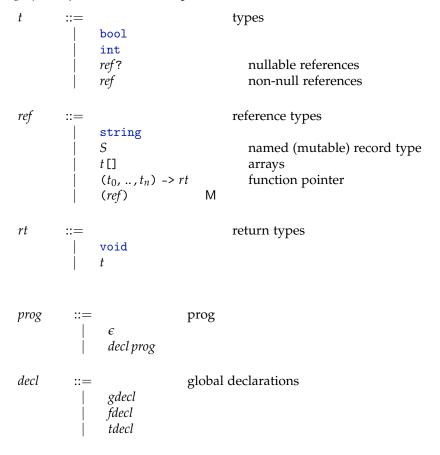
Oat v.2 Language Specification

252-0210-00L Compiler Design HS2021

November 21, 2021

1 Grammar

The following grammar defines the Oat syntax. All binary operations are *left associative* with precedence levels indicated numerically. Higher precedence operators bind tighter than lower precedence ones. Here *id* ranges over lower-case identifiers and *S* indicates an upper-case "struct name." The parts of the grammar marked with M are not part of the abstract syntax, but are included for parsing purposes (e.g. parentheses) or for use in explaining the typing judgments. Note that left-hand-sides (*lhs*) and global expressions (*gexp*) are just a subsets of expressions.



```
gdecl
                                                       global value declarations
                  global id = gexp;
arg
                                                       arg
                  t id
args
            ::=
                                                       args
                  arg_1, \dots, arg_n
fdecl
                                                       function declaration
            ::=
                  rt id(args) block
field
                                                       field declarations
                  t x
fields
                                                       fields
            ::=
                  field_1; ..; field_n
                  fields; t_{n+1} x_{n+1}; ...; t_m x_m
tdecl
                                                       struct declaration
                  struct S{ fields }
                                                          expressions
exp
                  ref null
                                                              type-annotated null value
                  boolean
                                                              boolean literals true or false
                  integer
                                                              64-bit integer literals
                                                              C-style strings
                  string
                                                              global or local variable
                  id
                  \operatorname{new} t[]\{exp_1, ..., exp_n\}
                                                              array literal value
                  new t [exp_1] \{id = > exp_2\}
                                                              array with "initializer"
                  exp_1[exp_2]
                                                              array index
                  length (exp)
                                                              built-in polymorphic length operation
                  \mathtt{new}\,S\{x_1 = exp_1\,;\,..\,;x_n = exp_n\}
                                                              struct initialization
                  exp.id
                                                              field projection
                  exp(exp_1, .., exp_n)
                                                              function call
                                                              unary operation
                  иор ехр
                  exp_1 bop exp_2
                                                              binary operation
                  (exp)
                                                     M
                                                     Μ
lhs
                                                          left-hand-sides for assignment
                  id
                  exp_1[exp_2]
                  exp.id
                                                          global expressions
gexp
                  ref null
                  boolean
```

```
integer
                     string
                     \operatorname{new} t[]\{exp_1, ..., exp_n\}
                     \mathbf{new} S\{x_1 = gexp_1; ...; x_n = gexp_n\}
                                                                      blocks
block
               ::=
                      \{stmt_1 .. stmt_n\}
                                                                      local variable declarations
vdecl
                      var id = exp
stmt
               ::=
                                                                      statements
                                                                         assignment statement
                      lhs = exp;
                      vdecl;
                                                                         variable declaration
                      return exp;
                                                                         return with value
                      return;
                                                                         return (void)
                                                                         call a void-returning function
                      exp(exp_1, .., exp_n);
                      if\_stmt
                                                                         conditionals
                      for(vdecls; exp_opt; stmt_opt) block
                      while(exp) block
if_stmt
               ::=
                                                                      if statements
                      if(exp) block else_stmt
                                                                         standard boolean if
                      if?(ref id = exp) block else_stmt
                                                                         possibly-null checked downcast
else_stmt
                                                                      else
                      \verb"else" block"
                      else if_stmt
```

```
bop
                        (left associative) binary operations
         ::=
                           precedence 90
                           precedence 90
                           precedence 100
                           precedence 60
                           precedence 60
                           precedence 70
                           precedence 70
               >
                           precedence 70
                           precedence 70
                           precedence 50 precedence 40
               &
               precedence 30 precedence 20
               [&]
               [1]
                           precedence 80
               <<
                           precedence 80
               >>
                           precedence 80
               >>>
                        unary operations
иор
               ! ~
```

