Formalizing Real World Programming Languages with Skeletal Semantics

Alan Schmitt

September 18, 2023

Wat

A lightning talk by Gary Bernhardt from CodeMash 2012

```
failbowl:~(master!?) $ jsc
> [] + []
> [] + {}
[object Object]
0
NaN
```





Sergio Maffeis

EPSRC Research Fellow, Imperial College London. Currently visiting INRIA Rocquencourt.

In collaboration with:

John Mitchell, Ankur Taly (Stanford University), Philippa Gardner, Gareth Smith (Imperial College London).

Rennes, November 4, 2011.

JSCert

Two JavaScript semantics in Coq descriptive given a program and a result, say if they are related executable given a program, compute the result

Correctness

If program P executes to v, then P and v are related

- 2 years, 8 people
- 18 klocs of Coq







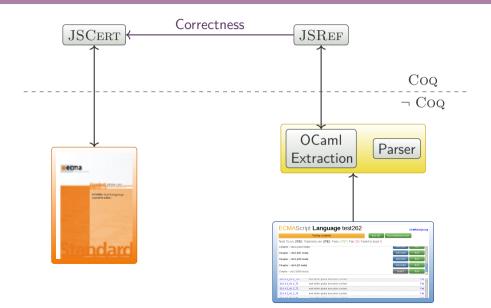












,

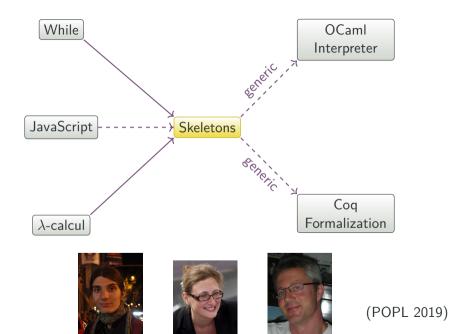
Reading JSCert

! red_stat_while : foral) S C labs e1 t2 o, red_stat S C (stat_while_1 labs e1 t2 resvalue_empty) o -> red stat S.C. (stat while labs et t2) o I red_stat_while_1 : forall S C labs e1 t2 rv u1 o. ed_stat_wmile_1 : foreit S C tabs el t2 rv yl 0, red_spec S C (spec_expr_get_value_conv spec_to_boolean el) yl -> red_stat S C (stat_while_2 labs el t2 rv yl) 0 -> red stat S C (stat while 1 labs el t2 rv) o I red_stat_while_2_false : forall 80 S C labs e1 t2 rv, red_stat 80 C (stat_while_2 labs e1 t2 rv (vret 8 false)) (out_ter 8 rv) 12.6.2 The while Statement I red_stat_while_2_true : forall S0 S C labs e1 t2 rv o1 o. The production IterationStatement: while (Expression) Statement is evaluated as follows: red_stat S C t2 o1 -> red_stat S C (stat_while_3 labs e1 t2 rv o1) o -> red_stat SB C (stat_while_2 labs e1 t2 rv (vret S true)) o Let V = empty. I red stat while 3 : forall rv SB S C labs e1 t2 rv' R o. 2. Repeat rv' = (If res_value R <> resvalue_empty then res_value R else rv) -> red_stat S C (stat_while_4 labs el t2 rv' R) o -> a. Let exprRef be the result of evaluating Expression. red_stat SB C (stat_while_3 labs e1 t2 rv (out_ter S R)) o If ToBoolean(GetValue(exprRef)) is false, return (normal, V, empty). I red_stat_while_4_continue : forall S C labs e1 t2 rv R o. c. Let stmt be the result of evaluating Statement. res_tupe R = restupe_continue /\ res_label_in R labs -> red stat S.C. (stat while 1 labs et t2 rv) n => d. If stmt value is not empty, let V = stmt value. red stat S C (stat while 4 labs et t2 rv P) o e. If stmt.type is not continue || stmt.target is not in the current label set, then I red_stat_while_4_not_continue : forall S C labs e1 t2 rv R o. If stmt, type is break and stmt, target is in the current label set, then "(res_table_in R_labs) -> red_stat SC (stat_while_5 labs e1 t2 rv R) o -> red_stat SC (stat_while_4 labs e1 t2 rv R) o 1 Return (normal V empty) If stmt is an abrupt completion, return stmt. I red_stat_while_5_break : forall S C labs e1 t2 rv R. ed_stat_wmlie_5_break : forall 5 U labs el t2 rv k, res_type R = restype_break /\ res_label_in R labs -> red_stat S C (stat_while_5 labs el t2 rv R) (out_ter S rv) I red stat while 5 not break : forall 5 C labs of t2 rv P o. (res_tupe R = restupe_break /\ res_label_in R labs) -> red stat S C (stat while 6 labs e1 t2 rv R) o -> red stat S C (stat while 5 labs el t2 rv P) o I red_stat_while_6_abort : forall S C labs e1 t2 rv R. res_type R <> restype_normal -> red_stat S C (stat_while_6 labs e1 t2 rv R) (out_ter S R) I red stat while 6 normal : forall S.C. labs et t2 ry R.o. res_type R = restype_normal -> red_stat S C (stat_while_1 labs e1 t2 rv) o -> red_stat S C (stat_while_6 labs e1 t2 rv R) o

