

Oliver Graf

THE

DROMEDAR

PROGRAMMING LANGUAGE

# Contents

1	Introduction	3
2	Typing System	3
3	Expressions	4
4	Handling Whitespace	4
5	Buildup of a Program	5
6	Garbage Collection	6
7	Formal Typing Rules	6
8	Formal Grammar	13

# 1 Introduction

## 1.1 Hello, World!

```
fn main -> void
    print_str("Hello, World!")
```

## 2 Typing System

Dromedar uses a static, strong and sound typing system – typing errors cannot happen at runtime. It knows primitive and reference types.

### 2.1 Primitives

There are four primitive types:

- **int** represents 64-bit 2's complement (signed) integers.
- **flt** represents 64-bit IEEE-754 standard floating point numbers.
- **char** represents 8-bit UTF-8 characters.
- **bool** represents a 1-bit value: **true** and **false**.

### 2.2 Reference Types

Reference types represent data structures that are laid over pointers to objects stored in the heap. References come in two types: Maybe-**null** and Definitely-not-**null** types. Operations like array subscript access are only possible with non-**null** types to ensure **null** safety.

With a non-null-type **t**, the type **t?** represents a reference type that allows **null** values. Primitives are non-nullable.

The following are reference types:

- **string** represents lists of characters.
- **[t]** represents an array with elements of type **t**.

Thus, e.g. **[[int]?)** represents a two-dimensional array of integers which is definitely non-**null**, whereas its rows may be **null**.

### 2.3 Mutability

In general, variables declared with the **let** keyword are immutable, whereas **mut** declarations create mutable variables.

For references, Dromedar uses a different notion of mutability: Allowing the object to point to new objects is handled with the **mut** declaration. However, even immutable objects are allowed to call methods which alter their internal state – for example changing an array's element.

### 2.4 Subtyping

Generally, variables and objects can be assigned values of *subtypes*. Every type is a subtype of itself, and e.g. **t** is a subtype of **t?**. Generally, a subtype is a restricted value set of its supertype – any subtype expression can be assigned to a variable of its supertype.

### 3 Expressions

Because Dromedar has a strong type system, it generally disallows any operations with operands of a non-specified type unless they are explicitly cast to the correct type before. This means that integer and floating point numbers cannot be added, multiplied, etc.

The following table describes precedence and type of all operators:

Operator	Name	Prec.	Assoc.	Types
-	Unary Negation	100	non-assoc.	<b>int</b> -> <b>int</b> <b>flt</b> -> <b>flt</b>
!	Logical Negation			<b>bool</b> -> <b>bool</b>
**	Exponentiation	90	right	<b>int,int</b> -> <b>int</b> <b>flt,flt</b> -> <b>flt</b> <b>int,flt</b> -> <b>flt</b> <b>flt,int</b> -> <b>flt</b>
*	Multiplication	80	left	<b>int,int</b> -> <b>int</b> <b>flt,flt</b> -> <b>flt</b> <b>int,flt</b> -> <b>flt</b> <b>flt,int</b> -> <b>flt</b>
+	String Addition	70	left	<b>string,string</b> -> <b>string</b>
+, -	Addition Subtraction	70	left	<b>int,int</b> -> <b>int</b> <b>flt,flt</b> -> <b>flt</b> <b>int,flt</b> -> <b>flt</b> <b>flt,int</b> -> <b>flt</b> <b>char,int</b> -> <b>char</b> <b>int,char</b> -> <b>char</b>
<<, >>, >>>	Left Shift Logical Right Shift Arithmetic Right Shift	60	left	<b>int,int</b> -> <b>int</b>
&	Bitwise And	60	left	<b>int,int</b> -> <b>int</b>
^	Bitwise Xor	50	left	<b>int,int</b> -> <b>int</b>
	Bitwise Or	40	left	<b>int,int</b> -> <b>int</b>
=, !=, >, <, >=, <=	Comparison	30	non-assoc.	[ <b>int/flt</b> ] -> <b>bool</b> [ <b>char</b> ] -> <b>bool</b>
=, !=, >, <, >=, <=	Structural Comparison	30	non-assoc.	[ <b>string</b> ] -> <b>bool</b>
==, !=	Reference Comparison	30	non-assoc.	[ <b>rt1...rtN</b> ] -> <b>bool</b>
&&	Logical And	20	left	<b>bool,bool</b> -> <b>bool</b>
	Logical Or	10	left	<b>bool,bool</b> -> <b>bool</b>

Consider the following example:

The expression `1+18-18+'a'` is well-typed (of type **char**) and gets parsed as `((1+18)-18)+'a'` and evaluated to `'b'`. `10 - 0.0` on the other hand is not well-typed as `-` cannot take an **int** and a **flt** operand as arguments.

Comparison operators work differently to other (binary) operators: Instead of comparing just two expressions, Dromedar allows chaining expressions to create one final boolean value: For example, `1 < 2 != 5 >= 5` holds because every single sub-expression (`1 < 2`, `2 != 5` and `5 >= 5`) holds. Every expression is only evaluated once for its side-effect.

Thus, `A op B op C` is not necessarily semantically equivalent to `A op B && B op C`.

### 4 Handling Whitespace

In order to keep the code simple and easy to look at, Dromedar uses significant whitespace: Blocks of code (such as bodies of **if** statements), are denoted by adding a level of indentation.

Every line is either empty (this includes lines containing only comments), or it contains code.

If a line contains code, the level of its indentation is determined by its relationship to the previous line and its environment:

Two neighboring lines of code within the same block of code must have exactly matching whitespace characters before their respective code starts. If a following line has a deeper level of indentation, it must match the whitespace characters of the previous line and then add a number of additional whitespace characters (space(s) and/or tab(s)).