Subtyping Rules

$$H \vdash t_1 \leq t_2$$

$$\overline{H \vdash \text{int}} \leq \text{int}$$

$$\overline{H \vdash \text{int}} \leq \text{int}$$

$$\overline{H \vdash \text{bool}} \leq \text{bool}$$

$$SUB_SUB_BOOL$$

$$\frac{H \vdash_r ref_1 \leq ref_2}{H \vdash ref_1?} \leq ref_2?$$

$$\frac{H \vdash_r ref_1 \leq ref_2}{H \vdash ref_1 \leq ref_2} = \text{SUB_SUB_REF}$$

$$\frac{H \vdash_r ref_1 \leq ref_2}{H \vdash ref_1 \leq ref_2?} = \text{SUB_SUB_NRREF}$$

$$\overline{H \vdash_r ref_1 \leq ref_2?} = \text{SUB_SUB_NRREF}$$

$$\overline{H \vdash_r ref_1 \leq ref_2?} = \text{SUB_SUB_NRREF}$$

$$\overline{H \vdash_r t \cap \subseteq t \cap \subseteq t \cap \subseteq t} = \text{SUB_SUBRARRAY}$$

$$\overline{H \vdash_r t \cap \subseteq t \cap \subseteq t} = \text{SUB_SUBRARRAY}$$

$$\overline{H \vdash_r t \cap \subseteq t \cap \subseteq t} = \text{SUB_SUBRARRAY}$$

$$\overline{H \vdash_r t \cap \subseteq t \cap \subseteq t} = \text{SUB_SUBRARRAY}$$

$$\overline{H \vdash_r t \cap \subseteq t} = \text{SUB_SUB_SUBRARRAY}$$

$$\overline{H \vdash_r t \cap \subseteq t} = \text{SUB_SUB_SUBRARRAY}$$

$$\overline{H \vdash_r t \cap \subseteq t} = \text{SUB_SUB_SUB_SUBRARRAY}$$

$$\overline{H \vdash_r t \cap \subseteq t} = \text{SUB_SUB_SUB_SUB_SUB_SUB_SUB_$$

 $H \vdash_{rt} rt_1 \leq rt_2$

 $H \vdash_r ref_1 \leq ref_2$

$$\frac{H \vdash_{rt} \texttt{void} \leq \texttt{void}}{H \vdash_{t_1} \leq t_2} \quad \texttt{SUB_SUBRETSVOID}$$

$$\frac{H \vdash t_1 \leq t_2}{H \vdash_{rt} t_1 \leq t_2} \quad \texttt{SUB_SUBRETRTTYP}$$

3 Well-formed types

 $H \vdash t$

 $H \vdash_r ref$

 $H \vdash_{rt} rt$

$$\frac{H \vdash_{\mathit{rt}} \mathtt{void}}{H \vdash_{\mathit{rt}} t} \quad \text{wf_rtypokvoidok}$$