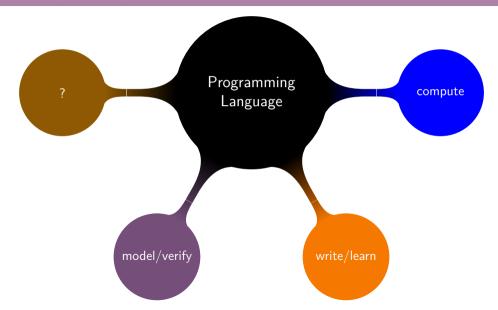
JSCert: The Problem

```
(** If statement (12.5) *)
red_stat_if : forall S C e1 t2 t3opt y1 o,
   red_spec S C (spec_expr_get_value_conv spec_to_boolean e1) y1 ->
   red_stat S C (stat_if_1 y1 t2 t3opt) o ->
   red_stat S C (stat_if e1 t2 t3opt) o
 red_stat_if_1_true : forall SO S C t2 t3opt o,
   red stat S C t2 o ->
   red stat SO C (stat if 1 (vret S true) t2 t3opt) o
red stat if 1 false : forall SO S C t2 t3 o.
   red stat S C t3 o ->
   red_stat SO C (stat_if_1 (vret S false) t2 (Some t3)) o
 red_stat_if_1_false_implicit : forall SO S C t2,
   red stat SO C (stat if 1 (vret S false) t2 None) (out ter S resvalue empty)
```

- 900 mutually inductive rules
- inversion during an induction runs out of memory



The Meaning of Programs



C

Goals

What

- Computable
- Readable
- Maintainable
- Usable

How

- Syntactic description of semantics
- No silo, always have an escape hatch
- Simple, to be extended & reused
- Supports non-determinism and partiality
- Incremental or incomplete specification

Why Not?

- OCaml, Haskell, Scheme
- Coq, Isabelle
- Lem, Ott
- K

Coupling too tight between definition and use

Skeletal Semantics

Specifying the Syntax of a Language

```
x \in Ident
e ::= n | x | e + e
    | e = e | !e
s ::= skip \mid x := e \mid s; s
     if e then s else s
     while e do s
```

 $n \in Lit$

Specifying the Syntax of a Language

```
(* unspecified type *)
n \in Lit
                              type lit
x \in Ident
                              type ident
e ::= n | x | e + e
   |e=e|!e
s ::= skip \mid x := e \mid s; s
    | if e then s else s
     while e do s
```

