

# Pure Garbage: The Compost Language Reference Manual

October 18, 2023

```
val name_email_map : (string * string) list =
[("Roger Burtonpatel", "roger.burtonpatel@tufts.edu");
 ("Randy Dang", "randy.dang@tufts.edu");
 ("Jasper Geer", "jasper.geer@tufts.edu");
 ("Jackson Warhover", "jackson.warhover@tufts.edu")]
```

# 1 Introduction

Compost is a statically-typed pure functional programming language with automatic compile-time memory management enabled by its uniqueness type system.

## 2 Notational Conventions

Code listings will appear in “verbatim” as follows:

```
(define foo ()
  bar)
```

Grammar rules are written in extended Backus-Naur format, as follows:

$$\begin{array}{l} rule \\ \quad | \end{array} ::= \begin{array}{l} (nonterminal \text{ terminal}) \\ \{ other-rule \} \end{array}$$

Parentheses are concrete syntax, but any pair of balanced parentheses may be freely exchanged for a pair of square brackets. For example, the following two declarations are indistinguishable in the abstract syntax:

```
(val x 1)
```

```
[val x 1]
```

Note that braces are used in a manner akin to the Kleene closure, that is, a term enclosed in brackets may be omitted or arbitrarily repeated.

## 3 Lexical Conventions

### 3.1 Whitespace

The following characters are considered as whitespace and, with one exception, ignored during tokenization: spaces, tabs, carriage returns, and newlines.

### 3.2 Comments

Comments are introduced by the character `;` and terminated by the newline character. Comments are treated as whitespace.

### 3.3 Literals

*literal* ::= *integer-literal*  
          | *symbol-literal*  
          | *boolean-literal*  
          | *unit-literal*

Literals introduce values of Compost’s primitive types. All literals are valid expressions.

### 3.4 Integer Literals

*integer-literal* ::= token composed only of digits, possibly prefixed with a `+` or `-`.

The `+` prefix denotes a positive integer and the `-` prefix denotes a negative integer. The characters `1 2 3 4 5 6 7 8 9 0` are considered digits.

### 3.5 Symbol Literals

*symbol-literal* ::= `'{ symbol-character }'`

*symbol-character* ::= any unicode code point other than `'` and the backslash character unless escaped with a backslash.

That is, any sequence of `'`-delimited unicode characters is a valid symbol literal, as long as every instance of `'` or backslash are preceded by a backslash. These escape sequences are replaced by their unescaped counterparts in the introduced symbol value. For example, the following are valid symbol literals:

```
'\'hello, world\''
```

```
'\\ is a backslash'
```

```
'I exist  
on multiple lines'
```

### 3.6 Other Literals

*boolean-literal* ::= `true` | `false`

*unit-literal* ::= `unit`

### 3.7 Reserved Words

The following tokens are considered reserved:

```
; ( ) [ ] : _ -> if val define datatype use case begin let dup int bool sym unit
```

## 4 Values

This section describes the kinds of values manipulated by Compost programs.

### 4.1 Integers

Integer values are 32-bit signed integers with a range of -2,147,483,648 to 2,147,483,647.

### 4.2 Symbols

Symbol values are interned immutable strings of unicode characters.

### 4.3 Booleans

Boolean values are either the boolean *true* or the boolean *false*.

### 4.4 Unit

Unit values are the value *unit*.

### 4.5 Variant Values

Variant values are either a constant constructor or a non-constant constructor applied to a series of value arguments. We write an arbitrary constant constructor *c* as *(c)* and an arbitrary non-constant constructor *d* applied to arguments  $v_1 \dots v_n$  as *(d v<sub>1</sub> ... v<sub>n</sub>)*.

Variant constructors are monomorphic, that is, for any constructor *c*, there exist types  $\tau_1 \dots \tau_n$  such that for any application of constructor *c* to arguments  $v_1 \dots v_n$ ,  $v_i$  must have type  $\tau_i$  for all  $i \in 1, 2, \dots, n$ .

### 4.6 Functions

Functions in Compost are first-class objects. Function values are mappings from ordered sets of values, to values. That is, a function *f*, when applied to values  $v_1 \dots v_n$ , produces a value  $v_r$ . Like variant constructors, functions are monomorphic, so the types of  $v_1 \dots v_n$  and  $v_r$  are fixed.