A line of code can only have a deeper indentation level of one step compared to the previous line. The first line of code is always a global instruction and as such has the lowest level of indentation. If it is indented, this level of indentation corresponds to a baseline indentation that every line of code must share.

Consider this valid example:

```
global x := 3           # baseline indentation of two spaces
                        # empty line -> indentation doesn't matter
fn main : args:[string] -> int
    Stdio.println("Hello, World!") # deeper indentation level
    return 0                    # same indentation level
```

Blocks with the same level of indentation can have different indentation strings, but they must still match their environments, as follows:

```
if <condition>
    BLOCK A
else
    BLOCK B
```

The two blocks have a different indentation level, but from context it is still clear that **BLOCK A** is a sub-block of the **if**-statement, whereas **BLOCK B** belongs to the **else**-statement.

## 5 Buildup of a Program

A program consists of a series of global statements – global variable declarations and function definitions.

### 5.1 Global Declarations

Global variables are all assigned a value at the point of their declaration. This value is evaluated statically – declaration expressions can only contain literals and global variables that were already declared previously.

Functions are also declared globally. They can call each other and themselves recursively within their respective function bodies.

### 5.2 Builtin Functions

The Dromedar standard library includes the following functions:

#### 5.2.1 String Operations

String operations are always preceded by the **Str** library name.

Name	Type	Effect
<b>Str.of_int</b>	<b>int</b> -> <b>string</b>	transforms an integer into a string
<b>Str.of_flt</b>	<b>flt</b> -> <b>string</b>	transforms a decimal number into a string
<b>Str.concat</b>	( <b>string, string</b> ) -> <b>string</b>	concatenates two strings

#### 5.2.2 I/O

Printing operations are always preceded by the **IO** library name.

Name	Type	Effect
<b>IO.print_str</b>	<b>string</b> -> <b>void</b>	prints a string to the console
<b>IO.print_int</b>	<b>int</b> -> <b>void</b>	prints an integer to the console
<b>IO.print_flt</b>	<b>flt</b> -> <b>void</b>	prints a real number to the console

## 6 Garbage Collection

Dromedar uses an algorithm that combines the mark/sweep and reference counting approaches. It is a *precise* garbage collection algorithm, meaning that it is capable of collecting all non-reachable objects and it will only attempt to GC exactly these (as opposed to conservative GC algorithms like the Boehm-Demers-Weiser garbage collector for C/C++).

Internally, each reference object is stored in a central garbage collection table that counts the numbers of program references that can reach a program, as well as the set of its children.

When a garbage collection run is triggered, the collector will free all objects that are not reachable. An object is deemed **reachable** if either its count of program references is nonzero, or if it is the child of a reachable object.

## 7 Formal Typing Rules

A typing rule takes the following shape:

$$\frac{\text{Hypotheses}}{S, \dots \vdash_{\text{type}} \text{grammar spec}} \text{NAME}$$

Here,  $S$  represents a stack of symbol definitions:  $S \in (\text{id} \times \text{type} \times \{c, m\})^n$  for some block depth  $n$  at any given point. In global context,  $S$  has only one layer.  $\text{id}$  corresponds to the set of names that variables can have (related to the Lexer symbol `%Identifier`),  $\text{type}$  to the set of types in a given program (related to the Parser symbol `Type`), and  $\{c, m\}$  to the mutability of the object:  $c$  represents an immutable value (as declared by **let**), and  $m$  a mutable one (declared by **mut**).

The symbol  $\in$  is defined as follows:  $s \in S \Leftrightarrow s$  is contained in *any* layer of  $S$ , whereas  $\in_0$  is true only if the symbol is at the top level of the symbol stack (i.e. defined in the same block). The operator  $\cup$  on  $S$  adds another binding to the top layer of the stack, whereas  $\sqcup$  adds another layer to the stack.

The following are the typing rules for Dromedar programs:

### 7.1 Subtyping Rules

#### 7.1.1 Cross-Typing

Cross types are types which aren't related by the subtype relation  $\preceq$  but still allow for some typesafe interaction – e.g. by assigning variables of one type to variables of another.

The crosstype relation, denoted by the  $\asymp$  operator, commutes.

$$\frac{}{\vdash_{\text{T}} \text{int} \asymp \text{flt}} \text{CROSSTYINTFLT}, \quad \frac{}{\vdash_{\text{T}} \text{flt} \asymp \text{int}} \text{CROSSTYFLTINT}$$

#### 7.1.2 Trivial Rule

$$\frac{}{\vdash_{\text{T}} \mathbf{t} \preceq \mathbf{t}} \text{SUBTYTRIVIAL}$$

#### 7.1.3 References

$$\frac{\vdash_{\text{T}} \mathbf{t1} \preceq \mathbf{t2}}{\vdash_{\text{T}} \mathbf{t1} \preceq \mathbf{t2?}} \text{SUBTYREFS}, \quad \frac{\vdash_{\text{T}} \mathbf{t1} \preceq \mathbf{t2}}{\vdash_{\text{T}} \mathbf{t1?} \preceq \mathbf{t2?}} \text{SUBTYREFS}$$

#### 7.1.4 Arrays

$$\frac{\vdash_{\text{T}} \mathbf{t1} \preceq \mathbf{t2}}{\vdash_{\text{T}} [\mathbf{t1}] \preceq [\mathbf{t2}]} \text{SUBTYFUNS}$$

#### 7.1.5 Functions

$$\frac{\vdash_{\text{T}} \mathbf{u1} \preceq \mathbf{t1}, \dots, \vdash_{\text{T}} \mathbf{un} \preceq \mathbf{tn} \quad \vdash_{\text{T}} \mathbf{rt} \preceq \mathbf{ru}}{\vdash_{\text{T}} (\mathbf{t1}, \dots, \mathbf{tn}) \rightarrow \mathbf{rt} \preceq (\mathbf{u1}, \dots, \mathbf{un}) \rightarrow \mathbf{ru}} \text{SUBTYFUNCS}$$

### 7.1.6 Subtypes and Supertypes

The functions `subtys` and `suptys` of types  $\rightarrow \mathcal{P}(\text{types})$  compute the sub- and supertype set of the input type, respectively.

