

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

There is a very simple programming language oriented to the arithmetic calculation of a calculator. In this language, programs consist of a sequence of expressions (there may be any expression). Valid expressions are sequences of operators and numbers ending with the = sign. An example of a valid program is: (only operations of +, - and * are valid)

$$+ * + 6 8 6 9 =$$

$$- + * - 45 23 2 5 2 =$$

There is no operator precedence. The expression is read from left to right. The result of the compiling process is shown on the screen and consists of the transformation of operations into equivalent sums and subtracts (the same operation, but without multiplications) and the result, following the example, the output on screen would show:

$$6 + 8 + 6 + 8 + 6 + 8 + 6 + 8 + 6 + 8 + 6 + 8 + 9 = 93$$

$$45 - 23 + 45 - 23 + 5 - 2 = 47$$

Numbers are positive integer.

It is required:

1. Define the grammar G that would generate valid statements of this programming language and the lexical analyzer.

Note: Note that there are no quartets for multiplication, they have to be implemented with sums.

Instruction	Meaning
(move, pos_1 , pos_2)	$pos_2 \leftarrow pos_1$
(push, pos_1 , ,)	incorporates the contents of pos_1 into the Stack
(pop, , , pos_1)	$pos_1 \leftarrow$ top of the Stack
(label, , , $label$)	defines a $label$
(goto, , , $label$)	go to a $label$
(return, , , reg)	go to the address in reg
(if, reg , , $label$)	go to $label$ if the content of reg is -1
(<, reg , , $label$)	go to $label$ if the content of reg is lower or equal to 0
(+, reg_1 , reg_2 , reg)	$reg \leftarrow reg_1 + reg_2$
(-, reg_1 , reg_2 , reg)	$reg \leftarrow reg_1 - reg_2$

Solution:

A grammar that generates the language of the problem is defined as follows:

$G = \{S, C, Z, E, E', O, O', U, V, T, W\}, \{ \grave{_}, ?, (,), =>, /=>, ->, \text{Id}, \text{Num}, +, -, /, *, ; \}, \{S\}$

- (1) $S ::= C S$
- (2) $S ::= E S$
- (3) $S ::= \lambda$
- (4) $E ::= E' \rightarrow V$
- (5) $E' ::= O U$
- (6) $O ::= \text{Id}$
- (7) $O ::= \text{Num}$
- (8) $U ::= O' E'$
- (9) $U ::= \lambda$
- (10) $O' ::= +$
- (11) $O' ::= -$
- (12) $O' ::= *$
- (13) $O' ::= /$
- (14) $V ::= \text{Id } T$
- (15) $T ::= \lambda$
- (16) $T ::= ; V$
- (17) $Z ::= E$
- (18) $Z ::= C$
- (19) $C ::= \grave{_}(E') \Rightarrow Z W$
- (20) $W ::= ?$
- (21) $W ::= /=> Z ?$