## 5 Names

Compost places relatively liberal constraints on the sequences of characters considered valid names.

*name* ::= any token that is not an *int-lit*, does not contain whitespace, a ' , bracket, or parenthesis, and is not a reserved word.

Names are bound to types, values, and variant constructors, and used to refer to them at various points in a program. Names are also used in the *use-declaration* syntactic form to refer to files.

## 6 Type Expressions

*type-expression* ::= *function-type*  
| *int-type*  
| *bool-type*  
| *sym-type*  
| *unit-type*  
| *datatype*

### 6.1 Primitive Types

*int-type* ::= **int**

*bool-type* ::= **bool**

*sym-type* ::= **sym**

*unit-type* ::= **unit**

**int** is the type of integer values.

**bool** is the type of boolean values.

**sym** is the type of symbol values.

**unit** is the type of unit values.

### 6.2 Function Types

*function-type* ::=  $(\rightarrow (\{ \textit{type} \}) \textit{type})$

$(\rightarrow (\textbf{t1} \dots \textbf{tn}) \textbf{tr})$  is the type of function values which map ordered sets of values  $v_1 \dots v_n$  of types  $\textbf{t1} \dots \textbf{tn}$  to values  $v_r$  of type  $\textbf{tr}$ .

## 6.3 Datatypes

*datatype* ::= *name*

Datatypes are the types of variant constructor values. Multiple variant constructors may share the same type. Datatypes and their constructors can be defined by the programmer with the following syntax:

*datatype-definition* ::= (**datatype** *name* ({ *variant-constructor-definition* } ))

*variant-constructor-definition* ::= (*name* ({ *type-expression* } ))

A *name* bound to the new type  $\tau_d$  appears directly following the **datatype** keyword, and this is followed by a list of variant constructor definitions. Each of these provides a *name* bound to the constructor, *c*, followed by a list of *type-expressions*  $\tau_1 \dots \tau_n$  typing its arguments. Given this definition, a variant value ( $c \ v_1 \ \dots \ v_n$ ) of type  $\tau_d$  may be introduced by applying function value *c* to  $v_1 \dots v_n$ , where the type of  $v_i$  is  $\tau_i$  for all  $i \in 1, 2, \dots, n$

The placement of a datatype or variant constructor's definition has no bearing on where it can be referenced, introduced, or eliminated. In fact, datatypes may be defined recursively, as in the following example:

```
(datatype int-list
  (
    [cons (int int-list)]
    [nil ()]
  )
)
```

This declaration can be read as: “an **int-list** is either **cons** applied to an **int** and an **int-list**, or **nil** applied to nothing”.

## 7 Expressions

*expr* ::= *literal*  
          | *case-expression*  
          | *if-expression*  
          | *begin-expression*  
          | *apply-expression*  
          | *let-expression*  
          | *dup-expression*  
          | *name-expression*

Meaningful computation is encoded in Compost as *expr* syntactic forms, or expressions. These appear either as the right-hand side of **val** declarations or as the bodies of functions.

We describe the semantics and typing rules of expressions largely informally but use formal notation to aid conciseness. Expressions are evaluated in an environment  $\rho$  mapping names to values. Initially, these environments contain the values and types of all globally bound names (functions, **val**-bound names).  $\rho[x \mapsto v]$  is the modified environment  $\rho$  in which name  $x$  is bound to value  $v$ .  $\rho[x]$  is the value mapped to by  $x$  in  $\rho$ .

There also exists a typing environment  $\Gamma$  mapping names to types. The same syntax is used to add bindings to  $\Gamma$  and denote the type mapped to by a name  $x$ . We also introduce a typing judgement  $\Gamma \vdash e : \tau$  which can be read “expression  $e$  has type  $\tau$  in context  $\Gamma$ ”. When  $\Gamma$  is used in a subsection, it refers to the environment in which that particular expression is typed, rather than the initial typing environment. This typing judgement is defined inductively on the structure of expressions by the following subsections.

Certain expressions will “consume” names, though note that only names mapped to variant values are consumed. Any names considered as consumed in a subexpression are considered consumed in the parent expression. Consumption is defined inductively on the structure of expressions by the following subsections.

Side effects are produced in evaluation order except in the case of **val**-bound names, which produce their associated expression’s side effects at **every** reference.

