

Oliver Graf

THE

DROMEDAR

PROGRAMMING LANGUAGE

Contents

1	Introduction	3
2	Typing System	3
3	Expressions	3
4	Handling Whitespace	4
5	Buildup of a Program	4
6	Formal Typing Rules	4
7	Formal Grammar	9

1 Introduction

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 of type **t** come in two shapes: **t**, a strictly non-**null** reference; and **t?**, a reference that may be **null**, where **t** is considered a subtype of **t?**.

Splitting up references that way has two advantages: On one hand, it enforces a good coding style and helps find Null pointer bugs before they even appear; on the other hand it accelerates the execution environment by allowing it to skip Null checks when dereferencing an object if it is of a strictly non-**null** type.

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.	int -> int flt -> flt bool -> bool
!	Logical Negation			
**	Exponentiation	90	right	int,int -> int flt,flt -> flt
*	Multiplication	80	left	int,int -> int flt,flt -> flt
+, -	Addition Subtraction	70	left	int,int -> int flt,flt -> flt char,int -> char int,char -> char
<<, >>, >>>	Left Shift Logical Right Shift Arithmetic Right Shift	60	left	int,int -> int
&	Bitwise And	60	left	int,int -> int
^	Bitwise Xor	50	left	int,int -> int
	Bitwise Or	40	left	int,int -> int
=, !=, >, <, >=, <=	Comparison	30	non-assoc.	[int] -> bool [flt] -> bool
&&	Logical And	20	left	bool,bool -> bool
	Logical Or	10	left	bool,bool -> bool

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.

6 Formal Typing Rules

A typing rule takes the following shape:

$$\frac{\text{Hypotheses}}{S, \dots \vdash \text{grammar spec}}^{\text{NAME}}$$

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

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:

6.1 Builtin Operators

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

6.1.1 Unary Operators

6.1.1.1 Arithmetic Negation

$$\frac{}{\vdash_O - :: \text{int} \rightarrow \text{int}}^{\text{TyUOPNEG}}, \quad \frac{}{\vdash_O - :: \text{flt} \rightarrow \text{flt}}^{\text{TyUOPNEG}}$$

6.1.1.2 Logical Negation

$$\frac{}{\vdash_O ! :: \text{bool} \rightarrow \text{bool}}^{\text{TyUOPNOT}}$$

6.1.2 Binary Operators

6.1.2.1 Power

$$\frac{}{\vdash_O ** :: (\text{int}, \text{int}) \rightarrow \text{int}}^{\text{TyBOPPOW}}, \quad \frac{}{\vdash_O ** :: (\text{flt}, \text{flt}) \rightarrow \text{flt}}^{\text{TyBOPPOW}}$$

6.1.2.2 Multiplication

$$\frac{}{\vdash_O * :: (\text{int}, \text{int}) \rightarrow \text{int}}^{\text{TyBOPMUL}}, \quad \frac{}{\vdash_O * :: (\text{flt}, \text{flt}) \rightarrow \text{flt}}^{\text{TyBOPMUL}}$$

6.1.2.3 Addition and Subtraction

•

$$\frac{}{\vdash_O + :: (\text{int}, \text{int}) \rightarrow \text{int}}^{\text{TyBOPADD}}, \quad \frac{}{\vdash_O + :: (\text{flt}, \text{flt}) \rightarrow \text{flt}}^{\text{TyBOPADD}}, \\ \frac{}{\vdash_O + :: (\text{int}, \text{char}) \rightarrow \text{char}}^{\text{TyBOPADD}}, \quad \frac{}{\vdash_O + :: (\text{char}, \text{int}) \rightarrow \text{char}}^{\text{TyBOPADD}}$$

•

$$\frac{}{\vdash_O - :: (\text{int}, \text{int}) \rightarrow \text{int}}^{\text{TyBOPSUB}}, \quad \frac{}{\vdash_O - :: (\text{flt}, \text{flt}) \rightarrow \text{flt}}^{\text{TyBOPSUB}}, \\ \frac{}{\vdash_O - :: (\text{int}, \text{char}) \rightarrow \text{char}}^{\text{TyBOPSUB}}, \quad \frac{}{\vdash_O - :: (\text{char}, \text{int}) \rightarrow \text{char}}^{\text{TyBOPSUB}}$$

6.1.2.4 Shift Operators

•

$$\frac{}{\vdash_O << :: (\text{int}, \text{int}) \rightarrow \text{int}}^{\text{TyBOPLSHIFT}}$$

•

$$\frac{}{\vdash_O >> :: (\text{int}, \text{int}) \rightarrow \text{int}}^{\text{TyBOPBITRSHIFT}}$$

•

$$\frac{}{\vdash_O >>> :: (\text{int}, \text{int}) \rightarrow \text{int}}^{\text{TyBOPBITAShift}}$$

