

CA4003 Compiler Construction

Assignment

Language Definition

David Sinclair

2020-2021

1 Overview

The language is not case sensitive. A nonterminal, X , is represented by enclosing it in angle brackets, e.g. $\langle X \rangle$. A terminal is represented without angle brackets. A **bold typeface** is used to represent terminal symbols in the language and reserved words, whereas a non-bold typeface is used for symbols that are used to group terminals and nonterminals together. Source code should be kept in files with the .cal extension, e.g. hello_world.cal .

2 Syntax

The reserved words in the language are **variable**, **constant**, **return**, **integer**, **boolean**, **void**, **main**, **if**, **else**, **true**, **false**, **while**, **begin**, **end**, **is** and **skip**.

The following are tokens in the language:

$, ; : := () + - \sim | \& = != < <= > >=$

Integers are represented by a string of one or more digits ('0'-'9') and may start with a minus sign ('-'), e.g. 123, -456. Unless it is the number '0', numbers may not start with leading '0's. For example, 0012 is illegal.

Identifiers are represented by a string of letters, digits or underscore character ('_') beginning with a letter. Identifiers cannot be reserved words.

Comments can appear between any two tokens. There are two forms of comment: one is delimited by $/^*$ and $*/$ and can be nested; the other begins with $//$ and is delimited by the end of line and this type of comments may not be nested.

$$\begin{aligned}
\langle \text{program} \rangle &\models \langle \text{decl_list} \rangle \langle \text{function_list} \rangle \langle \text{main} \rangle & (1) \\
\langle \text{decl_list} \rangle &\models (\langle \text{decl} \rangle ; \langle \text{decl_list} \rangle \mid \epsilon) & (2) \\
\langle \text{decl} \rangle &\models \langle \text{var_decl} \rangle \mid \langle \text{const_decl} \rangle & (3) \\
\langle \text{var_decl} \rangle &\models \mathbf{variable\ identifier} : \langle \text{type} \rangle & (4) \\
\langle \text{const_decl} \rangle &\models \mathbf{constant\ identifier} : \langle \text{type} \rangle := \langle \text{expression} \rangle & (5) \\
\langle \text{function_list} \rangle &\models (\langle \text{function} \rangle \langle \text{function_list} \rangle \mid \epsilon) & (6) \\
\langle \text{function} \rangle &\models \langle \text{type} \rangle \mathbf{identifier} (\langle \text{parameter_list} \rangle) \mathbf{is} & (7) \\
&\quad \langle \text{decl_list} \rangle \\
&\quad \mathbf{begin} \\
&\quad \langle \text{statement_block} \rangle \\
&\quad \mathbf{return} (\langle \text{expression} \rangle \mid \epsilon) ; \\
&\quad \mathbf{end} \\
\langle \text{type} \rangle &\models \mathbf{integer} \mid \mathbf{boolean} \mid \mathbf{void} & (8) \\
\langle \text{parameter_list} \rangle &\models \langle \text{nemp_parameter_list} \rangle \mid \epsilon & (9) \\
\langle \text{nemp_parameter_list} \rangle &\models \mathbf{identifier} : \langle \text{type} \rangle \mid \mathbf{identifier} : \langle \text{type} \rangle , \langle \text{nemp_parameter_list} \rangle & (10) \\
\langle \text{main} \rangle &\models \mathbf{main} & (10) \\
&\quad \mathbf{begin} \\
&\quad \langle \text{decl_list} \rangle \\
&\quad \langle \text{statement_block} \rangle \\
&\quad \mathbf{end} \\
\langle \text{statement_block} \rangle &\models (\langle \text{statement} \rangle \langle \text{statement_block} \rangle) \mid \epsilon & (11) \\
\langle \text{statement} \rangle &\models \mathbf{identifier} := \langle \text{expression} \rangle ; \mid & (12) \\
&\quad \mathbf{identifier} (\langle \text{arg_list} \rangle) ; \mid \\
&\quad \mathbf{begin} \langle \text{statement_block} \rangle \mathbf{end} \mid \\
&\quad \mathbf{if} \langle \text{condition} \rangle \mathbf{begin} \langle \text{statement_block} \rangle \mathbf{end} \\
&\quad \mathbf{else\ begin} \langle \text{statement_block} \rangle \mathbf{end} \mid \\
&\quad \mathbf{while} \langle \text{condition} \rangle \mathbf{begin} \langle \text{statement_block} \rangle \mathbf{end} \mid \\
&\quad \mathbf{skip} ; \\
\langle \text{expression} \rangle &\models \langle \text{fragment} \rangle \langle \text{binary_arith_op} \rangle \langle \text{fragment} \rangle \mid & (13) \\
&\quad (\langle \text{expression} \rangle) \mid \\
&\quad \mathbf{identifier} (\langle \text{arg_list} \rangle) \mid
\end{aligned}$$

