



Instituto Tecnológico y de Estudios Superiores de Monterrey

Diseño de Compiladores TC3048

Profesor:

Dr. Víctor de la Cueva

Nombre:

Sergio Gabriel Domínguez Cordero - A01365942

Introducción:

El compilador genera código ensamblador para el microprocesador MIPS, este lenguaje fue elegido por que ya se tenía conocimiento de este cuando se cursó la materia TC1016 Organización computacional

Manual de usuario:

1. Si ya tienes la librería Python-LEX-YACC (PLY) ve al paso 2 sino ve al apéndice a la documentación del análisis sintáctico.
2. Escribe un fragmento de código C- en un archivo con el nombre *sample.c-*
 - a. **Nota:** Este compilador no fue terminado por lo que solamente puede generar código para ciertos fragmentos de código en c-, los fragmentos de código con los que fue probado y se genera código ensamblador se encuentran en la sección **Código C- válido**
3. Crear un archivo con terminación .py con el siguiente contenido:

```
from globalTypes import *
from parser import *
from semantica import *
from cgen import *
f = open('sample.c-', 'r')
programa = f.read()           # lee todo el archivo a compilar
proglong = len(programa)      # longitud original del programa
programa = programa + '$'     # agregar un caracter $ que represente EOF
posicion = 0                  # posición del caracter actual del string

# función para pasar los valores iniciales de las variables globales
globales(programa, posicion, proglong)

AST = parser(true)
semantica(AST, true)
codeGen(AST, "file.s")
```

La letra 'p'
debe ser mayúscula

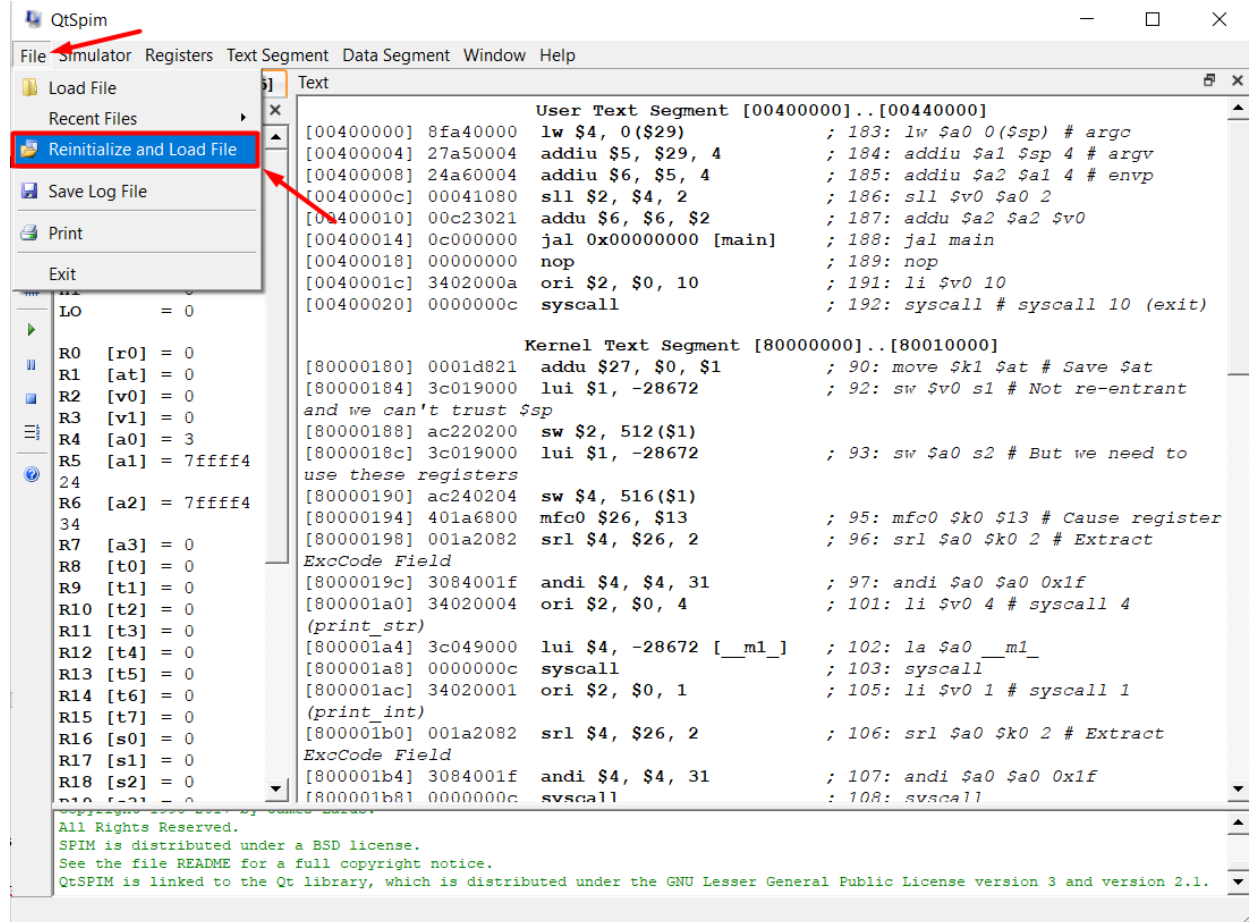
4. Para compilar el código en c- se tiene que ejecutar el siguiente comando:

```
python3 <nombre-de-archivo-creado-en-paso-3>
```

