



Tecnológico  
de Monterrey

---

# Instituto Tecnológico y de Estudios Superiores de Monterrey

Desarrollo de Aplicaciones Avanzadas

José Manuel Castillo Guajardo  
A01284149

## Profesores

Elda Quiroga

## Entregable

Entrega #1: Little Duck

## Fecha de entrega

10 / mayo / 2024

## Gramática

<TYPE> → int  
<TYPE> → float

<list\_ids> → id<mas\_ids>  
<mas\_ids> → ,<list\_ids>  
<mas\_ids> → ε

<VARS> → var <variables> : <TYPE> ; <mas\_var>  
<variables> → <list\_ids> :  
<mas\_var> → ε  
<mas\_var> → <variables>

<Programa> → program id ; <dec\_vars> <dec\_f> main <Body> end  
<dec\_v> → ε  
<dev\_v> → <VARS>  
<dec\_f> → ε  
<dec\_f> → <FUNCS> <mas\_f>  
<mas\_f> → ε  
<mas\_f> → <FUNCS> <mas\_f>

<Body> → { <mas\_statement> }  
<mas\_statement> → ε  
<mas\_statement> → <statement><mas\_statement>

<STATEMENT> → <ASSIGN>  
<STATEMENT> → <CONDITION>  
<STATEMENT> → <CYCLE>  
<STATEMENT> → <F\_CALL>  
<STATEMENT> → <PRINT>

<ASSIGN> → id = <EXPRESION>;

<EXPRESION> → <EXP> <mas\_expresion>  
<mas\_expresion> → > <EXP>  
<mas\_expresion> → < <EXP>  
<mas\_expresion> → != <EXP>  
<mas\_expresion> → ε

<EXP> → <TERMINO> <mas\_exp>  
<mas\_exp> → + <TERMINO>  
<mas\_exp> → - <TERMINO>  
<mas\_exp> → ε

<TERMINO> → <FACTOR> <mas\_termino>  
<mas\_termino> → \* <FACTOR>  
<mas\_termino> → / <FACTOR>

<mas\_termino> → ε

<CTE> → cte\_int

<CTE> → cte\_float

<PRINT> → print <EXPRESION> <mas\_print> );

<mas\_print> → , <EXPRESION> <mas\_print>

<mas\_print> → cte.string <mas\_print>

<mas\_print> → ε

<CYCLE> → while <BODY> do ( <EXPRESION> );

<CONDITION> → if ( <EXPRESION> ) <BODY> <mas\_condition> ;

<mas\_condition> → ε

<mas\_condition> else <BODY>

<F\_Call> → id ( <EXPRESION> <mas\_expresion> );

<mas\_expresion> → , <EXPRESION>

<mas\_expresion> → ε

<FUNCS> → void id ( <PARAMS> ) [ <VARS> <BODY> ] ;

<params> → , id : <TYPE> <params>

<params> → , id : <TYPE> <mas\_params>

<params> → ε

<mas\_params> → , <params>

<FACTOR> → <mas\_factor>

<mas\_factor> → ( <EXPRESION> )

<mas\_factor> → <signo> id

<mas\_factor> → <signo> <CTE>

<mas\_factor> → + <CTE>

<mas\_factor> → - <CTE>

<signo> → +

<signo> → -

<signo> → ε

## PLY (Python Lex-Yacc)

PLY es una implementación en Python de las herramientas lex y yacc que son comúnmente utilizadas para escribir parsers y compiladores. El parsing está basado en el algoritmo LALR utilizado en muchas herramientas yacc. PLY tiene un reporte y diagnóstico de errores robusto que ayuda a en la construcción efectiva del parser. Tiene soporte para lenguajes con cientos de reglas, como lo podría ser un lenguaje como C.

Para la realización de este proyecto se tomó en cuenta PLY y ANTLR para el desarrollo del scanner y el parser. Se decidió utilizar PLY porque es una librería que fue creada específicamente para el curso de introducción a los compiladores de la Universidad de Chicago en el 2001. Tiene únicamente las cualidades del core de LALR, esto ofreciendo una mayor rigidez en cuanto al diseño de la gramática. Esto ayuda a los estudiantes a aprender de manera más profunda el cómo verdaderamente funciona un compilador.

La documentación se puede encontrar en el siguiente enlace:

<https://ply.readthedocs.io/en/latest/>

## Gramática en PLY

```
# Palabras reservadas
reserved = {
    'if'      : 'IF',
    'else'    : 'ELSE',
    'program' : 'PROGRAM',
    'main'    : 'MAIN',
    'end'     : 'END',
    'var'     : 'VAR',
    'int'     : 'INT',
    'float'   : 'FLOAT',
    'print'   : 'PRINT',
    'void'    : 'VOID',
    'while'   : 'WHILE',
    'do'      : 'DO',
    'if'      : 'IF',
    'else'    : 'ELSE',
}

# Lista de tokens
tokens = [
    'ID',
    'SEMI_COLON',
    'OPEN_PAR',
    'CLOSE_PAR',
    'DIVIDE',
    'MULTIPLY',
    'ADD',
    'SUBTRACT',
    'LESS_THAN',
    'MORE_THAN',
    'NOT_EQUAL',
    'COLON',
    'L_BRACKET',
    'R_BRACKET',
    'LEFT_CURLY',
    'RIGHT_CURLY',
    'COMMA',
    'EQUAL',
    'CTE_STRING',
    'CTE_FLOAT',
    'CTE_INT',
] + list(reserved.values())
```

