**­­­­**

PL/0 User’s Guide

Authors: Travis Le and Allysia Freeman

Table of Contents

Table of Contents 2

Index of Figures and Tables 3

1.0: How to Program with PL/0 4

1.1 Data Types 4

1.2 Operators 5

1.3 Expressions 6

1.4 Statements 7

1.5 Procedures 8

2.0: Build, Compile, and Execute PL/0 programs 9

2.1 Building the Compiler 9

2.2 Executing the Compiler 9

3.0: EBNF Grammar of PL/0 11

4.0: Tokens and Reserved Words 12

5.0: Machine Code Instructions in PL/0 14

5.1.1 **The Stack** 14

5.1.2 **Machine Code Meanings** 14

6.0: Error Codes 15

Index of Figures and Tables

[Figure 1: An Example of a Stack 14](#_Toc424503019)

[Figure 2: A Stack with Three Activation Records 15](#_Toc424503020)

[Table 1: Extended Backus–Naur Form (EBNF) for PL/0 11](#_Toc424503015)

[Table 2: Tokens and Reserved Words f PL/0 13](#_Toc424503016)

[Table 3: Machine Code Instructions for PL/0 15](#_Toc424503017)

[Table 4: Error Codes 17](#_Toc424503018)

# **1.0: How to Program with PL/0**

* 1. Data Types
     1. **Integer**
     2. **Variable (local and global)**
     3. **Constant**
  2. Operators
     1. **Assignment**
     2. **Relational Operators**
     3. **Mathematical Operators**
  3. Expressions
  4. Statements
     1. **Assignments**
     2. **Read and Write**
     3. **While Do**
     4. **If, Then, Else**
  5. Procedures
     1. **Declaration**
     2. **Calling**
     3. **Recursion**
     4. **Nested Procedures**

**2.0: Build, Compile, and Execute PL/0 programs**

The Compiler is composed of the following files:

1. **compiler.c**

The main driver of the program, this handles all the other files and executes them as needed.

1. **header.h**

This holds all the declarations used often throughout all files. It also includes a section for redefining filenames. If you don’t know how to use #define in C programming, please don’t change these names.

1. **scanner.h**

This scans your input file and generates a list of lexemes (kinds of symbols).

1. **parser.h**

This takes the result from scanner.h and analyzes it for errors. It then generates machine code for the given input file. If it is found to be correct, it sends the generated machine code to vm.h

1. **vm.h**

This takes the machine code from parser.h and executes it.

1. **input.pl0**

This is your PL/0 code. It must be named this exactly, including the file extension .pl0. If you would like to change the input filename, edit *nameCode* in header.h, making sure to include the file’s extension. Complex file extensions are not supported, and it is strongly suggested to use either a plaintext or simple code editor to write your PL/0 code.

With the following output files, which are all used/generated internally:

1. **cleaninput.txt**

This is your PL/0 cleaned of all comments.

1. **lexemetable.txt**

This is a user friendly version of the lexeme list. To see all possible tokens and their meanings, look at section 4.

1. **lexemelist.txt**

The not so user friendly version of the lexeme table. This is used internally for analysis and machine code generation.

1. **mcode.txt**

The machine code that was generated.

1. **stacktrace.txt**

Includes the interpreted machine code, which is used internally, and a stack trace of when the code was executed. See section 5 for an explanation of the stack.

1. 1. Building the Compiler

It is assumed you are using a linuix-based terminal for all following instructions.

To build this compiler:

1. Make sure all files listed above are in the same directory, and named exactly as shown.
2. Open your terminal.
3. Run the command “**gcc compiler.c –o compile**”

You should now have a file called “**compiler.o**” in your directory, and are ready to execute.

* 1. Executing the Compiler

If you have closed your terminal, please re-open it in order to execute the compiler.

In order to execute the compiler, run the command “**./compile**”

Alternatively, the following flags are available for your use:

* **–l** which instructs the compiler to display the lexeme list to the screen. This is the internal interpretation of the input file.
* **–a** which instructs the compiler to display the generated machine code to the screen. The meaning of these codes and their properties are listed in section five.
* **–v** whichinstructs the compiler to display a stack trace for the execution of the input file. The stack is also explained in section five.

To use them, simply enter the flag after the execute command, making sure to put space between all flags. For example, in order to see the lexeme list, enter “**./compile -l**”. To see both the lexeme list and the stack trace, enter “**./compile –l -v**”, and in order to see all of them, enter “**./compile –l –a -v**”. They can be entered in any order.

Provided your PL/0 program is syntactically correct and can execute properly, the compiler will generate the output files described in the previous subsection.

And, if there are any read or write statements in your code, the console will either prompt you or print a value, respectively. Please enter a reasonable number and then press enter for read statements.