## 7.1 Case Expressions

*case-expression* ::= (**case** *expr* (**{** *case-branch* **}**))

*case-branch* ::= (*pattern* *expr*)

*pattern* ::= (*name* **{** *name* **|** **\_** **}**)  
| **\_**

Values of the form  $(c \ v_1 \ .. \ v_n)$  are eliminated by the *case-expression* syntactic form. Consider a case expression with  $n$  branches of the form:

```
(case e
  (
    [(c1 v11 v12 ...) e1]
    ...
    [(cn vn1 vn2 ...) en]
  )
)
```

### 7.1.1 Typing

We assert that the type of **e** must be a datatype. Suppose that  $\Gamma \vdash \mathbf{e} : \tau_d$ . Each  $c_i$  must be a variant constructor of  $\tau_d$ . For all  $i \in 1, 2, \dots, n$ , let  $\tau_{i1}, \tau_{i2}, \dots, \tau_{im}$  be the types of  $c_i$ ’s  $m$  arguments.

We assert that the number of names and wildcards following  $\mathbf{ci}$  must be precisely  $m$ , and that all must be fresh. Let  $\Gamma_i$  be  $\Gamma[\mathbf{vi}1 \mapsto \tau_{i1}, \dots, \mathbf{vim} \mapsto \tau_{im}]$ . Note that wildcards are not bound. We assert that  $\Gamma_i \vdash \mathbf{ei} : \tau_r$ . The type of this expression in context  $\Gamma$  is  $\tau_r$ .

### 7.1.2 Consumption

Names marked as consumed in  $\mathbf{e}$  are marked as consumed in all  $\mathbf{ei}$ .

### 7.1.3 Evaluation

Suppose evaluation of  $\mathbf{e}$  in environment  $\rho$  yields a value  $v = (c \ v_i \ \dots \ v_m)$ . If there exists some branch whose pattern is prefixed by  $c$ , it is evaluated in the environment  $\rho$  and its result is returned.

Suppose this branch is the **case-branch** containing the pattern prefixed by variant constructor  $\mathbf{ck}$ . Evaluation of this branch yields the result of evaluating  $\mathbf{ek}$  in the modified environment  $\rho' = \rho[\mathbf{vk}1 \mapsto v_1, \dots, \mathbf{vkm} \mapsto v_m]$ . Note that we do not bind wildcards in  $\rho'$ .

## 7.2 If Expressions

*if-expression* ::= (**if** *expr expr expr*)

Consider an if expression of the form:

(**if**  $\mathbf{e1} \ \mathbf{e2} \ \mathbf{e3}$ )

### 7.2.1 Typing

We assert that the  $\Gamma \vdash \mathbf{e1} : \mathbf{bool}$ . We further assert that  $\Gamma \vdash \mathbf{e2} : \tau_r$  and  $\Gamma \vdash \mathbf{e3} : \tau_r$ . The type of this expression in context  $\Gamma$  is  $\tau_r$ .

### 7.2.2 Consumption

Names marked as consumed in  $\mathbf{e1}$  are marked as consumed in  $\mathbf{e1}$  and  $\mathbf{e2}$ .

### 7.2.3 Evaluation

Suppose evaluation of  $\mathbf{e1}$  in environments  $\rho$  yields a boolean value  $v$ . If  $v$  is the value **true**, the expression  $\mathbf{e1}$  is evaluated in environment  $\rho$  and its result is returned. If  $v$  is the value **false**, the expression  $\mathbf{e2}$  is evaluated in environment  $\rho$  and its result is returned.

## 7.3 Begin Expressions

*begin-expression* ::= (**begin** { *expr* })

Consider a begin expression of the form:

(**begin**  $\mathbf{e1} \ \dots \ \mathbf{en}$ )

### 7.3.1 Typing

The type of this expression is the type of **en**.

### 7.3.2 Consumption

For all  $i \in 1, 2, \dots, n - 1$ , names marked as consumed in **ei** are marked as consumed in **ei** + 1.

### 7.3.3 Evaluation

Each **ei** is evaluated in environment  $\rho$  in order from 1... $n$ . We return the result of evaluating **en** in environment  $\rho$ .

