CMSC 330	Exam #2	Fall 2005

Do not open this exam until you are told. Read these instructions:

1. This is a closed book exam. No notes or other aids are allowed.

2. You must turn in your exam immediately when time is called at the end.

3. This exam contains 6 pages, including this one. Make sure you have all the pages. Each question’s point value is next to its number. Write your name on the top of all pages before starting the exam.

4. In order to be eligible for as much partial credit as possible, show all of your work for each problem, and clearly indicate your answers. Credit cannot be given for illegible answers.

5. If you finish at least 15 minutes early, bring your exam to the front when you are finished; otherwise, wait until the end of the exam to turn it in. Please be as quiet as possible.

6. If you have a question, raise your hand. If you feel an exam question assumes something that is not written, write it down on your exam sheet. Barring some unforeseen error on the exam, however, you shouldn’t need to do this at all, so be careful when making assumptions.

7. If you need scratch paper during the exam, please raise your hand. Scratch paper must be turned in with your exam, with your name and ID number written on it. Scratch paper will not be graded.

8. Small syntax errors will be ignored in any code you have to write on this exam, as long as the concepts are correct.

9. The Campus Senate has adopted a policy asking students to include the following handwritten statement on each examination and assignment in every course: “I pledge on my honor that I have not given or received any unauthorized assistance on this examination.” Therefore, just before turning in your exam, you are requested to write this pledge in full and sign it below:

________________________________________________________________________

Good luck!
1. [20 pts.] **Short Answer.**
   
   a. [9 pts.] Below is the output of the OCaml interpreter after we’ve typed in various definitions of the function $f$. In each case, write a definition of a function $f$ that has that type.
   
   i. `val f : 'a -> 'b -> 'a * 'a * 'b = <fun>`
   
   Answer: let $f$ $x$ $y$ = ($x$, $x$, $y$)
   
   ii. `val f : ('a -> 'b) -> ('b -> 'c) -> 'a -> 'c = <fun>`
   
   Answer: let $f$ $x$ $y$ $z$ = $y$ ($x$ $z$)
   
   iii. `val f : 'a -> 'b = <fun>`
   
   Answer: let rec $f$ $x$ = $f$ $x$
   
   Notice that because $f$ never terminates, its result type could be anything.

   b. [5 pts.] Define what it means for one name to **shadow** another, and give an example of shadowing in the language of your choice.
   
   **Answer:** Shadowing occurs when one name hides another name that would otherwise be in scope. For example, in OCaml, the inner $x$ shadows the outer $x$ in the following function:
   
   ```ocaml```
   let $f$ $x$ = (fun $x$ -> $x$ + $x$)
   ```
   
   Here's an example in C:
   
   ```c```
   void foo(int $x$) { int $x$ = 3; $x$ = $x$ + $x$; }
   ```
   
   See the lecture slides for more examples.

   c. [6 pts.] Suppose in C we define the following function:
   
   ```c```
   int and(int $x$, int $y$) { return $x$ && $y$; }
   ```
   
   Is it ever possible that calling `and($e_1$, $e_2$)` behaves differently than evaluating `$e_1$ && $e_2$` directly, where $e_1$ and $e_2$ are C expressions? Explain, and give an example if so.
   
   **Answer:** Yes. $e_1$ && $e_2$ is short-circuiting and won’t evaluate $e_2$ is $e_1$ is false. But functions are call-by-value, so `and($e_1$, $e_2$)` will always evaluate both $e_1$ and $e_2$. For example, suppose we define `int $f$(void) { printf("hello"); return 3; }`. Then 0 && $f$(x) will not print anything, but `and(0, $f$(x))` will.
2. [40 pts.] **Context–Free Grammars.**

   a. [10 pts.] Consider the short grammar we gave in class for expressions in the simply–typed lambda calculus:

```
S ::= U | V | λV:T.S | S S
T ::= int | T → T
U ::= 1 | 2 | 3
V ::= x | y | z
```

Here we have simplified the grammar a bit by restricting integers to just 1, 2, and 3, and variable names to just x, y, and z, and we just used different nonterminal names than in the grammar given in class.