Specifying the Syntax of a Language

```
(* unspecified type *)
n \in Lit
                           type lit
x \in Ident
                           type ident
                           (* specified type *)
                                                       (* specified type *)
e ::= n | x | e + e
                           type expr =
                                                       type stmt =
   |e=e|!e
                             Const lit
                                                       Skip
                            | Var ident
                                                       | Assign (ident, expr)
s ::= skip \mid x := e \mid s; s
                           | Plus (expr, expr)
                                                       | Seq (stmt, stmt)
                           | Equal (expr, expr)
                                                       | If (expr, stmt, stmt)
    if e then s else s
                             Not expr
                                                         While (expr, stmt)
    while e do s
```

$$b ::= tt \mid ff$$
 $v ::= n \mid b$
 $\sigma \in State$

$$\sigma, e \downarrow_e v$$
 $\sigma, s \downarrow_s \sigma$

```
b ::= tt \mid ff
v ::= n \mid b
\sigma \in State
```

```
type boolean = | True | False
type int
type value = | Int int | Bool boolean
type state
```

```
\sigma, e \Downarrow_e v
\sigma, s \Downarrow_s \sigma
```

```
(* specified term *)
val eval_expr ((st:state), (e:expr)) : value =
```

```
(* unspecified term *)
val int_of_lit: lit → int
```

```
(* specified term *)
            \sigma, n \downarrow _{\rho} n
                                  val eval expr ((st:state), (e:expr)) : value =
   \sigma, e_1 \downarrow_e n_1 \sigma, e_2 \downarrow_e n_2
                                     let Const n = e in
     \sigma, e_1 + e_2 \downarrow e_n n_1 + n_2
                                     let i = int of lit n in
                                     Int i
                                    let Plus(e1,e2) = e in
                                     let Int n1 = eval_expr (st, e1) in
                                     let Int n2 = eval_expr (st, e2) in
(* unspecified term *)
                                     let n = add (n1, n2) in
val int_of_lit: lit → int
                                    Int n
val add: (int, int) → int
```

```
(* specified term *)
           \sigma, n \downarrow _{\rho} n
                                 val eval expr ((st:state), (e:expr)) : value =
                                 branch
   \sigma, e_1 \downarrow_e n_1 \sigma, e_2 \downarrow_e n_2
                                    let Const n = e in
     \sigma, e_1 + e_2 \downarrow_e n_1 + n_2
                                    let i = int of lit n in
                                    Int i
                                 or
                                   let Plus(e1,e2) = e in
                                    let Int n1 = eval_expr (st, e1) in
                                    let Int n2 = eval_expr (st, e2) in
(* unspecified term *)
                                 let n = add (n1, n2) in
val int_of_lit: lit → int
                                Int n
                                 end
val add: (int, int) → int
```

```
\frac{\sigma, e \Downarrow_e ff}{\sigma, \text{while } e \text{ do } s \Downarrow_s \sigma}
\frac{\sigma, e \Downarrow_e \text{ tt} \qquad \sigma, s \Downarrow_s \sigma'}{\sigma', \text{while } e \text{ do } s \Downarrow_s \sigma''}
\frac{\sigma', \text{while } e \text{ do } s \Downarrow_s \sigma''}{\sigma, \text{while } e \text{ do } s \Downarrow_s \sigma''}
```

```
val eval stmt ((st:state), (s:stmt)) : state =
branch
  let While(e, s') = s in
  let Bool False = eval expr (st, e) in
  st
or
  let While(e, s') = s in
  let Bool True = eval_expr (st, e) in
  let st' = eval stmt (st, s') in
  eval_stmt (st', s)
or ...
end
```

```
\frac{\sigma, e \Downarrow_e ff}{\sigma, \text{while } e \text{ do } s \Downarrow_s \sigma}
\frac{\sigma, e \Downarrow_e \text{ tt} \qquad \sigma, s \Downarrow_s \sigma'}{\sigma', \text{while } e \text{ do } s \Downarrow_s \sigma''}
\frac{\sigma', \text{while } e \text{ do } s \Downarrow_s \sigma''}{\sigma, \text{while } e \text{ do } s \Downarrow_s \sigma''}
```

```
val eval stmt ((st:state), (s:stmt)) : state =
branch
  let While(e, s') = s in
  let Bool b = eval expr (st, e) in
  branch
    let False = b in st
  or
    let True = b in
    let st' = eval stmt (st, s') in
    eval stmt (st', s)
  end
or ...
end
```