## 7.4 Apply Expressions

*apply-expression* ::= (*expr* { *expr* })

Consider an apply expression of the form:

(**e** **e1** ... **en**)

### 7.4.1 Typing

We assert that  $\Gamma \vdash \mathbf{e} : (-> (t_1 \dots t_n) t_r)$ . Each **ei** must be of type  $t_i$  for  $i \in 1, 2, \dots, n$ . The type of this expression is  $t_r$ .

### 7.4.2 Consumption

For all  $i \in 1, 2, \dots, n - 1$ , names marked as consumed in **ei** are marked as consumed in **ei** + 1.

### 7.4.3 Evaluation

Each **ei** is evaluated in environment  $\rho$  in order from 1... $n$ . Let  $v_1 \dots v_n$  be the values returned by evaluating each **ei**.

We return the result of applying  $v_f$  to arguments  $v_1 \dots v_n$ .

## 7.5 Let Expressions

*let-expression* ::= (**let** ({ *let-binding* }) *expr*)

*let-binding* ::= (*name* *expr*)

Consider a let expression of the form:

```
(let
  (
    [x1 e1]
    ...
```



```

    [xn en]
  )
  e
)
```

### 7.5.1 Typing

We assert that each  $\mathbf{x}i$  for  $i \in 1, 2, \dots, n$  be fresh. Given that  $\Gamma_k : \mathbf{x}k : \tau_k$ , for  $k \in 1, 2, \dots, n-1$ , we say that  $\Gamma_{k+1} = \Gamma_k[\mathbf{x}k \mapsto \tau_k]$ . As a base case, let  $\Gamma_1 = \Gamma$ . The type of this expression is the type of  $\mathbf{e}n$  in context  $\Gamma_n$ .

### 7.5.2 Consumption

For any  $i \in 1, 2, \dots, n$  we mark any names consumed in  $\mathbf{e}i$  as consumed in both  $\mathbf{e}$  and all  $\mathbf{e}k$  for  $k > i$ .

### 7.5.3 Evaluation

For all  $i \in 1, 2, \dots, n$ , let  $\rho_{i+1} = \rho_i[\mathbf{x}i \mapsto v_i]$ , where  $v_i$  is the value returned by evaluating  $\mathbf{e}i$  in environment  $\rho_i$ . As a base case, let  $\rho_1 = \rho$ . We return the result of evaluation  $\mathbf{e}$  in environment  $\rho_{n+1}$ .

## 7.6 Name Expressions

*name-expression* ::= *name*

Consider a name expression of the form:

$\mathbf{n}$

### 7.6.1 Typing

We assert that  $\mathbf{n}$  be bound in  $\Gamma$ . We further assert that  $\mathbf{n}$  not be marked as consumed. The type of this expression is  $\Gamma[\mathbf{n}]$

### 7.6.2 Consumption

$\mathbf{n}$  is marked as consumed.

### 7.6.3 Evaluation

We return the value  $\rho[\mathbf{n}]$ .

## 7.7 Dup Expressions

*dup-expression* ::= (*dup name*)

Consider a dup expression of the form:

(*dup n*)

### 7.7.1 Typing

We assert that  $n$  be bound in  $\Gamma$  and that  $\Gamma[n]$  be a datatype. We further assert that  $n$  not be marked as consumed. The type of this expression is  $\Gamma[n]$

### 7.7.2 Consumption

$n$  is **not** marked as consumed.

### 7.7.3 Evaluation

We return the value  $\rho[n]$ .

## 8 Definitions

Syntactic forms in the *def* category are allowed only at the top level of a Compost program.

<i>def</i>	$::=$	<i>val-binding</i>
		<i>function-definition</i>
		<i>datatype-definition</i>
		<i>type-annotation</i>
		<i>use-declaration</i>

We retain the environment notation conventions from the previous section.

Compost maintains a global  $\Gamma_g$  and  $\rho_g$  which are mutated by type annotations, **val** bindings, and function definitions. Additional bindings may be added to these environments at code points. A change to either of these global environments at a given code point is reflected at all succeeding code points. To determine the initial  $\Gamma$  or  $\rho$  at a **val** binding or function definition, we take the  $\Gamma_g$  and  $\rho_g$  at its opening parenthesis.