**3.0: EBNF Grammar of PL/0**

|  |
| --- |
| **program** = block "**.**" **.**  **block** = const-declaration var-declaration procedure-declaration statement **.**  **const-declaration** = [“**const**” ident “**=**” number {“**,**” ident “**=**” number} “**;**”] **.**  **var-declaration** = [“**var**” ident {“**,**” ident} “**;**” ] **.**  **procedure-declaration** = {“**procedure**” ident “**;**” block “**;**”} **.**  **statement** = [ ident "**:=**" expression  | "**call**" ident  | "**begin**" statement {"**;**" statement } "**end**"  | "**if**" condition "**then**" statement  | "**while**" condition "**do**" statement  | "**read**" ident  | "**write**" ident] **.**  **condition** = "**odd**" expression  | expression rel-op expression **.**  **rel-op** ="**=**"|"**<>**"|"**<**"|"**<=**"|"**>**"|"**>=**" **.**  **expression** = [ "**+**"|"**-**"] term { ("**+**"|"**-**") term} **.**  **term** = factor { ("**\***" | "**/**") factor} **.**  **factor** = ident | number | "**(**" expression "**)**" **.**  **number** = digit {digit} **.**  **ident** = letter {letter |digit} **.**  **digit** = "**0**"|"**1**"|"**2**"|"**3**"|"**4**"|"**5**"|"**6**"|"**7**"|"**8**"|"**9**" **.**  **letter** = "**a**"|"**b**"|"**c**"|**...** |"**x**"|"**y**"|"**z**" **.** |

Table : Extended Backus–Naur Form (EBNF) for PL/0

The first word of each section is the name of a “syntactic class.”

This is read using the following rules:

1. **|** means “or, but not both”
2. **[ ]** means an optional item
3. **{ }** means repeat zero or more times
4. Special symbols are enclosed in quote marks
5. A period is used to indicate the end of the definition of a syntactic class

**4.0: Tokens and Reserved Words**

|  |  |  |  |
| --- | --- | --- | --- |
| **Symbol** | **Name** | **Value** | **Usage** |
|  | nulsym |  | n/a |
|  | identsym |  | used for constant, procedure, and variable names |
|  | numbersym |  | used for numbers |
| **+** | plussym |  | adds |
| **-** | minussym |  | subtracts |
| **\*** | multsym |  | multiplies |
| **/** | slashsym |  | divides |
| **odd** | oddsym |  | test if an expression is odd |
| **=** | eqlsym |  | constant definition or check is two expressions are equal |
| **<>** | neqsym |  | test that two expressions do not equal eachother |
| **<** | lessym |  | tests that left expression is less than right expression |
| **<=** | leqsym |  | tests that left expression is less than or equal to right expression |
| **>** | gtrsym |  | tests that left expression is greater than right expression |
| **>=** | geqsym |  | tests that left expression is greater than or equal to right expression |
| **(** | lparentsym |  | begin factor |
| **)** | rparentsym |  | end factor |
| **,** | commasym |  | separates identifiers in declarations |
| **;** | semicolonsym |  | ends a statement |
| **.** | periodsym |  | ends the program |
| **:=** | becomessym |  | assigns a value to a variable |
| **begin** | beginsym |  | begins a block of statements |
| **end** | endsym |  | ends a block of statements |
| **if** | ifsym |  | begins if statement, followed by condition |
| **then** | thensym |  | follows then, followed by statement |
| **while** | whilesym |  | begins while loop, followed by condition |
| **do** | dosym |  | follows while, followed by statement |
| **call** | callsym |  | calls a procedure |
| **const** | constsym |  | begins a constant declaration |
| **var** | varsym |  | begins a variable declaration |
| **procedure** | procsym |  | begins a procedure declaration |
| **write** | writesym |  | prints a value to the screen |
| **read** | readsym |  | asks the user to enter a value |
| **else** | elsesym |  | may follow if statements |

Table : Tokens and Reserved Words f PL/0

**5.0: Machine Code Instructions in PL/0**

* + 1. **The Stack**

This will be a basic explanation of the stack, which appears both in the output file stacktrace.txt and if you enter the flag “**-v**” when executing the compiler. A stack is utilized internally in order to properly execute the input code.

A **stack** is defined as a set of data sections in which data can be stored during execution. Everything is stored in it, including variable values and various results of mathematical operations.

The **lexical level** (referred to as “level”) is the number of activation records minus 1, since it starts at zero. The first is level 0, the second is level 1, and so on.