Higher Order

```
val eval_stmt ((st:state), (s:stmt)) : state = ...
(* is syntactic sugar for *)
val eval stmt : (state, stmt) → state =
  \lambda (st, s) : (state, stmt) \rightarrow ...
val app: nat \rightarrow (nat \rightarrow nat) \rightarrow nat =
  \lambda x: nat \rightarrow
  \lambda f: (nat \rightarrow nat) \rightarrow
  f x
```

Polymorphism¹

```
type list<a> =
| Nil | Cons (a, list<a>)
val map\langle a, b \rangle ((f: (a \rightarrow b)), (l: list\langle a \rangle)) : list\langle b \rangle =
  branch
     let Nil = 1 in Nil<b>
  or
     let Cons(x, xs) = 1 in
     let y = f x in
     let ys = map < a, b > (f, xs) in
     Cons<b>(y, ys)
  end
```

Monads

Language Monads

Polymorphism + first class functions is sufficient for monads type $st < a > = state \rightarrow (a, state)$ val ret < a > (v: a) : st < a > = λ s:state \rightarrow (v, s) val bind<a, b> $((w: st<a>), (f: a \rightarrow st)) : st =$ λ s:state \rightarrow let (v, s') = w s inlet w' = f v in w's'

GetValue(V)

- ReturnIfAbrupt(V).
- If V is not a Reference Record, return V.
- If IsUnresolvableReference(V) is true, throw a ReferenceError exception.
- If IsPropertyReference(V) is true, then
 - Let baseObj be ! ToObject(V.[[Base]]).
 - Return ? baseObj.[[Get]](V.[[ReferencedName]], GetThisValue(V)).
- Else,
 - Let base be V.[[Base]]
 - Assert: base is an Environment Record.
 - Return ? base.GetBindingValue(V.[[ReferencedName]], V.[[Strict]]).

State Monad

```
type st<a> := state → (a, state)
val ret < a > (v: a) : st < a > =
  \lambda s:state \rightarrow (v. s)
val bind<a, b> ((w: st<a>), (f: a \rightarrow st<b>)) : st<b> =
  \lambda s:state \rightarrow
  let (v, s') = w s in
  let w' = f v in
  w's'
val eval expr (e:expr) : st<value> = ...
val eval_stmt (s:stmt) : st<()> = ...
```

While in State Monad

```
val eval_stmt (s:stmt) : st<()> =
branch
  let While(e, s') = s in
  let w = eval expr e in
  bind<value, ()> (w, \lambda Bool b : value \rightarrow
  branch
    let False = b in ret<()> ()
  or
    let True = b in
    let w' = eval_stmt s' in
    bind<(),()> (w', \lambda () \rightarrow eval stmt s)
  end)
or ...
end
```

While in State Monad

```
val eval_stmt (s:stmt) : st<()> =
branch
 let While(e, s') = s in
  let Bool b =%bind eval_expr e in
  branch
    let False = b in ret<()> ()
  or
    let True = b in
    let () =%bind eval_stmt s' in
    eval stmt s
  end
or ...
end
```

While in State Monad

```
binder @ := bind
val eval stmt (s: stmt): st<()> =
branch
  let While (e, s') = s in
 let Bool b =@ eval_expr e in
  branch
    let False = b in ret<()> ()
  or
    let True = b in
    eval_stmt t';@
    eval_stmt t
  end
or ...
end
```