4 Typing Rules

```
\vdash bop_1, \dots, bop_i \colon\! t
                           \frac{}{\vdash <, <=, >, >=: (int, int) \rightarrow bool} \quad TYP\_CMPOPS

  ⊢ &, |: (bool, bool) -> bool
  TYP_BOOLOPS

 \vdash uop:t
                                                                        TYP_LOGNOT
                                            ⊢ !:(bool) -> bool
                                                                        TYP_BITNEG
                                             ⊢~ :(int) -> int
                                                                        TYP_NEG
                                                ⊢ -: (int) -> int
 H;G;L\vdash exp:t
                                           \frac{H \vdash ref}{H; G; L \vdash ref \text{ null } : ref?}
                                                                             TYP_NULL
                                                                              TYP_BOOL
                                           H;G;L \vdash boolean : bool
                                                                              TYP_INT
                                             \overline{H;G;L\vdash integer: int}
                                                                             TYP_STRING
                                         \overline{H;G;L\vdash string : string}
                                               \frac{\mathit{id} \colon\! t \in L}{H; G; L \vdash \mathit{id} \: \colon \: t} \quad \texttt{TYP\_LOCAL}
                                              \frac{id \notin L \quad id: t \in G}{H; G; L \vdash id: t} \quad \text{TYP\_GLOBAL}
                              H \vdash t
                              H;G;L \vdash exp_1 : t_1 ... H;G;L \vdash exp_n : t_n
                              H \vdash t_1 \leq t .. H \vdash t_n \leq t
                                  H;G;L \vdash \text{new } t[]\{exp_1, ..., exp_n\} : t[] TYP_CARR
                         H \vdash t
                         H;G;L\vdash exp_1: int
                         x \notin L H; G; L, x: \text{int} \vdash exp_2 : t' \quad H \vdash t' \leq t TYP_NEWARRAY
                              H;G;L \vdash \text{new } t[exp_1]\{x=>exp_2\} : t[]
                              H;G;L \vdash exp_1 : t[] \quad H;G;L \vdash exp_2 : int TYP_INDEX
                                          H;G;L\vdash exp_1[exp_2]:t
                                             H;G;L\vdash exp:t[]
                                      \overline{H;G;L\vdash length(exp): int} TYP_LENGTH
```

```
struct S\{t_1 x_1; ...; t_n x_n\} \in H
H;G;L\vdash exp_1:t_1' .. H;G;L\vdash exp_n:t_n'
H \vdash t_1' \leq t_1 \quad .. \quad H \vdash t_n' \leq t_n
fields may be permuted under new
 H;G;L \vdash \text{new } S\{x_1=exp_1; ...; x_n=exp_n\} : S TYP_STRUCTEX
        H;G;L\vdash exp:S
        struct S\{ \text{ fields } \} \in H \text{ } t \text{ } x \in \text{ fields} 
TYP_FIELD
                  H;G;L\vdash exp.x:t
   H;G;L \vdash exp : (t_1,..,t_n) \rightarrow t
   H;G;L \vdash exp_1 : t'_1 ... H;G;L \vdash exp_n : t'_n
   H \vdash t_1' \leq t_1 \quad .. \quad H \vdash t_n' \leq t_n
                                                     TYP_CALL
           H;G;L \vdash exp(exp_1,..,exp_n) : t
      \vdash bop: (t_1, t_2) \rightarrow t
      H;G;L\vdash exp_1:t_1 \quad H;G;L\vdash exp_2:t_2

TYP_BOP
               H;G;L \vdash exp_1 bop exp_2 : t
       H;G;L \vdash exp_1 : t_1 \quad H;G;L \vdash exp_2 : t_2
      H \vdash t_1 \leq t_2 \quad H \vdash t_2 \leq t_1
                                                               TYP_EQ
              H;G;L \vdash exp_1 == exp_2 : bool
      H; G; L \vdash exp_1 : t_1 \quad H; G; L \vdash exp_2 : t_2
      H \vdash t_1 \leq t_2 \quad H \vdash t_2 \leq t_1
                                                       TYP_NEQ
             H;G;L \vdash exp_1 != exp_2 : bool
          \vdash uop:(t) \xrightarrow{-> t} H; G; L \vdash exp:t
TYP_UOP
                  H;G;L\vdash uopexp:t
```

```
H;G;L_1 \vdash vdecl \Rightarrow L_2
                                                     \frac{H;G;L \vdash exp : t \quad x \not\in L}{H;G;L \vdash var \ x = exp \Rightarrow L,x:t} \quad \texttt{TYP\_DECL}
H;G;L_0 \vdash vdecls \Rightarrow L_i
                             \frac{H;G;L_0 \vdash vdecl_1 \Rightarrow L_1 \quad \dots \quad H;G;L_{n-1} \vdash vdecl_i \Rightarrow L_n}{H;G;L_0 \vdash vdecl_1, \dots, vdecl_n \Rightarrow L_n} \quad \text{TYP\_VDECLS}
H;G;L_1;rt \vdash stmt \Rightarrow L_2;returns
                                              lhs: t \in L or lhs not a global function id
                                              H;G;L\vdash lhs:t
                                              H;G;L\vdash exp:t'
                                              H \vdash t' \leq t
                                                    H; G; L; rt \vdash lhs = exp; \Rightarrow L; \bot TYP_ASSN
                                                 \frac{H;G;L_1 \vdash vdecl \Rightarrow L_2}{H;G;L_1;rt \vdash vdecl; \Rightarrow L_2;\bot} \quad \texttt{TYP\_STMTDECL}
                                        H;G;L \vdash exp : (t_1,..,t_n) \rightarrow void
                                        H;G;L\vdash exp_1:t_1' .. H;G;L\vdash exp_n:t_n'
                                       \frac{H \vdash t_1' \leq t_1 \quad .. \quad H \vdash t_n' \leq t_n}{H; G; L; rt \vdash exp(exp_1, ..., exp_n); \Rightarrow L; \bot} \quad \text{TYP\_SCALL}
                                                                H;G;L \vdash exp : bool
                                                                H;G;L;rt \vdash block_1;r_1
                                                                H;G;L;rt \vdash block_2;r_2
                                     H; \overline{G; L; rt} \vdash \text{if}(exp) \ block_1 \ \text{else} \ block_2 \ \Rightarrow \ L; r_1 \wedge r_2
                                     H;G;L \vdash exp : ref'?
                                     H \vdash ref' \leq ref
                            \frac{H;G;L,x:ref;rt\vdash block_1;r_1\quad H;G;L;rt\vdash block_2;r_2}{H;G;L;rt\vdash \  \, \texttt{if?}(ref\ x=exp)\  \, block_1\  \, \texttt{else}\  \, block_2\  \, \Rightarrow\  \, L;r_1\wedge r_2}
                                                            H;G;L \vdash exp : bool
                                                            H;G;L;rt \vdash block;r
                                            \overline{H;G;L;rt} \vdash \text{while}(exp) \ block \Rightarrow L;\bot TYP_WHILE
                                                         H;G;L_1 \vdash vdecls \Rightarrow L_2
                                                         H;G;L_2 \vdash exp : bool
                                                         H;G;L_2;rt \vdash stmt \Rightarrow L_3;\bot
                                                         H;G;L<sub>2</sub>;rt \vdash block;r
                                                                                                                                             TYP_FOR
                          \overline{H;G;L_1;rt\vdash for(vdecls; exp\_opt; stmt\_opt) block \Rightarrow L_1;\bot}
                                                   \frac{H;G;L\vdash exp:t'\quad H\vdash t'\leq t}{H;G;L;t\vdash \texttt{return}\,exp;\;\Rightarrow L;\top}
                                                                                                             TYP_RETVOID
                                                \overline{H;G;L;void}\vdash return; \Rightarrow L;\top
```