$$\langle \text{binary_arith_op} \rangle \models + \mid - \quad (14)$$

$$\langle \text{fragment} \rangle \models \text{identifier} \mid - \text{identifier} \mid \text{number} \mid \text{true} \mid \text{false} \mid \langle \text{expression} \rangle \quad (15)$$

$$\langle \text{condition} \rangle \models \sim \langle \text{condition} \rangle \mid (\langle \text{condition} \rangle) \mid \quad (16)$$

$$\langle \text{comp_op} \rangle \models = \mid != \mid < \mid <= \mid > \mid >= \quad (17)$$

$$\langle \text{arg_list} \rangle \models \langle \text{nemp_arg_list} \rangle \mid \epsilon \quad (18)$$

$$\langle \text{nemp_arg_list} \rangle \models \text{identifier} \mid \text{identifier}, \langle \text{nemp_arg_list} \rangle \quad (19)$$

3 Semantics

Declaration made outside a function (including **main**) are global in scope. Declarations inside a function are local in scope to that function. Function arguments are *passed-by-value*. Variables or constants cannot be declared using the **void** type. The **skip** statement does nothing.

The operators in the language are:

Operator	Arity	Description
<code>:=</code>	binary	assignment
<code>+</code>	binary	arithmetic addition
<code>-</code>	binary	arithmetic subtraction
<code>-</code>	unary	arithmetic negation
<code>~</code>	unary	logical negation
<code> </code>	binary	logical disjunction (logical or)
<code>&</code>	binary	logical conjunction (logical and)
<code>=</code>	binary	is equal to (arithmetic and logical)
<code>!=</code>	binary	is not equal to (arithmetic and logical)
<code><</code>	binary	is less than (arithmetic)
<code><=</code>	binary	is less than or equal to (arithmetic)
<code>></code>	binary	is greater than (arithmetic)
<code>>=</code>	binary	is greater than or equal to (arithmetic)

The following table gives the precedence (from highest to lowest) and associativity of these operators.

Operator(s)	Associativity	Notes
~	right to left	logical negation
-	right to left	arithmetic negation
+ -	left to right	addition & subtraction
< <= > >=	left to right	arithmetic comparison operators
= !=	left to right	equality & inequality operators
&	left to right	logical conjunction
	left to right	logical disjunction
:=	right to left	assignment

4 Examples

Three versions of the simplest non-empty file demonstrating that the language is case insensitive.

main	Main	MAIN
begin	begin	begin
end	eND	end

A simple file demonstrating comments.

```
main
begin
  // a simple comment
  /* a comment /* with /* several */ nested */ comments */
end
```

The simplest program that uses functions.

```
void func () is
begin
  return ();
end
```

```

main
begin
  func ();
end

```

A simple file demonstrating the different scopes.

```

variable i:integer;

integer test_fn (x:integer) is
  variable i:integer;
begin
  i := 2;
  return (x);
end

```

```

main
begin
  variable i:integer;

  i := 1;
  i := test_fn (i);
end

```

A file demonstrating the use of functions.

```

integer multiply (x:integer , y:integer) is
  variable result:integer;
  variable minus_sign : boolean;
begin
  // figure out sign of result and convert args to absolute values

  if (x < 0 & y >= 0)
  begin
    minus_sign := true;
    x := -x;
  end
  else
  begin
    if y < 0 & x >= 0

```

```

begin
    minus_sign := true;
    y := -y;
end
else
begin
    if (x < 0) & y < 0
    begin
        minus_sign := false;
        x := -x;
        y := -y;
    end
    else
    begin
        minus_sign := false;
    end
    end
end
end

result := 0;

while (y > 0)
begin
    result := result + x;
    y := y - 1;
end

if minus_sign = true
begin
    result := -result;
end
else
begin
    skip;
end

return (result);
end

main

```

```
begin
  variable arg_1:integer;
  variable arg_2:integer;
  variable result:integer;
  constant five:integer := 5;

  arg_1 := -6;
  arg_2 := five;

  result := multiply (arg_1 , arg_2);
end
```