They are used – among others – in the `EXPLITARR` rule.

$$\begin{aligned} \text{subtys} := & \begin{cases} \mathbf{int} & \mapsto \{\mathbf{int}\} \\ \mathbf{flt} & \mapsto \{\mathbf{flt}\} \\ \mathbf{char} & \mapsto \{\mathbf{char}\} \\ \mathbf{bool} & \mapsto \{\mathbf{bool}\} \\ \mathbf{string} & \mapsto \{\mathbf{string}\} \\ t? & \mapsto \{u, u? \mid u \in \text{subtys}(t)\} \\ [t] & \mapsto \{[u] \mid u \in \text{subtys}(t)\} \\ (t_1 \dots t_n) \rightarrow rt & \mapsto \{(u_1 \dots u_n) \rightarrow st \mid u_1 \dots u_n \in \text{subtys}(t_1 \dots t_n), st \in \text{subtys}(rt)\} \end{cases} \\ \text{suptys} := & \begin{cases} \mathbf{int} & \mapsto \{\mathbf{int}\} \\ \mathbf{flt} & \mapsto \{\mathbf{flt}\} \\ \mathbf{char} & \mapsto \{\mathbf{char}\} \\ \mathbf{bool} & \mapsto \{\mathbf{bool}\} \\ \mathbf{string} & \mapsto \{\mathbf{string}, \mathbf{string}?\} \\ t? & \mapsto \{u? \mid u \in \text{suptys}(t)\} \\ [t] & \mapsto \{[u], [u]? \mid u \in \text{suptys}(t)\} \\ (t_1 \dots t_n) \rightarrow rt & \mapsto \left\{ \begin{array}{l} (u_1 \dots u_n) \rightarrow st, ((u_1 \dots u_n) \rightarrow st)? \\ \mid u_1 \dots u_n \in \text{suptys}(t_1 \dots t_n), st \in \text{suptys}(rt) \end{array} \right\} \end{cases} \end{aligned}$$

## 7.2 Builtin Operators

Many builtin operators are overloaded, providing functionality for multiple input types.

### 7.2.1 Unary Operators

#### 7.2.1.1 Arithmetic Negation

$$\frac{}{\vdash_0 - :: \mathbf{int} \rightarrow \mathbf{int}} \text{TyUopNegInt}, \quad \frac{}{\vdash_0 - :: \mathbf{flt} \rightarrow \mathbf{flt}} \text{TyUopNegInt}$$

#### 7.2.1.2 Logical Negation

$$\frac{}{\vdash_0 ! :: \mathbf{bool} \rightarrow \mathbf{bool}} \text{TyUopNot}$$

### 7.2.2 Binary Operators

#### 7.2.2.1 Power

$$\begin{aligned} & \frac{}{\vdash_0 ** :: (\mathbf{int}, \mathbf{int}) \rightarrow \mathbf{int}} \text{TyBopPowInt}, \quad \frac{}{\vdash_0 ** :: (\mathbf{flt}, \mathbf{flt}) \rightarrow \mathbf{flt}} \text{TyBopPowFlt} \\ & \frac{}{\vdash_0 ** :: (\mathbf{int}, \mathbf{flt}) \rightarrow \mathbf{flt}} \text{TyBopPowIntFlt}, \quad \frac{}{\vdash_0 ** :: (\mathbf{flt}, \mathbf{int}) \rightarrow \mathbf{flt}} \text{TyBopPowFltInt} \end{aligned}$$

#### 7.2.2.2 Multiplication

$$\begin{aligned} & \frac{}{\vdash_0 * :: (\mathbf{int}, \mathbf{int}) \rightarrow \mathbf{int}} \text{TyBopMulInt}, \quad \frac{}{\vdash_0 * :: (\mathbf{flt}, \mathbf{flt}) \rightarrow \mathbf{flt}} \text{TyBopMulInt} \\ & \frac{}{\vdash_0 * :: (\mathbf{int}, \mathbf{flt}) \rightarrow \mathbf{flt}} \text{TyBopMulIntFlt}, \quad \frac{}{\vdash_0 * :: (\mathbf{flt}, \mathbf{int}) \rightarrow \mathbf{flt}} \text{TyBopMulFltInt} \end{aligned}$$

### 7.2.2.3 Addition and Subtraction

•

$$\begin{array}{c}
\frac{}{\vdash_0 + :: (\mathbf{int}, \mathbf{int}) \rightarrow \mathbf{int}} \text{TyBopAddInt}, \quad \frac{}{\vdash_0 + :: (\mathbf{flt}, \mathbf{flt}) \rightarrow \mathbf{flt}} \text{TyBopAddFlt}, \\
\frac{}{\vdash_0 + :: (\mathbf{int}, \mathbf{flt}) \rightarrow \mathbf{flt}} \text{TyBopAddIntFlt}, \quad \frac{}{\vdash_0 + :: (\mathbf{flt}, \mathbf{int}) \rightarrow \mathbf{flt}} \text{TyBopAddFltInt} \\
\frac{}{\vdash_0 + :: (\mathbf{int}, \mathbf{char}) \rightarrow \mathbf{char}} \text{TyBopAddCharR}, \quad \frac{}{\vdash_0 + :: (\mathbf{char}, \mathbf{int}) \rightarrow \mathbf{char}} \text{TyBopAddCharL} \\
\\
\frac{}{\vdash_0 + :: (\mathbf{string}, \mathbf{string}) \rightarrow \mathbf{string}} \text{TyBopAddString} \\
\\
\frac{\mathbf{t} = \min_{\prec}(\text{suptys}(\mathbf{t1}) \cup \text{suptys}(\mathbf{t2}))}{\vdash_0 + :: ([\mathbf{t1}], [\mathbf{t2}]) \rightarrow [\mathbf{t}]} \text{TyBopAddArr}
\end{array}$$