The Monad Zoo

- reader monad (environment)
- writer monad (log)
- option monad (exceptions, simpler control flow)
- state monad (heap)
- delimited continuation monad (generators, effects)

GetValue(V)

- ReturnIfAbrupt(V).
- If V is not a Reference Record, return V.
- If IsUnresolvableReference(V) is true, throw a ReferenceError exception.
- If IsPropertyReference(V) is true, then
 - Let baseObj be ! ToObject(V.[[Base]]).
 - Return ? baseObj.[[Get]](V.[[ReferencedName]], GetThisValue(V)).
- Else,
 - Let base be V.[[Base]]
 - Assert: base is an Environment Record.
 - Return ? base.GetBindingValue(V.[[ReferencedName]], V.[[Strict]]).

GetValue(V)

```
val getValue: (v: out<valref>) -> st<out<value>> =
 let result =0r
   let v =%returnIfAbrupt v in
   branch valref_Type(v, T_Reference); @false let Value v = v in ret v end; @
   let Reference v = v in
   branch isUnresolvableReference v;@true throw referenceError _getValue_ end;@
   branch let T = isPropertyReference v in
          let R Value v base = v. Base in let baseObj =! toObject(v base) in
          let baseObi =/o baseObi in
          let thisVal =? getThisValue(v) in
          let r =? baseObj. O Get (v. ReferencedName , thisVal) in ret r
          let F = isPropertyReference v in
   or
          let base = v. Base in
           assT ref Type (base, T R EnvironmentRecord) getValue ;@
           let R EnvironmentRecord base = base in let base =/er base in
           let r =? base. GetBindingValue (v. ReferencedName , v. Strict ) in ret r
   end
 in result
```

Delimited Continuations

In Languages

- generators (JavaScript, Python)
- effects (OCaml 5)
- and also...

JavaScript Function Calls

13 3 6 1 Runtime Semantics: Evaluation

```
• Return ? EvaluateCall(func, ref, arguments, tailCall)
13.3.6.2 EvaluateCall (func, ref, arguments, tailPosition)
 • Let result be Call(func, thisValue, argList)
10.2.10 FunctionDeclarationInstantiation (func, argumentsList)
 • Let iteratorRecord be CreateListIteratorRecord(argumentsList)
```

Iterators as Generators

7.4.9 CreateListIteratorRecord (list)

- Let closure be a new Abstract Closure with no parameters that captures list and performs the following steps when called:
 - For each element E of list, do
 - Perform ? Yield(E).
 - Return undefined.
- Let iterator be ! CreateIteratorFromClosure(closure, empty, %IteratorPrototype%).

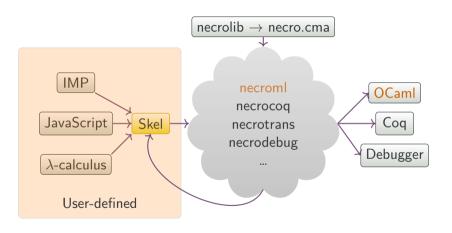
Whole Monad

```
type exc < a > = | Exc | Ok a
type rd < a > = env \rightarrow a
type st < a > = heap \rightarrow (a, heap)
type cont < a > = a -> cstack < a > -> a
type cstack<a> =
l Nil
Cons(cont<a>.cstack<a>)
type contM<a> = cont<a> -> cstack<a> -> a
tvpe ste<a> = st<exc<a>>
type k<a> = cont<ste<a>>
tvpe cs<a> = cstack<ste<a>>
type m<a> = rd<contM<ste<a>>>
```

```
val return<a> (v:a) : m<a> =
       \lambda:env \rightarrow \lambda k: k < a > \rightarrow \lambda ks: cs < a > \rightarrow \lambda h: heap <math>\rightarrow
       k (\lambda h':heap \rightarrow (0k < a > v, h')) ks h
val bind<a> (w:m<a>) (f:a \rightarrow m<a>) : m<a> =
       \lambda s:env \rightarrow \lambda k:k\langle a \rangle \rightarrow \lambda ks:cs\langle a \rangle \rightarrow \lambda h:heap \rightarrow
       w s (\lambdaste:ste<a> \rightarrow \lambdaks1:cs<a> \rightarrow \lambdah1:heap \rightarrow
              let (vo, h2) = ste h1 in
              match vo with
               | Exc \rightarrow k (\lambdah3:heap \rightarrow (Exc<a>,h3))
                                 ks1 h2
               | 0k v \rightarrow f v s k ks1 h2
              end) ks h
```

