Main topics of the week:
- A simple interpreter in ML
- Assignment in ML
- Type Inference

Grammar Basics
Suppose we want to have a grammar describing an arithmetic expression (say addition). The grammar might look like:

\[ \text{Expr ::= n | Expr + Expr} \]

We will assume that the goal is to have types so that we can define a function to perform the operation, i.e., a function that takes an Expr and returns an int. You might think we would do this in ML with the datatype:

- \text{datatype Expr = int | Expr * Expr;}

However, this is incomplete and gives a syntax error. To understand why ML doesn’t like this, consider the expression 5 – this is an int, so presumably it is an Expr. But an Expr is not an int. Rather we want to use constructors to show how to make an Expr from an int as well as from two Expr’s. To do this in ML, we define a data type as follows:

- \text{datatype Expr = N of int | P of Expr * Expr;}

(N is short for number, P is short for pair) This uses constructors to keep the types straight and is an instance of patterns. In particular, we cannot just use int since then int’s would be directly used in the function call and so would be evaluated by ML prior to the call. Note that these patterns are then used in the function definition, which is expressed as a disjunction of the disjuncts appearing in the type definition.

- \text{fun Interp (N x) = x}
- \text{val Interp = fn : Expr -> int}
- \text{Interp(P(N 4, N 5)); val it = 9 : int}
- \text{Interp(P(P(N 2, N 3), N 1)); val it = 6 : int}

Note that by defining things this way, evaluation is deferred until the call, i.e., the function performs the operation. Constructors are just ways of specifying types, they are actually implemented as functions that take the argument and just return it. The last of the expressions is essentially the arithmetic expression (2+3)+1. What we are really doing here is defining Interp to evaluate the addition operation, and using P to construct sum expressions.

Returning to the datatypes for the expression, we could expand it to cover two operations, e.g., addition and subtraction by (we shorten to just E instead of Expr):

- \text{datatype E = N of int | P of E*E | M of E*E;}
- \text{datatype E = M of E * E | N of int | P of E * E}

or

- \text{datatype binop = P | M;}
- \text{datatype E = N of int | B of binop*E*E;}
- \text{datatype E = B of binop * E * E | N of int}

Note these are all patterns of types. It’s a way to break up types into components. In a sense it suppresses evaluation and turns programs (code) into data (types).
Let’s try expanding the evaluation of the arithmetic expression, but have the “minus” operation produce an error.

```ml
- datatype E = N of int | P of E*E | M of E*E;
datatype E = M of E * E | N of int | P of E * E
- fun Interp(N x) = x
  = Interp(P(e1,e2)) = Interp(e1) + Interp(e2)
  = Interp = "error";
Error: right-hand-side of clause doesn't agree with function result type
```

Obviously, this doesn’t work. What is wrong is that there is a type mismatch – Interp sometimes returns an int, and sometimes a string. So we try defining a return type to capture this and adjusting the first and third clauses:

```ml
- datatype R = N of int | Err;
datatype R = Err | N of int
- fun Interp(N x) = N x
  = Interp(P(e1,e2)) = let val x = Interp(e1)
                        val y = Interp(e2)
                in case (x,y) of
                    (N1 p1, N1 p2) => N1(p1 + p2)
                  | _ => Err
                end
  = Interp = Err;
Error: operator and operand don’t agree
```

But we get the error since we don’t know how to “add” two R’s in the second clause. We correct this by introducing a new type for int in R for the return and adjusting the second clause to only perform the addition if we didn’t get an error on evaluating Interp on the components:

```ml
- datatype E = N of int | P of E*E | M of E*E;
datatype E = M of E * E | N of int | P of E * E
- datatype R = N1 of int | Err;
datatype R = Err | N1 of int
- fun Interp(N x) = N1 x
  = Interp(P(e1,e2)) = let val x = Interp(e1)
                        val y = Interp(e2)
                in case (x,y) of
                    (N1 p1, N1 p2) => N1(p1 + p2)
                  | _ => Err
                end
  = Interp = Err;
val Interp = fn : E -> R
- Interp(M(N 1, N 2));
val it = Err : R
- Interp(P(N 1, N 2));
val it = N1 3 : R
```