5. Se creará automáticamente un archivo llamado *file.s* el cual tiene código ensamblador que será ejecutado en el emulador QtSpim para el procesador MIPS.
6. Se requiere la instalación del emulador QtSpim:
<http://spimsimulator.sourceforge.net/>
7. Cuando se ejecute por primera vez QtSpim aparecerá la siguiente ventana (Imagen 1)

Elegir la opción File>Reinitialize and Load File

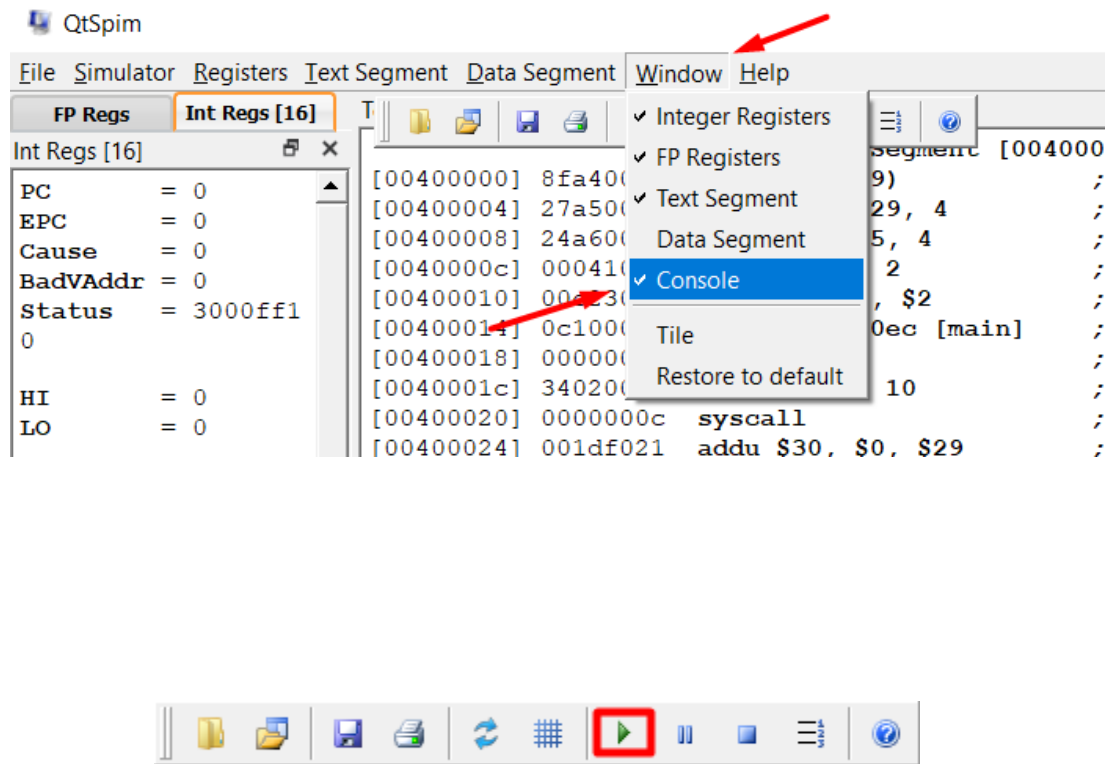
Imagen 1



8. Elegir el archivo con el nombre file.s y presionar el botón play

Nombre	Fecha de modificacion	tipo	lamano
.git	03/06/2020 07:19 p. m.	Carpeta de archivos	
.vscode	14/05/2020 08:26 a. m.	Carpeta de archivos	
__pycache__	03/06/2020 07:55 p. m.	Carpeta de archivos	
assembler	30/05/2020 11:07 a. m.	Carpeta de archivos	
pruebas	30/05/2020 11:07 a. m.	Carpeta de archivos	
compare.asm	03/06/2020 02:38 p. m.	Archivo ASM	4
file.s	03/06/2020 09:51 p. m.	Archivo S	3
salida.asm	03/06/2020 07:46 p. m.	Archivo ASM	4

9. El resultado será mostrado en la consola o estará a la espera de recibir una entrada por teclado (solo número entero).
10. En caso de que la consola no aparezca, presionar el botón que se muestra a continuación.