Necro ML

Necro Ecosystem



Necro ML

- Write the skeletal semantics
- Write a module implementing unspecified types and terms
- Ohoose how to interpret branches
- Run necroml and apply the MakeInterpreter functor
- Profit!

Necro ML Example

```
(* arith.sk *)
                                         open Arith (* file generated with necroml *)
                                         module Types = struct
type lit
                                           type lit = int
type value
                                          type value = int
                                         end
                                         module Input = struct
                                           include Unspec(Monads.ID)(Types)
val litToVal· lit → value
                                          let litToVal l = l
val add: (value, value) → value
                                          let add (11, 12) = 11 + 12
val sub: (value, value) → value
                                        let sub (11, 12) = 11 - 12
val mul: (value, value) → value
                                       let mul (11, 12) = 11 * 12
val div: (value, value) → value
                                        let div (11, 12) = 11 / 12
                                         end
(* next are specified types
   and terms *)
                                         module ArithInterp = MakeInterpreter(Input)
```

Interpretation Monads

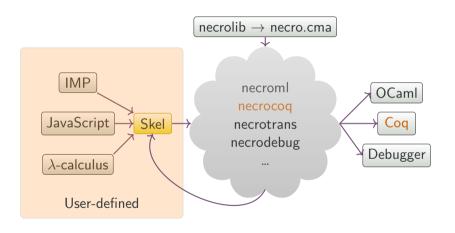
Terms are pure, skeletons are computations

```
module type MONAD = sig
  type 'a t
  val ret: 'a -> 'a t
  val bind: 'a t -> ('a -> 'b t) -> 'b t
  val branch: (unit -> 'a t) list -> 'a t
  val fail: string -> 'a t
  val apply: ('a -> 'b t) -> 'a -> 'b t
  val extract: 'a t -> 'a
end
```

- Main challenge: how to interpret branching?
- Instantiations: sequential, lists, non-deterministic, continuations

Necro Coq

Necro Ecosystem



A Deep Embedding in Coq (Skeleton.v)

```
Inductive term: Type :=
 term_constructor : string -> list type -> term -> term
 term_var: typed_var -> term
 term_tuple: list term -> term
 term func: pattern -> skeleton -> term
with skeleton: Type :=
 skel branch : type -> list skeleton -> skeleton
 skel match : term -> type -> list (pattern * skeleton) -> skeleton
 skel return : term -> skeleton
 skel apply : term -> list term -> skeleton
 skel letin : pattern -> skeleton -> skeleton -> skeleton.
```

Dynamic Semantics

• Concrete.v, natural (big-step) semantics using Coq induction

```
| i_letin: forall e e' p s1 s2 v w,
   interp_skel e s1 v ->
   add_asn e p v = Some e' ->
   interp_skel e' s2 w ->
   interp_skel e (skel_letin p s1 s2) w
```

- Concrete_ss.v, small-step semantics
- ConcreteRec.v, iterative semantics
- Concrete_ndam.v, non-deterministic abstract machine
- Concrete_am.v, backtracking abstract machine, can compute
- many equivalence proofs