•

$$\begin{array}{c}
\frac{}{\vdash_0 - :: (\mathbf{int}, \mathbf{int}) \rightarrow \mathbf{int}} \text{TyBopSubInt}, \quad \frac{}{\vdash_0 - :: (\mathbf{flt}, \mathbf{flt}) \rightarrow \mathbf{flt}} \text{TyBopSubFlt}, \\
\frac{}{\vdash_0 - :: (\mathbf{int}, \mathbf{flt}) \rightarrow \mathbf{flt}} \text{TyBopSubIntFlt}, \quad \frac{}{\vdash_0 - :: (\mathbf{flt}, \mathbf{int}) \rightarrow \mathbf{flt}} \text{TyBopSubFltInt} \\
\frac{}{\vdash_0 - :: (\mathbf{int}, \mathbf{char}) \rightarrow \mathbf{char}} \text{TyBopSubCharR}, \quad \frac{}{\vdash_0 - :: (\mathbf{char}, \mathbf{int}) \rightarrow \mathbf{char}} \text{TyBopSubCharL}
\end{array}$$

### 7.2.2.4 Shift Operators

•

$$\frac{}{\vdash_0 << :: (\mathbf{int}, \mathbf{int}) \rightarrow \mathbf{int}} \text{TyBopLShift}$$

•

$$\frac{}{\vdash_0 >> :: (\mathbf{int}, \mathbf{int}) \rightarrow \mathbf{int}} \text{TyBopBitRShift}$$

•

$$\frac{}{\vdash_0 >>> :: (\mathbf{int}, \mathbf{int}) \rightarrow \mathbf{int}} \text{TyBopBitAShift}$$

### 7.2.2.5 Bitwise Operators

•

$$\frac{}{\vdash_0 \& :: (\mathbf{int}, \mathbf{int}) \rightarrow \mathbf{int}} \text{TyBopBitAnd}$$

•

$$\frac{}{\vdash_0 \wedge :: (\mathbf{int}, \mathbf{int}) \rightarrow \mathbf{int}} \text{TyBopXor}$$

•

$$\frac{}{\vdash_0 | :: (\mathbf{int}, \mathbf{int}) \rightarrow \mathbf{int}} \text{TyBopBitOr}$$

### 7.2.2.6 Logical Operators

•

$$\frac{}{\vdash_0 \&\& :: (\mathbf{bool}, \mathbf{bool}) \rightarrow \mathbf{bool}} \text{TyBopLogAnd}$$

•

$$\frac{}{\vdash_0 \wedge\wedge :: (\mathbf{bool}, \mathbf{bool}) \rightarrow \mathbf{bool}} \text{TyBopLogXor}$$

•

$$\frac{}{\vdash_0 || :: (\mathbf{bool}, \mathbf{bool}) \rightarrow \mathbf{bool}} \text{TyBopLogOr}$$



### 7.2.2.7 Comparison Operators

$$\frac{}{\vdash_0 \{=, !=, >, <, >=, <= \} :: (\mathbf{int}, \dots, \mathbf{int}) \rightarrow \mathbf{bool}} \text{TyCmpListInt}$$

$$\frac{}{\vdash_0 \{=, !=, >, <, >=, <= \} :: (\mathbf{flt}, \dots, \mathbf{flt}) \rightarrow \mathbf{bool}} \text{TyCmpListFlt}$$

$$\frac{}{\vdash_0 \{=, !=, >, <, >=, <= \} :: (\mathbf{char}, \dots, \mathbf{char}) \rightarrow \mathbf{bool}} \text{TyCmpListChar}$$

$$\frac{\vdash_T t_1 \preceq \mathbf{string}, \dots, \vdash_T t_n \preceq \mathbf{string}}{\vdash_0 \{=, !=, >, <, >=, <= \} :: (t_1, \dots, t_n) \rightarrow \mathbf{bool}} \text{TyCmpListRefStr}$$

$$\frac{\vdash_T t_1 \preceq / \succeq t_2, \dots, \vdash_T t_{(n-1)} \preceq / \succeq t_n \quad t_1, \dots, t_n \text{ reference types}}{\vdash_0 \{==, !=\} :: (t_1, \dots, t_n) \rightarrow \mathbf{bool}} \text{TyCmpListRefVal}$$

## 7.3 Expressions

### 7.3.1 Subtyping Expression Rule

To reduce proof rule size, define the following rules:

$$\frac{S \vdash_E \text{exp} :: t_1 \quad \vdash_T t_1 \preceq t}{S \vdash_{ET} \text{exp} \preceq t} \text{ExpSubLocal}, \quad \frac{S \vdash_G \text{exp} :: t_1 \quad \vdash_T t_1 \preceq t}{S \vdash_{GT} \text{exp} \preceq t} \text{ExpSubGlobal}$$

$$\frac{S \vdash_E \text{exp} :: t_1, \quad \vdash_T t_1 \asymp t}{S \vdash_{ET} \text{exp} \asymp t} \text{ExpCrossLocal}, \quad \frac{S \vdash_E \text{exp} :: t_1, \quad \vdash_T t_1 \asymp t}{S \vdash_{GT} \text{exp} \asymp t} \text{ExpCrossGlobal}$$