Código C- válido

```
int suma(void) {
    int p;
    int q;
    p = 2;
    q = 15;
    output(p);
    return p + q;
}

void main(void) {
    output(suma());
}
```

```
int foo(void) {
```

```
    int p;  
    int q;  
    p = 1;  
    q = 8;  
    output(p);  
    return p + q;  
}  
  
void main(void) {  
    int y;  
    y = foo();  
    output(y);  
}
```

```
void main(void) {  
    int y;  
    int z;  
    y = input();  
    z = input();  
    y = (y * z)/3;  
    output(y);  
}
```

```
void main(void) {  
    int u;  
    int v;  
    int modulo;  
    u = input();  
    v = input();  
    modulo = u-u/v*v;  
    output(modulo);  
}
```

```
int foo(void) {  
    int p;  
    int q;  
    int res;  
    p = 5;  
    q = 6;  
    res = p * q + 13;  
    return res;  
}  
  
void main(void) {  
    output(foo());  
}
```

```
int suma(int p, int q) {
```

```

        return p + q;
    }

void main(void) {
    int y;
    int z;
    y = 4;
    z = 3;
    y = suma(y, z);
    output(y);
}

```

```

int foo(int p) {
    return p;
}

void main(void) {
    output(foo(12));
}

```

Apéndice:

- Documentación Análisis Léxico:

Expresiones Regulares

KEYWORD = else|if|int|return|void|while

SPECIAL_SYMBOLS = |+|~|*|/|<|<=|>|>=|==|=|;|,|(|)|[[|]]{|}

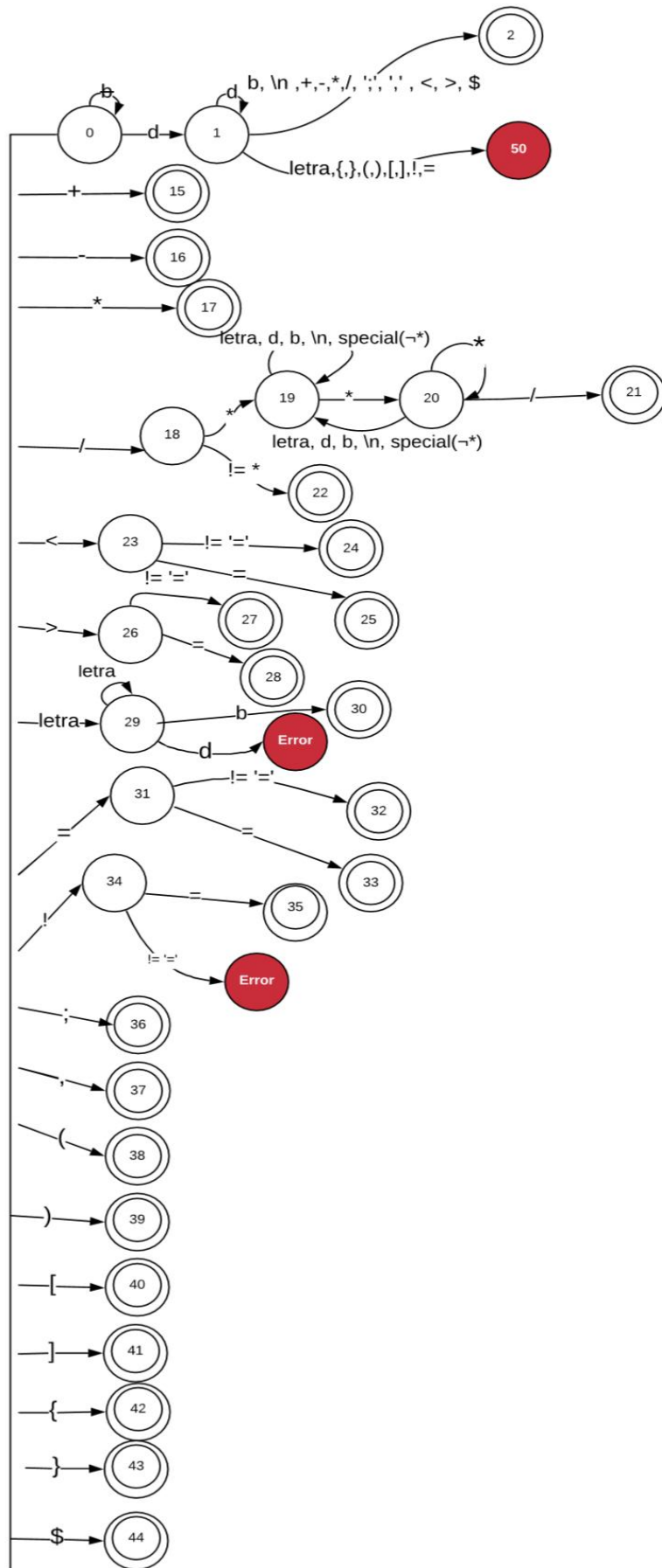
letra = [a-zA-Z]

COMMENTS = *(\n[a-zA-Z-0-9* \b])*~\

NUM = \s*[0-9][0-9]*

ID = letra letra*

Implementación de las expresiones regulares con un AFND



- Documentación Análisis Sintáctico

Descripción

Este analizador sintáctico fue implementado con la herramienta Python-LEX-YACC (PLY).

