

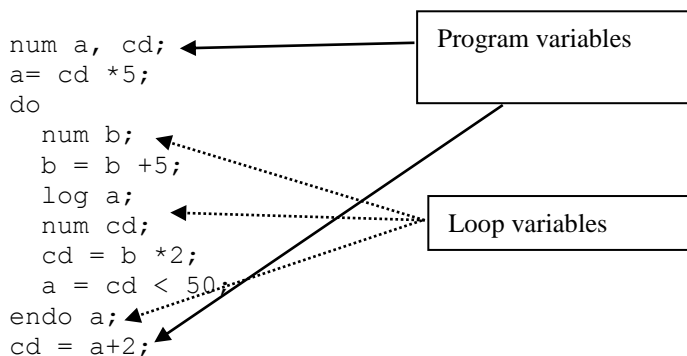
UNITS 2, 3 AND 4: LEXICAL ANALYSIS AND GRAMMAR DESIGN FOR THE SYNTAX ANALYSIS

Build a compiler for the following programming language. There are three types of statements: declaration, arithmetic / logical expression and loop and two types of data: numeric and logical. The sentences are described as follows:

- Declaration
 - **type name_variable [, name_variable]* ;**
 - where type can be **num** or **log** and **name_variable** is a char string with a maximum number of 8 characters
- Arithmetic/Logic expressions
 - **name_variable = expression_arithmetic ;**
 - **name_variable = expression_logic ;**
 - Arithmetic expressions can contain variables of type **num** and numbers with the operators: +, -, *, /
 - The logical expression relates variables of type **num** with numbers through logical operators: <, >, =, #. The possible results of the evaluation of the logical expression is V or F. The variable to which it is assigned must be type **log**.
- Loop
 - **do [sentence]+ endo [expression_logic | variable_logic] ;**
 - where sentences can be declaration, expression or loop.

The sentences of the loop are executed at least once, and the loop is repeated while the logical expression in endo takes the value V. The variables are local to the loop where they are declared, if variables are used in the logical expression of the endo then there must be been declared inside the loop. The variables declared in the main body of the program are considered local to it.

Example of a program without errors:



It is required:

1. Define the G grammar that would generate valid sentences of this programming language.

Solution

A grammar that generates the given language can be defined as follows:

$$G = \{S, A, B, D, E, V, X\}, \{S\}, \{ ;, \text{type, var, do, endo, =, o, p, " ;" }\}$$

- (1) $S ::= A ; S$
- (2) $S ::= A ;$
- (3) $A ::= D$
- (4) $A ::= B$
- (5) $A ::= E$
- (6) $D ::= \text{type } V$
- (7) $V ::= \text{var}$
- (8) $V ::= \text{var } , V$
- (9) $B ::= \text{do } S \text{ endo } X$
- (10) $E ::= \text{var } = X$
- (11) $X ::= o$
- (12) $X ::= o p X$

The token "type" represents the strings "num" and "log", the token "var" to the set of characters that identifies a variable, "o" is an operand (variable or number) and "p" is an operator, either of arithmetic or logical type.

After left-factoring, the modified G' is:

$$G' = \{S, S', A, B, D, E, V, V', X, X'\}, \{S\}, \{ ;, \text{type, var, do, endo, =, o, p, " ;" }\}$$

- (1) $S ::= A ; S'$
- (2) $S' ::= S$
- (3) $S' ::= \lambda$
- (4) $A ::= D$
- (5) $A ::= B$
- (6) $A ::= E$
- (7) $D ::= \text{type } V$
- (8) $V ::= \text{var } V'$
- (9) $V' ::= , V$
- (10) $V' ::= \lambda$
- (11) $B ::= \text{do } S \text{ endo } X$
- (12) $E ::= \text{var } = X$
- (13) $X ::= o X'$
- (14) $X' ::= p X$
- (15) $X' ::= \lambda$