Conclusion

Current Status

- Programming Languages
 - WebAssembly (Thomas Rubiano)
 - JavaScript (Adam Khayam)
 - Python (Martin Andrieux)
- Many applications
 - Generation of OCaml interpreter and Coq formalization (Victoire Noizet)
 - In-browser debugger (Victoire Noizet)
 - Certified interpreter (extracted from Coq) (Guillaume Ambal)
 - Skel to Skel transformation
 - big-step to small-step (Guillaume Ambal)
 - big-step to abstract machines (Martin Andrieux)
 - Generation of abstract analyzers (Vincent Rébiscoul)
 - Hoare Logic (Laura-Andrea Schimbător)
 - Abstract machines for process calculi (Sergueï Lenglet)

Future Work

- More languages
 - Rust
 - Esterel
- Skel improvements
 - include support
 - type inference
- New backends
 - generic compilation
 - symbolic execution

Questions?

Many thanks to Guillaume Ambal, Martin Andrieux, Martin Bodin, Nathanaëlle Courant, Enzo Crance, Philippa Gardner, Olivier Idir, Thomas Jensen, Adam Khayam, Sergueï Lenglet, Victoire Noizet, Vincent Rébiscoul, Thomas Rubiano

https://skeletons.inria.fr

Extra Slides

Pattern Matching

```
val eval_expr ((st:state), (e:expr)) : value =
match e with

| Const n -> let i = int_of_lit n in Int n

| Plus(e1,e2) ->
    let Int n1 = eval_expr (st, e1) in
    let Int n2 = eval_expr (st, e2) in
    let n = add (n1, n2) in
    Int i
end
```

Choice

```
val insert<a> ((e:a), (l: list<a>)): list <a> =
  branch
    Cons < a > (e, 1)
 or
    let Cons(e', l') = l in
    let l'' = insert < a > (e, l') in
    Cons<a>(e', 1'')
  end
val permut<a> (1: list<a>): list <a> =
 match 1 with
  | Nil → Nil<a>
  | Cons(e,es) → let es' = permut<a> es in insert<a>(e, es')
  end
```

Skel, Formally

```
TERM t ::= x \mid C \mid t \mid (t, ..., t) \mid \lambda p : \tau \cdot S

PATTERN p ::= x \mid \_ \mid C \mid p \mid (p, ..., p)

SKELETON S ::= t \mid t \mid \text{let } p = S \text{ in } S \mid \text{let } p : \tau \text{ in } S

\mid \oplus (S..S) \mid \mathcal{M}(t) (p \rightarrow S..p \rightarrow S) \mid t

TYPE SPEC r ::= \text{type } b \mid \text{type } b := \tau \mid \text{type } b = "\mid "C\tau ..."\mid "C\tau

TERM SPEC r' ::= \text{val } x : \tau \mid \text{val } x : \tau = t
```

Note: this is almost in administrative normal form

Existentials

$$\frac{\Gamma, x : \tau \vdash m : \nu}{\Gamma \vdash \lambda x \cdot m : \tau \to \nu}$$

Existentials

$$\frac{\Gamma, x : \tau \vdash m : \nu}{\Gamma \vdash \lambda x \cdot m : \tau \to \nu}$$

Predicate

```
input: whole judgement, output: unit
val ctype ((gamma: env), (t: term), (tp: ltype)) : () =
  branch
  let Lam (x, m) = t in
  let Arrow (tau,nu) = tp in
  let gamma' = ext_env (gamma, x, tau) in
  ctype (gamma', m, nu)
```

Existentials

$$\frac{\Gamma, x : \tau \vdash m : \nu}{\Gamma \vdash \lambda x \cdot m : \tau \to \nu}$$

Algorithmic

```
input: typing env and term, output: type
val ctype ((gamma: env), (t: term)) : ltype =
branch
   let Lam (x, m) = t in
   let tau : ltype in
   let gamma' = ext_env (gamma, x, tau) in
   let nu = ctype (gamma', m) in
   Arrow (tau, nu)
```