### 7.3.2 Assignability

The rule  $\vdash_A$  defines when an expression can be assigned a value;  $\vdash_{AC}$  is equivalent to  $\vdash_A$  – except that it also allows immutable objects whose children (e.g. array elements) can still be modified.

$$\frac{(\mathbf{id}, \_, m) \in S}{S \vdash_A \mathbf{id}} \text{ExpAssnID}, \quad \frac{S \vdash_{AC} e_1}{S \vdash_A e_1[e_2]} \text{ExpAssnSub}$$

$$\frac{(\mathbf{id}, \_, \_) \in S}{S \vdash_{AC} \mathbf{id}} \text{ExpAssnID}', \quad \frac{S \vdash_{AC} e_1}{S \vdash_{AC} e_1[e_2]} \text{ExpAssnSub}'$$

### 7.3.3 Global Expressions

The rule  $\vdash_G$  describes global expressions, which are restricted in a way that they can be computed at compile time. Global variables are also strictly non-**null**.

The following copied rules use  $\vdash_G$  instead of  $\vdash_E$

#### 7.3.3.1 Literals

- $\text{GExpLitInt} := \text{ExpLitInt}$
- $\text{GExpLitFlt} := \text{ExpLitFlt}$
- $\text{GExpLitChar} := \text{ExpLitChar}$
- $\text{GExpLitBool} := \text{ExpLitBool}$

#### 7.3.3.2 Other Rules

- $\text{GExpID} := \text{ExpID}$
- $\text{GExpUOp} := \text{ExpUOp}$
- $\text{GExpBOp} := \text{ExpBOp}$
- $\text{GExpCmpList} := \text{ExpCmpList}$

Note that global variable declarations cannot feature function calls or **null** declarations.

### 7.3.4 Literals

•

$$\frac{}{\vdash_E n :: \mathbf{int}} \text{EXPLITINT}$$

•

$$\frac{}{\vdash_E f :: \mathbf{flt}} \text{EXPLITFLT}$$

•

$$\frac{}{\vdash_E c :: \mathbf{char}} \text{EXPLITCHAR}$$

•

$$\frac{}{\vdash_E \mathbf{true} :: \mathbf{bool}} \text{EXPLITBOOLTRUE}, \frac{}{\vdash_E \mathbf{false} :: \mathbf{bool}} \text{EXPLITBOOLFALSE}$$

•

$$\frac{}{\vdash_E s :: \mathbf{string}} \text{EXPLITSTRING}$$

- In arrays, the typechecker looks for the common subtypes of all array literal elements and looks for the one type  $\mathbf{t}$  that is a supertype of all elements and which is a subtype of all other such subtypes (the minimum of the subtypes given the  $\preceq$  relation on types).

$$\frac{S \vdash_E e1 :: t1, \dots, S \vdash_E en :: tn \quad \mathbf{t} = \min_{\preceq}(\bigcap_{i=1}^n \text{suptys}(\mathbf{t}_i))}{S \vdash_E [e1, \dots, en] :: [\mathbf{t}]} \text{EXPLITARR}$$

Because the graph connecting types and subtypes is a forest of trees, if the intersection of supertypes is nonempty there is a unique solution  $\mathbf{t}$ .

### 7.3.5 Null

$$\frac{\mathbf{t} \text{ is a non-null ref. type}}{S \vdash_E \mathbf{null of t} :: \mathbf{t?}} \text{EXPNULL}$$

### 7.3.6 Identifiers

$$\frac{(id, \mathbf{t}, \_) \in S}{S \vdash_E id :: \mathbf{t}} \text{EXPID}$$

### 7.3.7 Unary Operations

Unary and Binary Operations do not need to do subtype checking, as they operate only on primitives.

$$\frac{S \vdash_E \text{exp} :: t1 \quad \vdash_O \text{op} :: t1 \rightarrow t}{S \vdash_E \text{op exp} :: t} \text{EXPUOP}$$

### 7.3.8 Binary Operations

$$\frac{S \vdash_E e1 :: t1 \quad S \vdash_{ET} e2 :: t2 \quad \vdash_O \text{op} :: (t1, t2) \rightarrow t}{S \vdash_E e1 \text{ op } e2 :: t} \text{EXPBOP}$$

### 7.3.9 Function Calls

$$\frac{S \vdash_E f :: (t1, \dots, tn) \rightarrow rt \quad rt \neq \mathbf{void} \quad S \vdash_{ET} e1 \sqsubseteq t1 \vee e1 \bowtie t1, \dots, S \vdash_{ET} en \sqsubseteq tn \vee en \bowtie tn}{S \vdash_E f(e1, \dots, en) :: rt} \text{EXPFUNC}$$

### 7.3.10 Subscript Access

•

•

$$\frac{S \vdash_E s :: \mathbf{string} \quad S \vdash_E i :: \mathbf{int}}{S \vdash_E s[i] :: \mathbf{char}} \text{EXPARRSUB}$$

$$\frac{S \vdash_E e :: [t] \quad S \vdash_E i :: \mathbf{int}}{S \vdash_E e[i] :: t} \text{EXPARRSUB}$$

### 7.3.11 Comparison Lists

$$\frac{\vdash_O \text{op1} :: (t_0, t_1) \rightarrow t, \dots, \vdash_O \text{opn} :: (t_{n-1}, t_n) \rightarrow t \quad S \vdash_{ET} (e_0, \dots, e_n) [\leq \mid \bowtie] (t_0, \dots, t_n)}{S \vdash_E e_0 \text{ op1 } \dots \text{ opn } e_n :: t} \text{EXPCMPLIST}$$

## 7.4 Statements

In the statement typing rules, a statement rule produces a tuple  $(S, r)$  where  $S$  stands for the newly updated context, and  $r \in \{\perp, \top\}$ , where  $\perp$  means that a statement might not return and  $\top$  means that a statement definitely returns.

