \Downarrow m_1 = m_2} \quad (9)$$

$$\frac{e_1 \Downarrow \text{tt} \quad e_2 \Downarrow v_2}{\text{if } e_1 \text{ then } e_2 \text{ else } e_3 \Downarrow v_2} \quad (10)$$

$$\frac{e_1 \Downarrow \text{ff} \quad e_3 \Downarrow v_3}{\text{if } e_1 \text{ then } e_2 \text{ else } e_3 \Downarrow v_3} \quad (11)$$

$$\frac{e_1 \Downarrow v_1 \quad e_2[x \mapsto v_1] \Downarrow v_2}{\text{let } x \text{ be } e_1 \text{ in } e_2 \Downarrow v_2} \quad (12)$$

$$\frac{e_1 \Downarrow (\text{fix } f(x:t_1):t_2 \text{ is } e'_1) \quad e_2 \Downarrow v_2 \quad (e'_1[x \mapsto v_2])[f \mapsto (\text{fix } f(x:t_1):t_2 \text{ is } e'_1)] \Downarrow v}{e_1 e_2 \Downarrow v} \quad (13)$$

$$\frac{e_1 \Downarrow v_1 \quad e_2 \Downarrow v_2}{\langle e_1, e_2 \rangle \Downarrow \langle v_1, v_2 \rangle} \quad (14)$$

$$\frac{e \Downarrow \langle v_1, v_2 \rangle}{\text{fst } e \Downarrow v_1} \quad (15)$$

$$\frac{e \Downarrow \langle v_1, v_2 \rangle}{\text{snd } e \Downarrow v_2} \quad (16)$$

$$\frac{e_1 \Downarrow v_1 \quad e_2 \Downarrow v_2}{(e_1 :: e_2) \Downarrow (v_1 :: v_2)} \quad (17)$$

$$\frac{e_1 \Downarrow \text{nil}[t] \quad e_2 \Downarrow v_2}{(\text{case } e_1 \text{ of nil } \Rightarrow e_2 \mid x::y \Rightarrow e_3) \Downarrow v_2} \quad (18)$$

$$\frac{e_1 \Downarrow (v_1 :: v_2) \quad (e_3[x \mapsto v_1])[y \mapsto v_2] \Downarrow v_3}{(\text{case } e_1 \text{ of nil } \Rightarrow e_2 \mid x::y \Rightarrow e_3) \Downarrow v_3} \quad (19)$$

Define the function

```
val evalSimple : exp -> exp
```

such that `evalSimple( $\lceil e \rceil$ )` returns  $\lceil v \rceil$  when  $e \Downarrow v$ , and raises the `Error` exception if the expression cannot be evaluated. Your code should have essentially the same form as the natural semantics. For example, to evaluate the expression  $e_1 + e_2$ , we recursively evaluate  $e_1$  and  $e_2$  to get  $v_1$  and  $v_2$  respectively; after checking that these are (tagged) integer constants the interpreter can return their (tagged) sum. To evaluate the expression `let  $x$  be  $e_1$  in  $e_2$` , we recursively evaluate  $e_1$  to get a  $v_1$ , substitute  $v_1$  into  $e_2$  (using `substClosed`) and return the result of evaluating the substituted expression.

Be careful when implementing the evaluation for `if` and for `case`; in both cases you should evaluate  $e_1$  first, and only evaluate one of  $e_2$  or  $e_3$  as appropriate.

The file `assign5.sml` contains the code for several cases of this function. Fix this code so that it returns the correct answer for the remaining cases instead of raising the `Unimplemented` exception.

## 4 Environments and Dynamic Scope

As discussed in class, a more efficient approach to implementation is to avoid substitution by simply remembering the current values of variables (that is, by maintaining an environment which maps variable names to values). This can also be modeled directly with a natural

semantics, as described in class. The relation  $(E, e) \Downarrow v$  defined by the following inference rules can be read as “ $e$  evaluates to  $v$  if the values of its free variables are given by the environment  $E$ ”. The notation  $E, x=v$  stands for the environment  $\eta$  extended by associating the value  $v$  with the variable  $x$ . The notation  $E(x)$  stands for the value which the environment  $E$  contains for the variable  $x$ .