An **activation record**, or **AR**, is defined as a section of the stack used for the main code of a block. Each procedure gains its own activation record when called, and though they cannot be shared, they can be accessed by any level, at any level.

|  |
| --- |
| pc bp sp stack  10 SIO 0 1 11 1 7 0 0 0 0 0 4 2  11 STO 0 4 12 1 6 0 0 0 0 2 4 |

Figure : An Example of a Stack

We are going to focus on the section of numbers at the end, the representation of the stack, after the 7 and 6 for both lines respectively. However, here are some quick definitions for the other numbers, by position:

1. Line of code being executed
2. The operation being executed, see section 5.1.2 for meanings
3. The level for the operation being executed
4. The value or line for the operation being executed
5. The line number of the next line to be executed, or *program counter* (pc)
6. The position of the bottom of the topmost activation record, or *base pointer* (bp)
7. The number of elements in all activation records, or *stack pointer* (sp)

The first four spaces (shown above as three zeros) are reserved, and their meanings are:

1. Reserved.
2. The *static link*, or the positon of the activation record that is its most senior parent. Unless a special case, this will most likely have a value of one in any record after the first.
3. The *dynamic link*, or the position of the previous activation record.
4. The *return address*. This is the line number of the code to execute after a procedure is returned.

Every position after these zeroes are used for variables, and since constants cannot be changed and procedure names do not hold values, they do not get a position (though they *are* acknowledged during execution). Variables are stored in the activation record where they were declared. We see that there are two positions after the zeros in the figure, and also a mysterious 2.

In the first line we see the result of a read, the 2, and in the second line we see the read’s value being stored into whatever variable was asked to be read, position 1. The 4 is already stored in another variable’s value, in position 2.

|  |
| --- |
| pc bp sp stack  21 INC 0 5 22 13 17 0 0 0 0 6 9 | 0 1 1 191 2 1 | 0 1 7 111 0  22 SIO 0 1 23 13 18 0 0 0 0 6 9 | 0 1 1 191 2 1 | 0 1 7 111 0 2 |

Figure : A Stack with Three Activation Records

Here we can see these things in action. Each “|” denotes a separation between activation records. This is not used internally, and is only displayed in order to make stacks easier to read.

We can see that the second record is pointing to the first with its values of “1 1 191” and has two new variables, “2 1”. It will execute line number 191 when returned. The third record is pointing to the second with its values of “1 7 111”, with one new variable, “0”. It will execute line number 111 when returned.

* + 1. **Machine Code Meanings**

|  |  |  |
| --- | --- | --- |
| **1** |  |  |
| **2** |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| **3** |  |  |
| **4** |  |  |
| **5** |  |  |
| **6** |  |  |
| **7** |  |  |
| **8** |  |  |
| **9** |  |  |

Table : Machine Code Instructions for PL/0

**6.0: Error Codes**

|  |  |  |
| --- | --- | --- |
| Error Number | Error Message | Explanation |
| **0** | Program is syntactically correct. | n/a |
| **1** | Invalid file input | Make sure the input file is named correctly or edit header.h's filename definitions |
| **2** | Use “=” not “:=” | Don’t use := when you’re not assigning a value to a variable. |
| **3** | Use “:=” not “=” | Always use := when assigning a value to a variable. |
| **4** | “=” expected after const declaration | When declaring a constant, it must be given a value. |
| **5** | Number expected after “=” with const | Constant declarations must be followed by =. |
| **6** | "then" expected after "if" | “then” must follow after an “if” statement. |
| **7** | "do" expected after "while" | “do” must follow after a “while” statement. |
| **8** | const, var, and procedure must be followed by an identifier | You must give a name to every constant, varable, and procedure when declaring it. |
| **9** | “:=” expected after identifier | Missing “:=” after a variable name. |
| **10** | Ident expected after “call” | The proper syntax is “call name;” |
| **11** | Relational operator expected | Missing a relational test. |
| **12** | Assignment to constants and procedures not allowed | You cannot assign a value to procedure names or constants. They are static. |
| **13** | Semicolon needed between statements | Missing a semicolon after a statement. |
| **14** | Cannot begin statement with this symbol | Check your program’s syntax. |
| **15** | Undeclared variable detected | Declare the variable before using it. |
| **16** | Unclosed parenthesis detected | All parenthesis must be closed. |
| **17** | Invalid operator | See operators section for correct implementation. |
| **18** | Invalid symbol | The symbol is not supported by PL/0. |
| **19** | ";" expected | Missed a “;” in code |
| **20** | Number too long | A number is too long. Make it shorter. |
| **21** | Identifier too long | The name of a variable is too long. Make it shorter. |
| **22** | Generated code too long | Your program is too complicated for the compiler. Shorten or simplify your code. |
| **23** | Compiler has run out of memory | n/a |
| **24** | Period expected. | A period must be at the end of the program. |
| **25** | Var or const detected more than once | You can only declare variables and constants once per level. Merge the declarations into one. |

Table : Error Codes