To prevent potential mistakes, the typechecker prevents statements which are deemed unreachable at compile time. The logical operators  $\vee$  and  $\wedge$  operate as if  $\perp \equiv 0$  and  $\top \equiv 1$ .

### 7.4.1 Local Variable Declarations

$$\frac{(id, \_, \_) \notin_0 S \quad S \vdash_E \text{exp} :: t}{S, rt \vdash_S \mathbf{let} \text{id} := \text{exp} \Rightarrow (S \cup (id, t, c)), \perp} \text{STMTVDECLCONST}$$

$$\frac{(id, \_, \_) \notin_0 S \quad S \vdash_{ET} \text{exp} \leq t \vee \text{exp} \bowtie t}{S, rt \vdash_S \mathbf{let} \text{id}:t := \text{exp} \Rightarrow (S \cup (id, t, c)), \perp} \text{STMTVTDECLCONST}$$

$$\frac{(id, \_, \_) \notin_0 S \quad S \vdash_E \text{exp} :: t}{S, rt \vdash_S \mathbf{mut} \text{id} := \text{exp} \Rightarrow (S \cup (id, t, m)), \perp} \text{STMTVDECLMUT}$$

$$\frac{(id, \_, \_) \notin_0 S \quad S \vdash_{ET} \text{exp} \leq t \vee \text{exp} \bowtie t}{S, rt \vdash_S \mathbf{mut} \text{id}:t := \text{exp} \Rightarrow (S \cup (id, t, m)), \perp} \text{STMTVTDECLMUT}$$

### 7.4.2 Assignments

$$\frac{S \vdash_A \text{lhs} \quad S \vdash_E \text{lhs} :: t \quad S \vdash_{ET} \text{exp} \leq t \vee \text{exp} \bowtie t}{S, rt \vdash_S \text{lhs} := \text{exp} \Rightarrow S, \perp} \text{STMTASSN}$$

### 7.4.3 Expression Statements

$$\frac{S \vdash_E \text{exp} :: t}{S, rt \vdash_S \text{exp} \Rightarrow S, \perp} \text{STMTEXPR} \quad \frac{S \vdash_E (t_1, \dots, t_n) \rightarrow \mathbf{void}}{S, rt \vdash_S \text{exp}(a_1, \dots, a_n) \Rightarrow S, \perp} \text{STMTEXPRVOID}$$

### 7.4.4 If Statements

$$\frac{S \vdash_E c :: \mathbf{bool} \quad S, rt \vdash_S b_1 \Rightarrow S_1, R_1 \quad S, rt \vdash_S b_2 \Rightarrow S_2, R_2}{S, rt \vdash_S \mathbf{if} \ c \ b_1 \ \mathbf{else} \ b_2 \Rightarrow S, R_1 \wedge R_2} \text{STMTIF}$$

### 7.4.5 Checked Null Casts

$$\frac{S \vdash_E \text{exp} :: t? \quad S \cup \{(r, t, c)\}, rt \vdash_S b_1 \Rightarrow S', R_1 \quad S, rt \vdash_S b_2 \Rightarrow S'', R_2}{S, rt \vdash_S \mathbf{dennull} \ r := \text{exp} \ b_1 \ \mathbf{else} \ b_2 \Rightarrow S, R_1 \wedge R_2} \text{STMTNULLCAST}$$

### 7.4.6 While Statements

$$\frac{S \vdash_E c :: \mathbf{bool} \quad S, rt \vdash_S b \Rightarrow S', R}{S, rt \vdash_S \mathbf{while} \ c \ b \Rightarrow S, \perp} \text{STMTWHILE}$$

#### 7.4.7 Do-While Statements

$$\frac{S \vdash_E c :: \mathbf{bool} \quad S, \text{rt} \vdash_S b \Rightarrow S', R}{S, \text{rt} \vdash_S \mathbf{do} \ b \ \mathbf{while} \ c \Rightarrow S, R} \text{STMTDoWHILE}$$

#### 7.4.8 For Statements

$$\frac{S \vdash_E \text{estart} :: \mathbf{int} \quad S \vdash_E \text{eend} :: \mathbf{int} \quad S \cup \{(\text{id}, \mathbf{int}, c)\}, \text{rt} \vdash_S b \Rightarrow S', R}{S, \text{rt} \vdash_S \mathbf{for} \ \text{id} := \text{estart} \ \% \text{ForSpecifier} \ \text{eend} \ b \Rightarrow S, \perp} \text{STMTFor}$$

#### 7.4.9 Return Statements

$$\frac{S \vdash_{ET} \text{exp} \trianglelefteq \text{rt} \vee \text{exp} \bowtie \text{rt}}{S, \text{rt} \vdash_S \mathbf{return} \ \text{exp} \Rightarrow S, \top} \text{STMTReturnEXP}, \quad \frac{}{S, \mathbf{void} \vdash_S \mathbf{return} \Rightarrow S, \top} \text{STMTReturn}$$

#### 7.4.10 Blocks

$$\frac{S \sqcup \{\}, \text{rt} \vdash_S s1 \Rightarrow S_1, \perp, \quad S_1, \text{rt} \vdash_S s2 \Rightarrow S_2, \perp, \quad \dots, \quad S_{n-1}, \text{rt} \vdash_S s_n \Rightarrow S_n, R}{S, \text{rt} \vdash_S s1 \ \dots \ s_n \Rightarrow S_n, R} \text{STMTBlock}$$