Alternatively, we could simply raise an exception:

```ml
- exception Err;
exception Err
- datatype E = N of int | P of E*E | M of E*E;
datatype E = M of E * E | N of int | P of E * E
- fun Interp(N x) = x
  = Interp(P(e1,e2)) = Interp(e1) + Interp(e2)
  = Interp = raise Err;
val Interp = fn : E -> int
```

**Changing without Assignment**

Since ML does not support assignment to variables (we’ll see side effects later), what can we do without assignment? Suppose we have a tree structure with integer values and we want to write a function that changes the values in the tree by adding one to each value. We can do this as follows:

```ml
- datatype Tree = Leaf of int | Node of int*Tree*Tree;
datatype Tree = Leaf of int | Node of int * Tree * Tree
```

Atkins – Spring 2004 - 2 - CIS 425 Programming Languages
The thing about this function is that it doesn’t really change anything – it actually produces a new copy of the tree with the values adjusted. This may seem inefficient, but the philosophy is that the programmer should try to program as elegantly as possible and not worry about this type of economy of storage or efficiency. The job of the compiler is to make this efficient. After all, the compiler is in a better position to know if doing an update in place is possible without violating the integrity of the program – it knows when things are referenced, for example.

Let’s reconsider the tree type, which is monomorphic, i.e., applies only to trees of integers. We can make it polymorphic as follows:

```ml
datatype 'a Tree = Leaf of 'a | Node of 'a * 'a Tree * 'a Tree;
```

Note that we use the notation (called alpha) of ‘a to indicate the unspecified type. Similarly, we can change the function to be polymorphic like this:

```ml
- fun change (f, Leaf x) = Leaf (f(x))
  = | change (f, Node (x, l, r)) = Node (f(x), change (f,l), change (f,r));
val change = fn : ('a -> 'b) * 'a Tree -> 'b Tree
```

Note that the return value of change is a beta tree, that is, ML cannot guarantee that it is of the same type as the tree passed to it. Also note that we have added an argument to change, namely a function from one unspecified type to another. We can use an anonymous function, e.g.,

```ml
- change(fn(x)=>true, Leaf 5);
val it = Leaf true : bool Tree
```

**Grammar for the Interpreter**

We want to write a simple interpreter to be able to experiment with the mechanisms of different scoping. This interpreter will be a function in ML that evaluates an expression. The expression will be given in abstract syntax. That is, we don’t want to mess with parsing input, so we will provide the input expression in an abstract form. So the first thing to do is to figure out a reasonable grammar for our language fragment and express it as datatypes in ML. We want our language to allow variables, constants, the addition operation, function definitions, local blocks, and function application. So we start with the grammar:

```ml
datatype E =
  Variable of identifier
| Literal of int
| Plus of E * E
| Lambda of identifier * E
| Lett of ((identifier * E) list) * E
| App of E * E;
```

Note that we have the components we just listed. These are all the forms an expression can take. A Variable is just an identifier (a string). The constants here (Literal) are just integers. We have one operation, Plus, that is a binary operation of two expressions. The function definition consists of a single parameter and an expression. The local block consists of a list of identifiers and expressions, followed by an expression. And finally, the application of a function consists of two expressions – the function name and the parameter.
A simple interpreter

We need a notion of an activation frame, so we define a type to encapsulate the environment. The environment entry is really just an identifier and a value, and the environment is just a list of these entries. This is defined mutually with a result type which we use for the value since we want to allow for errors and closure when we look up values in the environment.

```
datatype env = Env of (identifier * result) list
and   result = Value of int
         | Closure of E * env
         | Error of int;
```

We will define several utility functions for accessing and adding to the environment: lookup_env will search for a given id in a given environment, raising a not found exception if nothing is there.