6.1.2.5 Bitwise Operators

-

$$\frac{}{\vdash_O \& :: (\text{int}, \text{int}) \rightarrow \text{int}}^{\text{TyBOPBITAND}}$$

-

$$\frac{}{\vdash_O \wedge :: (\text{int}, \text{int}) \rightarrow \text{int}}^{\text{TyBOPXOR}}, \quad \frac{}{\vdash_O \wedge :: (\text{bool}, \text{bool}) \rightarrow \text{bool}}^{\text{TyBOPXOR}}$$

-

$$\frac{}{\vdash_O | :: (\text{int}, \text{int}) \rightarrow \text{int}}^{\text{TyBOPBITOR}}$$

6.1.2.6 Logical Operators

-

$$\frac{}{\vdash_O \&\& :: (\text{bool}, \text{bool}) \rightarrow \text{bool}}^{\text{TyBOPLOGAND}}$$

-

$$\frac{}{\vdash_O || :: (\text{bool}, \text{bool}) \rightarrow \text{bool}}^{\text{TyBOPLOGOR}}$$

6.1.2.7 Comparison Operators

$$\frac{}{\vdash_O \{=, !=, >, <, >=, <= \} :: [\text{int}] \rightarrow \text{bool}}^{\text{TyCmpLIST}}, \quad \frac{}{\vdash_O \{=, !=, >, <, >=, <= \} :: [\text{flt}] \rightarrow \text{bool}}^{\text{TyCmpLIST}},$$

$$\frac{}{\vdash_O \{=, !=, >, <, >=, <= \} :: [\text{char}] \rightarrow \text{bool}}^{\text{TyCmpLIST}},$$

6.2 Expressions

6.2.1 Assignability

The rule \vdash_A defines when an expression can be assigned a value.

$$\frac{(\text{id}, \text{t}) \in S}{S \vdash_A \text{id}}^{\text{EXPASSN}}$$

6.2.2 Global Expressions

The rule \vdash_G describes global expressions, which are restricted in a way that they can be computed at compile time.

Copied rules always use \vdash_G instead of \vdash_E

6.2.2.1 Literals

- $\text{GEXPLITINT} := \text{EXPLITINT}$
- $\text{GEXPLITFLT} := \text{EXPLITFLT}$
- $\text{GEXPLITCHAR} := \text{EXPLITCHAR}$
- $\text{GEXPLITBOOL} := \text{EXPLITBOOL}$

6.2.2.2 Other Rules

- $\text{GEXPID} := \text{EXPID}$
- $\text{GEXPUP} := \text{EXPUP}$
- $\text{GEXPBOP} := \text{EXPBOP}$
- $\text{GEXPCMLIST} := \text{EXPCMLIST}$

Note that global variable declarations cannot feature function calls.

6.2.3 Literals

•

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

•

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

•

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

•

$$\frac{}{\vdash_E \mathbf{true} :: \mathbf{bool}} \text{EXPLITBOOL}, \quad \frac{}{\vdash_E \mathbf{true} :: \mathbf{bool}} \text{EXPLITBOOL}$$

6.2.4 Identifiers

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

6.2.5 Unary Operations

$$\frac{S \vdash \text{exp} :: t' \quad S \vdash \text{op} :: t' \rightarrow t}{S \vdash_E \text{op} \text{ exp} :: t} \text{EXPUP}$$

6.2.6 Binary Operations

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

6.2.7 Function Calls

$$\frac{S \vdash f :: (t1, \dots, tn) \rightarrow rt \quad S \vdash e1 :: t1, \dots, S \vdash en :: tn}{S \vdash_E f(e1, \dots, en) :: rt} \text{EXPFUNC}$$

6.2.8 Comparison Lists

$$\frac{S \vdash e0 :: t0, \dots, S \vdash en :: tn \quad S \vdash \text{op1} :: (t0, t1) \rightarrow t, \dots, S \vdash \text{opn} :: (t(n-1), tn) \rightarrow t}{S \vdash_E e0 \text{ op1 } \dots \text{ opn } en :: t} \text{EXPCMPLIST}$$

6.3 Statements

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

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

6.3.1 Local Variable Declarations

$$\frac{S \vdash \text{exp} :: t}{S, rt \vdash_S \text{let } id := \text{exp} \Rightarrow (S \cup (id, t)), \perp} \text{STMTVDECL}$$

6.3.2 Assignments

$$\frac{S \vdash \text{exp} :: t \quad S \vdash \text{lhs} :: t \quad S \vdash_A \text{lhs}}{S, \text{rt} \vdash_S \text{lhs} := \text{exp} \Rightarrow S, \perp} \text{STMTASSN}$$

6.3.3 Expression Statements

$$\frac{S \vdash \text{exp} :: t}{S, \text{rt} \vdash_S \text{exp} \Rightarrow S, \perp} \text{STMTEXPR}$$