### 7.5 Global Statements

#### 7.5.1 Global Function Declaration

$$\frac{S \sqcup \{(a1, t1, c), \dots, (an, tn, c)\}, \text{rt} \vdash_S b \Rightarrow S', \top \text{ (or } \text{rt} = \mathbf{void}) \quad a1, \dots, an \text{ distinct}}{S \vdash_G \mathbf{fn} \ \text{id} : a1:t1, \dots, an:tn \rightarrow \text{rt} \ b \Rightarrow S} \text{GSTMTFDECL}$$

#### 7.5.2 Global Variable Declaration

$$\frac{(id, \_, \_) \notin S \quad S \vdash_G \text{exp} :: t \quad t \text{ non-null}}{S \vdash_G \mathbf{global} \ \text{id} := \text{exp} \Rightarrow S \cup \{(id, t, c)\}} \text{GSTMTVDECLCONST}$$

$$\frac{(id, \_, \_) \notin S \quad S \vdash_{GT} \text{exp} \trianglelefteq t \quad t \text{ non-null}}{S \vdash_G \mathbf{global} \ \text{id}:t := \text{exp} \Rightarrow S \cup \{(id, t, c)\}} \text{GSTMTVDECLCONST}$$

$$\frac{(id, \_, \_) \notin S \quad S \vdash_G \text{exp} :: t \quad t \text{ non-null}}{S \vdash_G \mathbf{global} \ \mathbf{mut} \ \text{id} := \text{exp} \Rightarrow S \cup \{(id, t, m)\}} \text{GSTMTVDECLMUT}$$

$$\frac{(id, \_, \_) \notin S \quad S \vdash_{GT} \text{exp} \trianglelefteq t \quad t \text{ non-null}}{S \vdash_G \mathbf{global} \ \mathbf{mut} \ \text{id}:t := \text{exp} \Rightarrow S \cup \{(id, t, m)\}} \text{GSTMTVDECLMUT}$$

#### 7.5.3 Program

$$\frac{S_0 \vdash_G \text{gs1} \Rightarrow S_1, \dots, S_{n-1} \vdash_G \text{gsn} \Rightarrow S_n}{S_0 \vdash_G \text{gs1} \ \dots \ \text{gsn} \Rightarrow S_n} \text{GSTMTPROGRAM}$$

### 7.6 Context Buildup

#### 7.6.1 Global Function Declarations

$$\frac{(id, \_) \notin S}{S \vdash_{GCF} \mathbf{fn} \ \text{id} : a1:t1, \dots, an:tn \rightarrow \text{rt} \ b \Rightarrow S \cup (id, (t1, \dots, tn) \rightarrow \text{rt}, c)} \text{GSTMTFCtxtFDECL}$$

$$\frac{}{S \vdash_{GCF} \mathbf{global} \ \text{id} := \text{exp} \Rightarrow S} \text{GSTMTFCtxtVDECLCONST}$$

$$\frac{}{S \vdash_{GCF} \mathbf{global} \ \text{id}:t := \text{exp} \Rightarrow S} \text{GSTMTFCtxtVTDECLCONST}$$

$$\frac{}{S \vdash_{GCF} \mathbf{global} \ \mathbf{mut} \ \text{id} := \text{exp} \Rightarrow S} \text{GSTMTFCtxtVDECLMUT}$$

$$\frac{}{S \vdash_{GCF} \mathbf{global} \ \mathbf{mut} \ \text{id}:t := \text{exp} \Rightarrow S} \text{GSTMTFCtxtVTDECLMUT}$$

## 7.6.2 Program: Functions

$$\frac{S_0 \vdash_{\text{GCF}} \text{gs1} \Rightarrow S_1, \dots, S_{n-1} \vdash_{\text{GCF}} \text{gsn} \Rightarrow S_n}{S_0 \vdash_{\text{GC}} \text{gs1} \dots \text{gsn} \Rightarrow S_n} \text{GSTMTCTXTFUNCS}$$

## 7.7 Rule for Program Typechecking

Let  $S^*$  be the starting context which contains the builtin context. It looks as follows:

$$S^* := \{\} \sqcup \left\{ \begin{array}{ll} (\text{string\_of\_int}, & \mathbf{int} \rightarrow \mathbf{string}, & c), \\ (\text{string\_of\_flt}, & \mathbf{flt} \rightarrow \mathbf{string}, & c), \\ (\text{string\_concat}, & (\mathbf{string}, \mathbf{string}) \rightarrow \mathbf{string}, & c), \\ (\text{io\_print}, & \mathbf{string} \rightarrow \mathbf{void}, & c) \end{array} \right\}$$

$$\frac{S^* \vdash_{\text{GC}} \text{prog} \Rightarrow S' \quad S' \vdash_{\text{G}} \text{prog} \Rightarrow S}{\vdash \text{prog}} \text{PROG}$$

# 8 Formal Grammar

## 8.1 Lexer Grammar

The Lexer Grammar is specified using **regular expressions**:

```

LiteralInt    ::= /[1-9]\d*/
LiteralFlt    ::= /\d+\.\d*/
LiteralChar   ::= /'([^'\\"|(\\"\\nrt'))'/
LiteralBool   ::= /true|false/
LiteralStr    ::= /"([^"\"|(\\"\\nrt'))*"/

Identifier    ::= /[a-zA-Z][a-zA-Z0-9_]*/

global        ::= /global/
fn            ::= /fn/
let           ::= /let/
mut           ::= /mut/

int           ::= /int/
flt           ::= /flt/
char          ::= /char/
bool          ::= /bool/
string        ::= /string/
void          ::= /void/

null          ::= /null/
dennull       ::= /dennull/

of            ::= /of/

if            ::= /if/
elif          ::= /elif/
else          ::= /else/
do            ::= /do/
while         ::= /while/
for           ::= /for/

return        ::= /return/

Dash          ::= /\-/
Bang          ::= /!/