Prove that this grammar is ambiguous.

**Answer:**

There are two sources of ambiguity in this grammar. One is the production for function application: $S ::= S S$

The other is the production for function types $T ::= T → T$

Ambiguity could be shown by showing two leftmost or two rightmost derivations for the same string (using either source of ambiguity), or two parse trees for the same string (which may have been quicker than writing two derivations).

Here is one ambiguity, shown by having two parse trees for the same string $x \ y \ z$:

Two leftmost or two rightmost derivations could be shown for the same string, or for any other ambiguous string, as well. The two leftmost derivations for this string are:

$S \Rightarrow S \ S \Rightarrow S \ S \ S \Rightarrow x \ S \ S \Rightarrow x \ y \ S \Rightarrow x \ y \ z$

and

$S \Rightarrow S \ S \Rightarrow x \ S \Rightarrow x \ S \ S \Rightarrow x \ y \ S \Rightarrow x \ y \ z$
b. [30 pts.] Now, write a grammar for the same language from part (a) that is **not ambiguous**. In particular,

- Your grammar should generate exactly the same expressions as from part (a)—no more, and no less.
- Function definitions extend as far to the right as possible. For example, \( \lambda x: \text{int}.x \ 3 \) should have the interpretation \( \lambda x: \text{int}.(x \ 3) \).
- Function application is **left–associative**. For example, your grammar should interpret \( \lambda x: \text{int}. \ x \ 1 \ y \ 2 \) as \( \lambda x: \text{int}.(( (x \ 1) \ y) \ 2) \).
- Function types are **right–associative**. For example, your grammar should interpret \( \text{int} \to \text{int} \to \text{int} \) as \( \text{int} \to (\text{int} \to \text{int}) \).

(Note that lambda calculus expressions in this problem may not contain parentheses; we just used them for clarification above.)

Some examples (this is for illustration, and does **not** illustrate all possible valid expressions):

<table>
<thead>
<tr>
<th>expression</th>
<th>interpretation your grammar should give the expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>x 1</td>
<td>x 1</td>
</tr>
<tr>
<td>x 1 y 2 z 3</td>
<td>(((((x 1) y) 2) z) 3</td>
</tr>
<tr>
<td>( \lambda y: \text{int} \to \text{int} \cdot y )</td>
<td>( \lambda y: \text{int} \to \text{int} \cdot y )</td>
</tr>
<tr>
<td>( \lambda x: \text{int} \to \text{int} \cdot \lambda y: \text{int} \cdot x \ y \ x )</td>
<td>( \lambda x: \text{int} \to (\lambda y: \text{int}.((x \ y) \ x)) )</td>
</tr>
</tbody>
</table>

In order for it to be graded as accurately and quickly as possible, your grammar should use consecutive nonterminals beginning with S (S, T, U, etc.).

**Answer:**

It turned out that the second requirement above, that function definitions extend as far to the right as possible, is a little tricky to enforce, so we just ignored it in grading, and only required that your answer be unambiguous and enforce function application to be left–associative and function types to be right–associative. Here’s a correct solution:

\[
\begin{align*}
S & ::= T \ U \mid T \mid U \\
T & ::= T \ W \mid W \\
U & ::= \lambda X: V . S \\
V & ::= \text{int} \mid \text{int} \to V \\
W & ::= 1 \mid 2 \mid 3 \mid X \\
X & ::= x \mid y \mid z
\end{align*}
\]
3. [40 pts.] **Evaluating Java Bytecode.** In this problem you will write an OCaml function to evaluate Java bytecodes. The Java Virtual Machine, which executes bytecodes, is stack-based, so that bytecode instructions operate on a stack, pushing and popping values, which for this problem will only be integers. Java bytecodes can also access a set of local variables, which are numbered.