Now we’re ready to look at the implementation of the interpreter. This is defined as a function interp1 (we wrap it with a function interp to handle the environment).

```
fun interp1(exp,env) =
  case exp of
    Variable(id)     => lookup_env(env,id)
  | Literal(Int)   => Value(Int)
  | Lambda(id,exp) => Closure(Lambda(id,exp),env)
  | Plus(E1, E2)   =>
    let val V1 = interp1(E1,env)
    val V2 = interp1(E2,env)
    in
    case (V1,V2) of (Value(X), Value(Y)) =>  Value(X+Y)
    | _ => raise err(BadArguments)
    end
  | App(E1, E2)  =>
    let val V1 = interp1(E1,env)
    val V2 = interp1(E2,env)
    in
    case V1 of Closure(Lambda(id,exp),clos_env) =>
      let val new_env = extend_env(clos_env,id,V2)
      in
      interp1(exp, new_env)
      end
    | _ => raise err(BadOperator)
    end
  | Lett(id_exp_list, E) =>
    let val id_result_list =
      map(fn (id,exp) => (id,interp1(exp,env))) id_exp_list
    in
    let val new_env = extend_env_all(env, id_result_list)
    in
    interp1(E, new_env)
    end
  end
```

So what can we notice about this interpreter? Given a Literal, it evaluates it. Given a Variable, it looks up the value. Given an addition operation, it invokes the interpreter to produce the values of the operands, then adds them. What about a function definition? Notice that it evaluates to a Closure – this is a way of capturing the environment for the function definition so that we can implement static scope. If we want dynamic scope, we could do away with the closure altogether. The Closure is used in the function application, where we call the interpreter to evaluate the function in the environment we get by evaluating the second expression (the
parameter to the function). Finally, the local block is achieved by using map to evaluate each expression in the block and adding it to the environment.

### Type Inference

Now we look at the chain of reasoning ML goes through to determine types. Of course, for a given expression, we can let ML tell us the type, but it is useful to see how this is done.

Consider:

```ml
fun apply (f, x) = f x;
```

What can we say about types here? We start with a first approximation:

\[ A \times B \rightarrow C \]

This just says that apply takes two arguments and gives back a value. We can observe that the right hand side is syntactically a function, hence

\[ A = E \rightarrow F \]

However, we can now also see from x that

\[ B = E \]

And from the right hand side that

\[ C = F \]

Putting these together, we reduce to

\[ A \times B \rightarrow C \]

\[ (E \rightarrow F) \times B \rightarrow C \]

\[ (E \rightarrow F) \times E \rightarrow C \]

\[ (E \rightarrow F) \times E \rightarrow F \]

That is, the function apply takes a function \( T_1 \rightarrow T_2 \) and a value \( T_1 \), and returns a value \( T_2 \), which is just what ML tells us:

```ml
-fun apply (f, x) = f x;
val apply = fn : ('a -> 'b) * 'a -> 'b
```

Here’s another example.

```ml
fun foo(nil, l) = l
  | foo(x::xs, l) = x::foo(xs, l);
```

Our first approximation is

\[ A \times B \rightarrow C \]

Further examination gives us:

\[ A = A_1 \text{ list} \quad \text{(from } x::xs) \]

\[ C = B_1 \text{ list} \quad \text{(from } x::\text{foo}(\ldots)) \]

\[ C = B \quad \text{(from first case)} \]

\[ A_1 = B_1 \quad \text{(from second case)} \]

Putting these together gives:

\[ A \times B \rightarrow C \]

\[ A_1 \text{ list } \times B \rightarrow B_1 \text{ list} \]

\[ A_1 \text{ list } \times B_1 \text{ list } \rightarrow B_1 \text{ list} \]

\[ A_1 \text{ list } \times A_1 \text{ list } \rightarrow A_1 \text{ list} \]

Which we verify in ML:

```ml
-fun foo(nil, l) = l
  | foo(x::xs, l) = x::foo(xs, l);
val foo = fn : 'a list * 'a list -> 'a list
```
(This is actually how we would code an append function.)

Type inference is an example of a theorem prover, where reasoning is done on what is known to reduce to the least number of unspecified types.

One more example:

\[
\text{fun foo}(f, \text{nil}) = \text{nil} \\
| \text{foo}(f, x::y) = (f x) :: \text{foo}(f, y);
\]

Here we start with

\[
A \times B \rightarrow C
\]

And proceed to observe

\[
\begin{align*}
B &= B1 \text{ list} & \text{(from args of second clause)} \\
A &= D \rightarrow E & \text{(from body of second clause)} \\
D &= B1 & \text{(from body of second clause)} \\
C &= C1 \text{ list} & \text{(from body of second clause)} \\
E &= C1 & \text{(since } f x \text{ is first elt of } C1 \text{ list)}
\end{align*}
\]