```

```

Star      ::= /\*/
Plus      ::= /\+/
LShift    ::= /<</
RShift    ::= />>/
AShift    ::= />>>/
Bitand    ::= /\&/
Xor       ::= /\^/
Bitor     ::= /\|/
Logand    ::= /\&&/
Logor     ::= /\||/

Equal     ::= /=/
NotEqual  ::= /!=/
Greater   ::= />/
Less      ::= /</
GreaterEq ::= />=/
LessEq    ::= /<=/

RefEqual  ::= /==/
RefNotEqual ::= /!==/

Assign    ::= /:=/

Colon     ::= /:/
Arrow     ::= /\->/

Comma     ::= /\,/

Dots      ::= /\.\.\./
DotsPipe  ::= /\.\.\|/
PipeDots  ::= /\|\.\.\./
PipeDotPipe ::= /\|\.\.\|/

LParen    ::= /\(/
RParen    ::= /\)/
LBrack    ::= /\[/
RBrack    ::= /\]/

QuestionMark ::= /\?/

```

## 8.2 Parser Grammar

The following grammar specification uses a preceding % for lexer tokens.

```

Program      ::=
    | €
    | GlobalStatement Program

GlobalStatement ::=
    | GVDeclaration
    | GFDeclaration

GVDeclaration ::=
    | %Global      %Identifier          %Assign GlobalExpression
    | %Global      %Identifier %Colon Type %Assign GlobalExpression
    | %Global %Mut %Identifier          %Assign GlobalExpression
    | %Global %Mut %Identifier %Colon Type %Assign GlobalExpression

GFDeclaration ::=
    | %Fn %Identifier          %Arrow ReturnType Block
    | %Fn %Identifier %Colon FArguments %Arrow ReturnType Block

```

FArguments	<pre> ::=   %Identifier %Colon Type   %Identifier %Colon Type %Comma FArguments </pre>
ReturnType	<pre> ::=   %Void   Type </pre>
Block	<pre> ::=   €   Statement Block </pre>
Statement	<pre> ::=   VDeclaration   AssignStmt   IfStmt   NullCastStmt   WhileStmt   DoWhileStmt   ForStmt   ExprStmt   ReturnStmt </pre>
VDeclaration	<pre> ::=   %Let %Identifier %Assign Expression   %Let %Identifier %Colon Type %Assign Expression   %Mut %Identifier %Assign Expression   %Mut %Identifier %Colon Type %Assign Expression </pre>
AssignStmt	<pre> ::=   LHS %Assign Expression </pre>
IfStmt	<pre> ::=   %If Expression Block ElifStmt </pre>
ElifStmt	<pre> ::=   %Elif Expression Block ElifStmt   %Else Block   € </pre>
NullCastStmt	<pre> ::=   %Dnull %Id %Assign Expression Block   %Dnull %Id %Assign Expression Block %Else Block </pre>
WhileStmt	<pre> ::=   %While Expression Block </pre>
DoWhileStmt	<pre> ::=   %Do Block %While Expression </pre>
ForStmt	<pre> ::=   %For %Id %Assign Expr ForSpecifier Expr Block </pre>
ForSpecifier	<pre> ::=   %Dots   %DotsPipe   %PipeDots   %PipeDotPipe </pre>
ExprStmt	<pre> ::= </pre>

	Expression
	Expression %Colon FArgs
ReturnStmt	::=   %Return   %Return Expression
LHS	::=   %Identifier
Type	::=   %Int   %Flt   %Char   %Bool   RefType   RefType %QuestionMark
RefType	::=   %Str   %LBrack Type %RBrack
GlobalExpression	::= Expression*
Expression	::=   %Null %Of RefType   %LBrack %RBrack %Of Type   ExprPrec30
ExprPrec20	::=   ExprPrec30   ExprPrec30 %Equal ExprPrec20   ExprPrec30 %NotEqual ExprPrec20   ExprPrec30 %Greater ExprPrec20   ExprPrec30 %Less ExprPrec20   ExprPrec30 %GreaterEq ExprPrec20   ExprPrec30 %LessEq ExprPrec20   ExprPrec30 %RefEqual ExprPrec20   ExprPrec30 %RefNotEqual ExprPrec20
ExprPrec30	::=   ExprPrec40   ExprPrec40 %Bitor ExprPrec30
ExprPrec40	::=   ExprPrec50   ExprPrec50 %Xor ExprPrec40
ExprPrec50	::=   ExprPrec60   ExprPrec60 %Bitand ExprPrec50
ExprPrec60	::=   ExprPrec70   ExprPrec70 %LShift ExprPrec60   ExprPrec70 %RShift ExprPrec60   ExprPrec70 %AShift ExprPrec60
ExprPrec70	::=   ExprPrec80   ExprPrec80 %Plus ExprPrec70   ExprPrec80 %Minus ExprPrec70
ExprPrec80	::=   ExprPrec90   ExprPrec90 %Star ExprPrec80
ExprPrec90	::=



```

ExprPrec100      | ExprPrec100
                  | ExprPrec100 %StarStar ExprPrec90
ExprPrec100 ::=
                  | SimpleExpression
                  | %Dash ExprPrec100
                  | %Bang ExprPrec100

SimpleExpression ::=
                  | BaseExpression Application
                  | BaseExpression

BaseExpression  ::=
                  | %LParen Expression %RParen
                  | %LBrack CommaExpList %RBrack
                  | %LiteralInt
                  | %LiteralFlt
                  | %LiteralChar
                  | %LiteralBool
                  | %LiteralStr
                  | %Identifier

Application     ::=
                  | %LParen CommaExpListNE %RParen
                  | %LBrack Expression %RBrack

CommaExpList    ::=
                  | €
                  | CommaExpListNE

CommaExpListNE  ::=
                  | Expression
                  | Expression %Comma CommaExpListNE

```