H;G;L; $rt \vdash block$;returns

$$\frac{H;G;L_0;rt \vdash_{ss} stmt_1..stmt_n \Rightarrow L_n;r}{H;G;L_0;rt \vdash \{stmt_1..stmt_n\};r} \quad \text{TYP_BLOCK}$$

H; G; L_0 ; $rt \vdash_{ss} stmt_1 ... stmt_n \Rightarrow L_n$; returns

$$\begin{array}{c} H;G;L_{0};rt \vdash stmt_{1} \Rightarrow L_{1};\bot\\ \dots\\ H;G;L_{n-2};rt \vdash stmt_{n-1} \Rightarrow L_{n-1};\bot\\ H;G;L_{n-1};rt \vdash stmt_{n} \Rightarrow L_{n};r\\ \hline H;G;L_{0};rt \vdash_{ss} stmt_{1}..stmt_{n-1} stmt_{n} \Rightarrow L_{n};r \end{array} \ \ \ {\tt TYP_STMTS}$$

H; $G \vdash_s tdecl$

$$\frac{H \vdash t_1 \quad .. \quad H \vdash t_i \quad x_1 ... x_i \, \mathbf{distinct}}{H; G \vdash_{S} \, \mathbf{struct} \, S\{ \, t_1 \, x_1 \, ; \, .. \, ; t_i \, x_i \, \}} \quad \mathsf{TYP_TDECLOK}$$

H; $G \vdash_f fdecl$

$$\frac{H;G;x_1:t_1,...,x_i:t_i;rt\vdash block;\top x_1..x_i \mathbf{distinct}}{H;G\vdash_f rtf(t_1x_1,...,t_ix_i) \ block}$$
 TYP_FDECLOK