### 8.1 Type Annotations

<i>type-annotation</i>	$::=$	$(: \textit{ name type-expression})$
------------------------	-------	--------------------------------------

Type annotations constrain the type of globally bound names. Each bound name, whether bound by **val** or **define**, must have an associated type annotation. Consider a type annotation of the form:

$(: \textit{ n t})$

If  $n$  is bound by **define**, then we bind  $n$  to  $t$  in  $\Gamma_g$  at the first character of the file. If  $n$  is bound by **val**, then we bind  $n$  to  $t$  in  $\Gamma_g$  at the closing parenthesis of the **val** binding.

## 8.2 Val Bindings

*val-binding* ::= (**val** *name* *exp*)

Consider a **val** binding of the form:

(**val** *x* *e*)

Let  $\Gamma, \rho$  be  $\Gamma_g, \rho_g$  at the opening parenthesis of the binding. Let  $\Gamma_c$  be  $\Gamma_g$  at the closing parenthesis of the binding.

We assert that *x* be free in  $\rho$  and bound in  $\Gamma_g$ . Given  $\Gamma_c[\mathbf{n}] = \tau$ , we assert that  $\Gamma \vdash \mathbf{e} = \tau$ .

Let *v* be the result of evaluation *e* in environment  $\rho$ . We bind *x* to *v* in  $\rho_g$  at the closing parenthesis.

Note that if *e* produces a side effect, it is produced **only** when *n* is referenced and **every** time *x* is referenced. That is, references to **val**-bound names behave as zero-arity function calls rather than references to **let**-bound names. The secret sauce here is that **val** bindings are simply type-checked macros.

## 8.3 Function Definitions

*function-definition* ::= (**define** *name* (*{ name }*) *exp*)

Consider a function definition of the form:

(**define** *x* (*x1 ... xn*) *e*)

Let  $\Gamma, \rho$  be  $\Gamma_g, \rho_g$  at the opening parenthesis of the binding.

We assert that *x* be bound in  $\Gamma$  and free in  $\rho$ . We assert that  $\Gamma[x] = (-> (\tau_1 \dots \tau_n) \tau_r)$ . We assert that  $\Gamma[\mathbf{x1} \mapsto \tau_1, \dots, \mathbf{xn} \mapsto \tau_n] \vdash \mathbf{e} : \tau_r$ .

We bind *x*, in  $\rho_g$  at the first character of the file, to the function value that, when applied to arguments  $v_1, \dots, v_n$ , returns the result of evaluating *e* in the environment  $\rho[\mathbf{x1} \mapsto v_1, \dots, \mathbf{xn} \mapsto v_n]$ .

## 8.4 Use Declarations

*use-declaration* ::= (**use** *filename*)

Use declarations are a thinly-veiled preprocessor directive which are replaced by the contents of the named file.

## 9 Initial Basis

Compost includes a minimal initial basis providing those functions not possible or practical to define in terms of the rest of the core Compost language. A type annotation and a description will be provided for each such function.

## 9.1 Equality

`(: =i (-> (int int) bool))`

Integer equality.

`(: =b (-> (bool bool) bool))`

Boolean equality.

`(: =s (-> (sym sym) bool))`

Symbol equality.

`(: =u (-> (unit unit) bool))`

Unit equality. Always returns `true`.

## 9.2 Arithmetic

`(: + (-> (int int) int))`

Two's complement addition.

`(: - (-> (int int) int))`

Two's complement subtraction.

`(: * (-> (int int) int))`

Two's complement multiplication.

`(: / (-> (int int) int))`

Two's complement div.

`(: % (-> (int int) int))`

Two's complement modulus.

`(: neg (-> (int) int))`

Two's complement negation.

### 9.3 Comparison

```
(: > (-> (int int) bool))
```

Returns **true** if the first argument is greater than the second. Returns **false** otherwise.

```
(: < (-> (int int) bool))
```

Returns **true** if the first argument is less than the second. Returns **false** otherwise.

```
(: >= (-> (int int) bool))
```

Returns **true** if the first argument is greater than or equal to the second. Returns **false** otherwise.

```
(: <= (-> (int int) bool))
```

Returns **true** if the first argument is less than or equal to the second. Returns **false** otherwise.

### 9.4 Logic

```
(: not (-> (bool bool) bool))
```

Logical NOT.

```
(: and (-> (bool bool) bool))
```

Logical AND.

```
(: or (-> (bool bool) bool))
```

Logical OR.

```
(: xor (-> (bool bool) bool))
```

Logical XOR.