6.3.4 If Statements

$$\frac{S \vdash c :: \mathbf{bool} \quad S, \text{rt} \vdash b1 \Rightarrow S_1, R_1 \quad S, \text{rt} \vdash b2 \Rightarrow S_2, R_2}{S, \text{rt} \vdash_S \mathbf{if} \ c \ b1 \ \mathbf{else} \ b2 \Rightarrow S, R_1 \wedge R_2} \text{STMTIF}$$

6.3.5 While Statements

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

6.3.6 Do-While Statements

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

6.3.7 Return Statements

$$\frac{S \vdash \text{exp} :: \text{rt}}{S, \text{rt} \vdash \mathbf{return} \ \text{exp} \Rightarrow S, \top \quad S, \text{rt} \vdash \mathbf{return} \Rightarrow S, \top} \text{STMTRETURN}$$

6.3.8 Blocks

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

6.4 Global Statements

6.4.1 Global Function Declaration

$$\frac{S \sqcup \{(a1, t1), \dots, (an, tn)\}, \text{rt} \vdash b, \top}{S \vdash_G \mathbf{fn} \ id : a1:t1, \dots, an:tn \rightarrow \text{rt} \ b \Rightarrow S} \text{GSTMTFDECL}$$

6.4.2 Global Variable Declaration

$$\frac{(id, any) \notin S \quad S \vdash_G \text{exp} :: t}{S \vdash_G \mathbf{global} \ id := \text{exp} \Rightarrow S \cup \{(id, t)\}} \text{GSTMTVDECL}$$

6.4.3 Program

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

6.5 Context Type-Checking

6.5.1 Global Function Declaration

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

$$\frac{}{S \vdash_{\text{GCF}} \mathbf{let} \text{id} := \text{exp} \Rightarrow S} \text{GSTMTCTXTFDECL}$$

6.5.2 Program: Functions

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

6.5.3 Program

$$\frac{\{\} \sqcup \{\} \vdash_{\text{GCF}} \mathbf{prog} \Rightarrow S' \quad S' \vdash_{\text{G}} \mathbf{prog} \Rightarrow S}{\vdash_{\text{G}} \mathbf{prog} \Rightarrow S} \text{PROG}$$

7 Formal Grammar

7.1 Lexer Grammar

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

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

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

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

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

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

```
return ::= /return/
```

```
Dash ::= /\-/
Bang ::= /!/
Star ::= /\*/
Plus ::= /\+/
LShift ::= /<</
RShift ::= />>/
AShift ::= />>>/
```

```

Bitand      ::= /&/
Xor         ::= /\^/
Bitor       ::= /\|/
Logand      ::= /&&/
Logor       ::= /\|\|/

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

Assign      ::= /:=/

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

Comma       ::= /\,/

LParen      ::= /\(/
RParen      ::= /\)/

```

7.2 Parser Grammar

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

```

Program      ::=
    | €
    | GlobalStatement Program

GlobalStatement ::=
    | GVDeclaration
    | GFDeclaration

GVDeclaration ::=
    | %Global %Identifier %Assign GlobalExpression

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

FArguments   ::=
    | %Identifier %Colon Type
    | %Identifier %Colon Type %Comma FArguments

ReturnType   ::=
    | %Void
    | Type

Block        ::=
    | €
    | Statement Block

Statement    ::=
    | VDeclaration
    | AssignStmt
    | IfStmt
    | WhileStmt
    | DoWhileStmt
    | ExprStmt

```

	ReturnStmt
VDeclaration	::= %Let %Identifier %Assign Expression
AssignStmt	::= LHS %Assign Expression
IfStmt	::= %If Expression Block ElifStmt
ElifStmt	::= %Elif Expression Block ElifStmt %Else Block €
WhileStmt	::= %While Expression Block
DoWhileStmt	::= %Do Block %While Expression
ExprStmt	::= Expression Expression %Colon FArgs
ReturnStmt	::= %Return %Return Expression
LHS	::= %Identifier
Type	::= %Int %Flt %Char %Bool
GlobalExpression	::= Expression*
Expression	::= ExprPrec30
ExprPrec20	::= ExprPrec30 ExprPrec30 %Equal ExprPrec20 ExprPrec30 %NotEqual ExprPrec20 ExprPrec30 %Greater ExprPrec20 ExprPrec30 %Less ExprPrec20 ExprPrec30 %GreaterEq ExprPrec20 ExprPrec30 %LessEq 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 %StarStar ExprPrec90
ExprPrec100 ::=
| SimpleExpression
| %Dash ExprPrec100
| %Bang ExprPrec100
SimpleExpression ::=
| %LParen Expression %RParen
| %LiteralInt
| %LiteralFlt
| %LiteralChar
| %LiteralBool
| %Identifier
| %Identifier Application
Application ::=
| %LParen FArgs %RParen
FArgs ::=
| €
| NonEmptyFArgs
NonEmptyFArgs ::=
| Expression
| Expression %Comma NonEmptyFArgs

```