Las palabras reservadas y los tokens se definen en el lenguaje como se aprecia en las figuras. Definiendo en un diccionario cada palabra reservada y en una lista los tokens que se utilizan para el lenguaje.

```
# Define reglas para cada token
t_SEMI_COLON = r';'
t_OPEN_PAR = r'\('
t_CLOSE_PAR = r'\)'
t_DIVIDE = r'/'
t_MULTIPLY = r'\*'
t_ADD = r'\+'
t_SUBTRACT = r'\-'
t_LESS_THAN = r'<'
t_MORE_THAN = r'>'
t_NOT_EQUAL = r'!='
t_COLON = r':'
t_L_BRACKET = r'\['
t_R_BRACKET = r'\]'
t_LEFT_CURLY = r'\{'
t_RIGHT_CURLY = r'\}'
t_COMMA = r','
t_EQUAL = r'='
```

Después de esto se tienen que definir las reglas de expresión regular para cada uno de los tokens. De esta manera el scanner los puede identificar.

```
def p_programa(p):
    """programa : PROGRAM ID SEMI_COLON dec_v dec_f MAIN body END
    """

def p_vars(p):
    """vars : VAR variables COLON type SEMI_COLON mas_var
    | empty"""

def p_variables(p):
    """variables : list_ids
    """

def p_mas_var(p):
    """mas_var : variables
    | empty"""
```

En PLY cada regla tiene su propia función en la que se escribe la gramática de la misma. Como se observa p\_programa sería el equivalente a <Programa> → program id ; <dec\_vars> <dec\_f> main <Body> end. En esta las palabras en mayúscula se refieren a los tokens o palabras reservadas. Las palabras en minúscula se refiere a otras reglas gramaticales (otra función). En este caso también el uso de “|” indica un “or” en la regla gramatical. La palabra “empty” sería el equivalente al  $\epsilon$ .

## Test Plan

El test plan que se desarrolló es el siguiente, dónde se prueba el scanner y el parser desarrollado, probando las reglas gramaticales definidas.

```
program HelloWorld;

var i : int;

void myFunc(param1 : float) [
    var x : int;
    {
        x = 0;
        do {
            x = x + 1;
            print(x);
        } while (x < 6);
    }
];

void myFunc2(param1 : float) [
    {
        if (param1 > 2) {
            print("Greater than");
        } else {
            print("Not greater than");
        };
    }
];

main
{
    i = 45 * 2 + 1 / 33;
    myFunc(45);
    myFunc2(1);
    print("Hello World!");
}
end
```

Cómo resultado obtenemos la detección correcta de cada token o palabra reservada (no se muestran todos en la imagen) y en el análisis semántico obtenemos un None, que significa que no hubo errores ya que es un test correcto.

```
LexToken(CTE_INT,6,12,182)
LexToken(CLOSE_PAR,')',12,183)
LexToken(SEMI_COLON,';',12,184)
LexToken(RIGHT_CURLY,}',13,190)
LexToken(R_BRACKET,']',14,192)
LexToken(SEMI_COLON,';',14,193)
LexToken(VOID,'void',16,196)
LexToken(ID,'myFunc',16,201)
LexToken(OPEN_PAR,'(',16,208)
LexToken(ID,'param1',16,209)
LexToken(COLON,':',16,216)
LexToken(FLOAT,'float',16,218)
LexToken(CLOSE_PAR,')',16,223)
LexToken(L_BRACKET,['',16,225)
LexToken(LEFT_CURLY,{',17,231)
LexToken(IF,'if',18,241)
LexToken(OPEN_PAR,'(',18,244)
LexToken(ID,'param1',18,245)
LexToken(MORE_THAN,'>',18,252)
LexToken(CTE_INT,2,18,254)
LexToken(CLOSE_PAR,')',18,255)
LexToken(LEFT_CURLY,{',18,257)
LexToken(PRINT,'print',19,271)
LexToken(OPEN_PAR,'(',19,276)
```

```
        if (param1 > 2) {
            print("Greater than");
        } else {
            print("Not greater than");
        };
    }
];

main
{
    i = 45 * 2 + 1 / 33;
    myFunc(45);
    myFunc2(1);
    print("Hello World!");
}
end
None
```

Por otro lado, si quitamos un punto y como de una de las líneas en main, saldría el siguiente error. Un error de semántica en myFunc2, ya que previamente debería tener ese punto y como que se eliminó.

```
main
{
    i = 45 * 2 + 1 / 33;
    myFunc(45)
    myFunc2(1);
    print("Hello World!");
}
LexToken(SEMI_COLON, ';', 31, 459)
LexToken(RIGHT_CURLY, '}', 32, 461)
LexToken(END, 'end', 33, 463) "Not greater than"
Semantic error LexToken(ID, 'myFunc2', 62, 422)
```