$$\frac{}{(E, v) \Downarrow v} \quad (20)$$

$$\frac{}{(E, x) \Downarrow E(x)} \quad (21)$$

$$\frac{(E, e_1) \Downarrow \mathbf{m}_1 \quad (E, e_2) \Downarrow \mathbf{m}_2}{(E, e_1+e_2) \Downarrow \mathbf{m}_1+\mathbf{m}_2} \quad (22)$$

$$\frac{(E, e_1) \Downarrow \mathbf{m}_1 \quad (E, e_2) \Downarrow \mathbf{m}_2}{(E, e_1-e_2) \Downarrow \mathbf{m}_1-\mathbf{m}_2} \quad (23)$$

$$\frac{(E, e_1) \Downarrow \mathbf{m}_1 \quad (E, e_2) \Downarrow \mathbf{m}_2}{(E, e_1 * e_2) \Downarrow \mathbf{m}_1 \mathbf{m}_2} \quad (24)$$

$$\frac{(E, e_1) \Downarrow \mathbf{m}_1 \quad (E, e_2) \Downarrow \mathbf{m}_2}{(E, e_1 < e_2) \Downarrow \mathbf{m}_1 < \mathbf{m}_2} \quad (25)$$

$$\frac{(E, e_1) \Downarrow \mathbf{m}_1 \quad (E, e_2) \Downarrow \mathbf{m}_2}{(E, e_1 = e_2) \Downarrow \mathbf{m}_1 = \mathbf{m}_2} \quad (26)$$

$$\frac{(E, e_1) \Downarrow \mathbf{tt} \quad (E, e_2) \Downarrow v_2}{(E, \text{if } e_1 \text{ then } e_2 \text{ else } e_3) \Downarrow v_2} \quad (27)$$

$$\frac{(E, e_1) \Downarrow \mathbf{ff} \quad (E, e_3) \Downarrow v_3}{(E, \text{if } e_1 \text{ then } e_2 \text{ else } e_3) \Downarrow v_3} \quad (28)$$

$$\frac{(E, e_1) \Downarrow v_1 \quad ((E, x=v_1), e_2) \Downarrow v_2}{(E, \text{let } x \text{ be } e_1 \text{ in } e_2) \Downarrow v_2} \quad (29)$$

$$\frac{(E, e_1) \Downarrow (\text{fix } f(x:t_1):t_2 \text{ is } e'_1) \quad (E, e_2) \Downarrow v_2 \quad ((E, x=v_2, f=(\text{fix } f(x:t_1):t_2 \text{ is } e'_1)), e'_1) \Downarrow v}{(E, e_1 e_2) \Downarrow v} \quad (30)$$

$$\frac{(E, e_1) \Downarrow v_1 \quad (E, e_2) \Downarrow v_2}{(E, \langle e_1, e_2 \rangle) \Downarrow \langle v_1, v_2 \rangle} \quad (31)$$

$$\frac{(E, e) \Downarrow \langle v_1, v_2 \rangle}{(E, \text{fst } e) \Downarrow v_1} \quad (32)$$

$$\frac{(E, e) \Downarrow \langle v_1, v_2 \rangle}{(E, \text{snd } e) \Downarrow v_2} \quad (33)$$

$$\frac{(E, e_1) \Downarrow v_1 \quad (E, e_2) \Downarrow v_2}{(E, e_1 :: e_2) \Downarrow v_1 :: v_2} \quad (34)$$

$$\frac{(E, e_1) \Downarrow \text{nil}[t] \quad (E, e_2) \Downarrow v_2}{(E, \text{case } e_1 \text{ of nil} \Rightarrow e_2 \mid x::y \Rightarrow e_3) \Downarrow v_2} \quad (35)$$

$$\frac{(E, e_1) \Downarrow (v_1 :: v_2) \quad ((E, x=v_1, y=v_2), e_3) \Downarrow v_3}{(E, \text{case } e_1 \text{ of nil} \Rightarrow e_2 \mid x::y \Rightarrow e_3) \Downarrow v_3} \quad (36)$$

Write the function

```
evalDyn : (exp env) * exp -> exp
```

which takes an environment (mapping variable names to values) and an expression, and evaluates that expression using dynamic scoping as in the inference rules above. This function should also raise the `Error` exception if the expression cannot be evaluated.

The file `assign5.sml` contains the code for several cases of this function. Fix this code so that it returns the correct answer for the remaining cases instead of raising the `Unimplemented` exception.