And reducing gives

\[
\begin{align*}
A \times B \rightarrow C \\
(D->E) \times B \rightarrow C \\
(D->E) \times B1 \text{ list} \rightarrow C \\
(D->E) \times D \text{ list} \rightarrow C \\
(D->E) \times D \text{ list} \rightarrow C1 \text{ list} \\
(D->E) \times D \text{ list} \rightarrow E \text{ list}
\end{align*}
\]

Which we verify with ML:

\[
\begin{align*}
- \text{fun foo}(f, \text{nil}) = \text{nil} \\
| \text{foo}(f, x::y) = (f x) :: \text{foo}(f, y); \\
\text{val foo} = \text{fn} : ('a \rightarrow 'b) * 'a \text{ list} \rightarrow 'b \text{ list}
\end{align*}
\]

and this is how we would define a map function on a list that produces a possibly different list type.

Suppose we had the expression

\[
\text{foo}(\text{fn } x:\text{int} \rightarrow \text{true}, [\text{true}, \text{false}]);
\]

Error: operator and operand don't agree

Note that this is not well typed since the second argument should be a list of int, not bool.

How about the function

\[
\text{fun foo } f = (f 1, f \text{ true});
\]

Error: operator and operand don't agree

Here the problem is we can’t change \( f \) midstream in the body. Once it is seen as one type, it must remain as that type. \( F \) cannot be both a function taking an int and a function taking a bool simultaneously. I.e., the polymorphism becomes monomorphism in the body. However, we can do something along these lines with:

\[
\begin{align*}
- \text{let val } f = \text{fn } x \rightarrow x \\
= \text{in} \\
= (f 1, f \text{ true}) \\
= \text{end;}
\text{val it} = (1, \text{true}) : \text{int} * \text{bool}
\end{align*}
\]
Here we have essentially specified a polymorphic template for f, but f has not been bound to a specific function, where in the previous example, the call to foo would bind f for the body.

Consider

```ml
- fun f g = let val x = g 1
  =               val y = g true
  =               in
  =               x = y
  = end;
```

ML gives us a typecheck error since g is constrained to type int->’a by val x and also to bool->’a by val y. However, we can try:

```ml
- let fun f x = x
  = in
  =    f 1;  f true
  = end;
```

and this will typecheck. The difference is that in the first example, g gets bound by the first use in val x and so is inconsistent with the next expression of val y. However, in the second example, f gets bound in each expression separately, so there is no inconsistency.

Side Effects in ML and calling conventions

We have seen there is no assignment in ML as we are used to in C. However, in ML we can declare reference types. A reference is like a pointer, actually it has the type of a polymorphic function - ref: ‘a -> a ref. There is also an operator in ML for de-referencing - !: ‘a ref -> ‘a, and also a special operator for assignment - := which has type ‘a ref -> ‘a -> int The following example illustrates these:

```ml
- val x = ref 5;
  val x = ref 5 : int ref
- ! x;
  val it = 5 : int
- x := !x + 2;
  val it = () : unit
- ! x;
  val it = 7 : int
- x := x + 1;
  Error: operator and operand don't agree
```

Contrast this to imperative languages where we have expressions like x = x + 1. ML requires that you be explicit about assignment, so the dereferencing on x is necessary to get a consistent type to add an integer to. Another way of looking at this is that ML distinguishes L values and R values. In C a variable can be an L value or R value, depending on its context in an expression. In ML, only a reference can be an L value, and it cannot be an R value. This means extra work for the programmer, but it also means the compiler can do a better job. The compiler is able to see what is mutable and what is not and use this information to advantage in optimization.

Side effects and call by name

Here’s a little example that you probably don’t want to run:
This runs forever because (f 5) keeps evaluating to f 5. If this were call by name, the evaluation would be delayed to the body of the function zero, so we wouldn’t loop on evaluation. In call by value, before we get to the body of zero, we must evaluate f 5, but since this keeps resulting in another value to evaluate, it goes on forever.

Consider the code:

```plaintext
val a = ref 10;
fun zero x = (a := !a + x);
fun f x = (x; x; !a);
f ( zero( 5));
```

Let’s analyze in a call by value language: replace zero(5) by (a := !a +5) and evaluate and pass to f. This evaluation causes a to refer to 15, and even though f evaluates 15 twice, the final value of a returned by the expression is just 15.