H; $G \vdash prog$

H; $G_1 \vdash_g prog \Rightarrow G_2$

$$\begin{array}{ccc} \overline{H; G \vdash_{g} \epsilon \Rightarrow G} & \text{TYP_GEMPTY} \\ \\ \underline{H; G_{1} \vdash_{g} prog \Rightarrow G_{2}} \\ \overline{H; G_{1} \vdash_{g} tdecl prog \Rightarrow G_{2}} & \text{TYP_GTDECL} \end{array}$$

$$H; G_1; \vdash gexp : t \quad gexp \ contains \ no \ global \ variables$$

$$x \notin G_1 \quad H; G_1, x : t \vdash_g prog \Rightarrow G_2$$

$$H; G_1 \vdash_g \text{global } x = gexp; prog \Rightarrow G_2$$

$$TYP_GGDECL$$

$$\frac{H; G_1 \vdash_g prog \Rightarrow G_2}{H; G_1 \vdash_g fdecl prog \Rightarrow G_2} \quad \texttt{TYP_GFDECL}$$

 $H \vdash fdecl \Rightarrow id:t$

$$\frac{H \vdash_{rt} rt \quad H \vdash t_1 \quad .. \quad H \vdash t_n}{H \vdash rt f(t_1 x_1, .., t_n x_n) \ block \Rightarrow f \colon (t_1, .., t_n) \rightarrow rt} \quad \text{TYP_FTYP}$$

$$H; G_1 \vdash_f prog \Rightarrow G_2$$

$$\begin{array}{c} \overline{H;G\vdash_f \epsilon\Rightarrow G} & \text{TYP_FEMPTY} \\ \\ \overline{H;G_1\vdash_f prog\Rightarrow G_2} \\ \overline{H;G_1\vdash_f tdecl\,prog\Rightarrow G_2} & \text{TYP_FTDECL} \\ \\ \overline{H;G_1\vdash_f prog\Rightarrow G_2} \\ \overline{H;G_1\vdash_f gdecl\,prog\Rightarrow G_2} & \text{TYP_FGDECL} \\ \\ H\vdash fdecl\Rightarrow f:t \\ f\not\in G_1 & H;G_1,f:t\vdash_f prog\Rightarrow G_2 \\ \hline H;G_1\vdash_f fdecl\,prog\Rightarrow G_2 & \text{TYP_FFDECL} \\ \end{array}$$

 $H_1 \vdash_s prog \Rightarrow H_2$

$$\overline{H \vdash_s \epsilon \Rightarrow H} \quad \text{TYP_SEMPTY}$$

$$\underline{S \not\in H_1 \quad H_1, \text{struct } S\{ \text{ fields } \} \quad \vdash_s prog \Rightarrow H_2} \quad \text{TYP_STDECL}$$

$$\underline{H_1 \vdash_s prog \Rightarrow H_2} \quad \text{TYP_SGDECL}$$

$$\underline{H_1 \vdash_s prog \Rightarrow H_2} \quad \text{TYP_SGDECL}$$

$$\underline{H_1 \vdash_s prog \Rightarrow H_2} \quad \text{TYP_SFDECL}$$

$$\underline{H_1 \vdash_s prog \Rightarrow H_2} \quad \text{TYP_SFDECL}$$

⊢ prog

$$\frac{\cdot \vdash_{s} prog \Rightarrow H \quad H; G_{0} \vdash_{f} prog \Rightarrow G_{1} \quad H; G_{1} \vdash_{g} prog \Rightarrow G_{2} \quad H; G_{2} \vdash prog}{\vdash prog} \quad \text{TYP_PROG}$$

Notes:

- The context G_0 mentioned in the rule for typechecking a complete, top-level program is the "initial context" which should contain bindings for all of the OAT built-in functions.
- The type system processes the program in several passes: (1) collect up all the structure type definitions and make sure their names don't clash using the \vdash_s rules, (2) add all the function identifiers and their types to the global context using the \vdash_f rules, again ensuring no name clashes, (3) type-check the global value declarations and add them to the context using the \vdash_g rules, and (4) process all the declarations one more time to examine all the struct fields to make sure their types are well formed and to typecheck the function bodies.

The rules therefore allow the types of structs to be mutually recursive, and for global values to mention function pointers as constants.

• We use ⊥ to indicate that a statement might not return, and ⊤ to indicate that a statement definitely returns. When typechecking the list of statements that make up a block, only the last statement is allowed to definitely return (all the others must possibly not return). Note that TYP_FDECLOK requires that the block making up a function body definitely return.