The following table describes the eight opcodes you will implement for this problem. The right-most column is a succinct picture of what the instructions do. Stacks are written with the top of the stack to the left and . . . used to indicate the portion of the stack that doesn’t change. V is the local variable lookup table, and V(n) refers to the value which variable number n maps to in V. The ⇒ means that the initial stack’s contents are shown to the left of the arrow, and the stack’s contents after the instruction is executed are shown to the right of the arrow. So, for example, the notation “. . . ⇒ V(n), . . .” in the 2nd to last row of the table means “look up n in the table V, and push the result onto the stack.” The last row’s notation means “remove the top value from the stack, and store whatever that was in V(n).”

<table>
<thead>
<tr>
<th>Instruction</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>iconst x</td>
<td>Push the value x onto the stack</td>
<td>. . . ⇒ x, . . .</td>
</tr>
<tr>
<td>iadd</td>
<td>Add two ints on top of the stack</td>
<td>x, y, . . . ⇒ x + y, . . .</td>
</tr>
<tr>
<td>dup</td>
<td>Duplicate the value on the top of the stack</td>
<td>x, . . . ⇒ x, x, . . .</td>
</tr>
<tr>
<td>nop</td>
<td>Do nothing</td>
<td>. . . ⇒ . . .</td>
</tr>
<tr>
<td>pop</td>
<td>Pop the top value off of the stack and discard it</td>
<td>x, . . . ⇒ . . .</td>
</tr>
<tr>
<td>swap</td>
<td>Swap the top two operand stack values</td>
<td>x, y, . . . ⇒ y, x, . . .</td>
</tr>
<tr>
<td>iload n</td>
<td>Push the value in local variable #n onto the stack</td>
<td>. . . ⇒ V(n), . . .</td>
</tr>
<tr>
<td>istore n</td>
<td>Pop the top value off of the stack and store it in local variable #n</td>
<td>x, . . . ⇒ . . . V(n) := x</td>
</tr>
</tbody>
</table>

Below we give an OCaml data type `instr` representing Java bytecodes. You job is to write two functions. First, write

\[
\text{eval_instr : int list * instr -> int list}
\]

which, given a stack and an instruction, returns a new stack according to the behavior of the instruction. The stack is a list of integers with the head of the list as the top of the stack. Then write

\[
\text{eval : int list * instr list -> int list}
\]

which, given a stack and a list of instructions, returns a new stack which is the result of evaluating each of the instructions in turn, in order from the first instruction in the list to the end of the list. (You should find this second function much easier to write.)

- It is an error to try to read a local variable that has not been written to, and your solution may do anything if an instruction tries to do this.
- It is an error to try to pop from an empty stack, and your solution may do anything if an instruction tries to do this.
- You **may not use OCaml library functions or arrays** for this question. In particular, if you choose to use associative lists to represent V, then you will need to give code that implements them.

Here is the data type for instructions:

\[
\text{type instr =}
\]

\[
\text{IConst of int | IAdd | Dup | Nop | Pop | Swap | ILoad of int | IStore of int}
\]
Here is a sample run of eval:

```ocaml
# let l = [IConst 3; IConst 4; IAdd];;
val l : instr list = [IConst 3; IConst 4; IAdd]
# eval ([], l);
- : int list = [7]
```

Write your functions eval_instr and eval, as well as any other functions or declarations you may need, here:

**Answer:**

(* returns the value corresponding to the first occurrence of n in assoc list *)

```ocaml
let rec lookup n = function
  | [] -> raise Not_found (* optional *)
  | ((k, v)::t) -> if n = k then v else lookup v t

let v = ref []

let eval_instr = function
  | (s, IConst x) -> x::s
  | (x::y::s, IAdd) -> (x+y)::s
  | (x::s, Dup) -> x::x::s
  | (s, Nop) -> s
  | (x::s, Pop) -> s
  | (x::y::s, Swap) -> y::x::s
  | (s, ILoad n) -> (lookup n (!v))::s
  | (x::s, IStore n) -> begin
      v := (n, x)::(!v); (* add new value of n *)
      s (* evaluate to result stack *)
    end

let rec eval = function
  | (s, []) -> s
  | (s, i::is) -> eval (eval_instr (s, i), is)
```