If ML were call by name, then the zero(5) would not get evaluated until the evaluation of the expression defining f. That is, it would get evaluated twice, and then the value of a printed out would be 20.

We have already noted that call by name and call by value may differ in efficiency. What we see here is that side effects may cause call by name and call by value to differ in results.

There is also a notion of call by need. What happens here is that the evaluation is only done if we need it, and the value is stored in case we need it again. For pure languages with no side effects, call by need and call by name are equivalent, but in the presence of side effects, they can produce different results.

Finally, we must be careful about the implementation of call by name. If our static scope is implemented by simple substitution, then it is possible for a value to be captured, and in this case it would give the effect of dynamic scope.

Let’s consider free variable capture again. Look at the following example:

```plaintext
let
  fun f x = let
    val y = 10
    in
    x + y
  end
  in
  let
    val y = 20
    in
    f (y + 1)
  end
end
```

In call by name with substitution, the evaluation is lazy, i.e., delayed. So if we substitute for f in f(y+1), we get essentially: let val y = 20 in let val y = 10 in (y+1) + y end end. And what happens is that y resolves to just the y=10, i.e., the y in ‘y+1’ is captured. So we need to fix this. The fix
A simple interpreter is to rename the first y to y1. Then in the expansion we just noted, we would see (y+1)+y1, which would give us the desired result. (Of course ML is call by value, not call by name with substitution, so we do get the desired result in real ML.)

Recall in our earlier discussion of L-value versus R-value that ML requires explicit distinction between the two; nothing is inferred as in C. Let’s consider call by value some more. In a call to a function where we have call by value, the value of the variable in the caller cannot change – the function cannot have side effects on the caller’s value (through the formal parameters, anyway). Local variables containing copies of the caller’s values are used in the function. Contrast to call by reference, where the function has access to the caller’s variable and so can affect its value. In this case, no local variables are used, and the function affects the caller’s values directly as execution proceeds. Although C is call by value, we get the effect of call by reference by passing addresses of variables and declaring the function interface to expect values which are pointer types, thus allowing the function access to the caller’s variables.

Now we consider another model: call by value result. In this model, there is a copy-in phase where the values are copied to local variables for use in the function body, just as in normal call by value. After execution, there is a copy-out phase, where the values in the local variables are copied back to the caller’s variables. So it’s somewhere in between call by value and call by reference. In most cases, it gives the same results as call by reference, but we’ll see where that is not always true. The idea of call by value result is to encapsulate the function, making its working opaque to the caller. Side note – we say that call by value is “eager” since it evaluates arguments even if they are not needed.

Now consider the following C code fragment:

```c
int i = 10;
void foo(int x, int y) { i = y; }
void main() {
    int A[20];
    i = 2; A[i] = 99;
    foo(i, A[i]);
}
```

If we had call by reference, the value of i would be 99 after execution.
If we had call by value result, the value of i would be 2 after execution. This happens because the formal parameter x passed to foo is not changed by the body of foo, hence the copy-out phase will restore i to the value (2) it had at the beginning of the call. So we see that call by reference and call by value result can indeed differ. This happens here because of aliasing, that is there are several names for the same variable – as the global variable i and as the formal parameter to foo. The language ADA has call by value result, but it turns out to just be implemented by call by reference. The claim is that aliasing is incorrect programming, so it doesn’t matter whether call by value result gives the right behavior! That is, in all “correct” programs, i.e., those with no aliasing, call by value result and call by reference would produce the same thing, so it’s okay to implement this way. And you wonder whether the government uses tax dollars effectively….

Here’s an example in Algol:

```algol
var x : int;
x = 0;
function p(y : int)
begin
    y = 1; x = 0;

Atkins – Spring 2004 - 9 -   CIS 425 Programming Languages
```
First note that in Algol, as in most procedural languages, the declaration and initialization are separate. Contrast to functional languages like ML, where the two are inseparable. Consider how the code fragment behaves:

- Call by value: $x$ is 0
- Call by reference: $x$ is 0 (it changes to 1, then back to 0)
- Call by value result: $x$ is 1 (the copy-out sets the formal parameter to its final value – in this case $y$ is the local variable, ends up with 1, and that is copied back to the caller’s parameter, $x$.)