<https://www.dabeaz.com/ply/ply.html>

Antes de probar el script, realizar las instalaciones necesarias

Instalación de librería

Elegir una opción

```
$ pip install ply
```

<https://pypi.org/project/ply/>

```
$ conda install -c conda-forge ply
```

<https://anaconda.org/conda-forge/ply>

Gramática usada

Se uso la gramática que se encuentra en el archivo de la definición del lenguaje C- en BNF, no fue necesario realizar la transformación a EBNF.

Documentación Análisis Semántico

ID es una literal entera

$\vdash ID: Int$

$\vdash e1: Int \quad \vdash e2: Int$

$\vdash e1 + e2: Int$

$\vdash e1: Int \quad \vdash e2: Int$

$\vdash e1 - e2: Int$

$\vdash e1: Int \quad \vdash e2: Int$

$\vdash e1 * e2: Int$

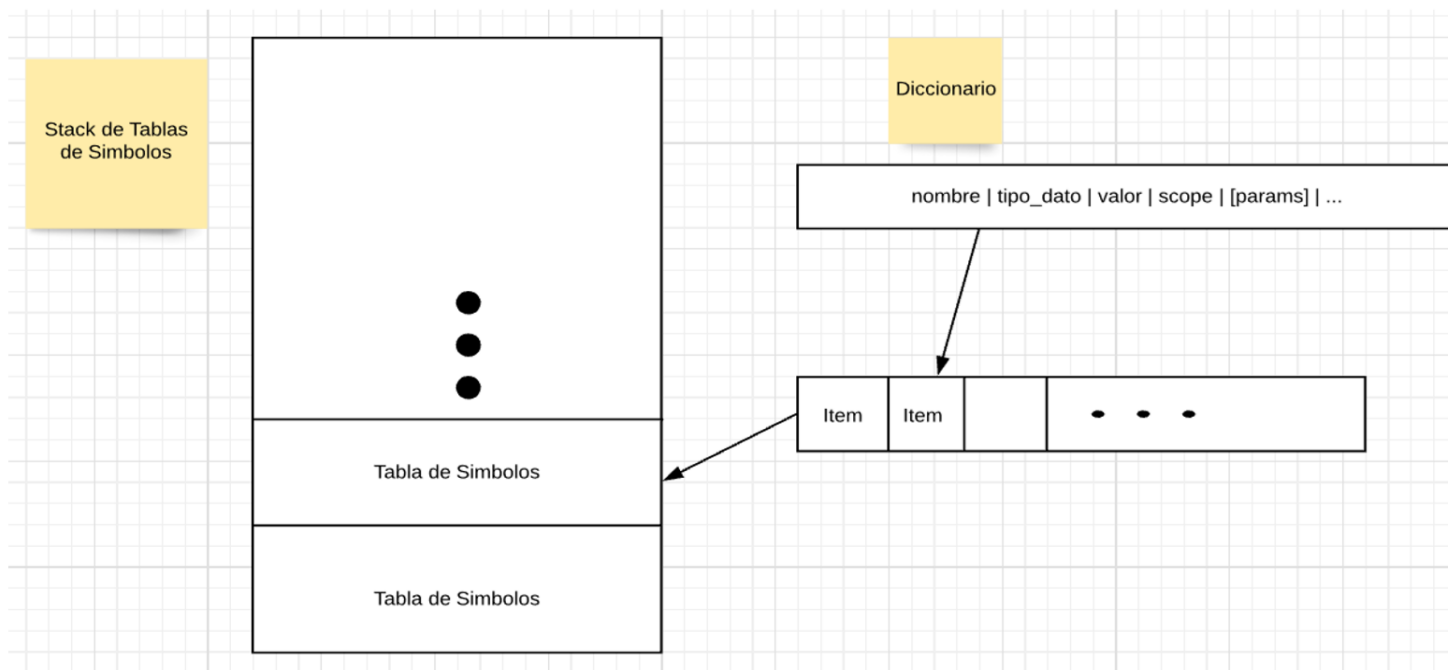
$\vdash e1: Int \quad \vdash e2: Int$

$\vdash e1 / e2: Int$

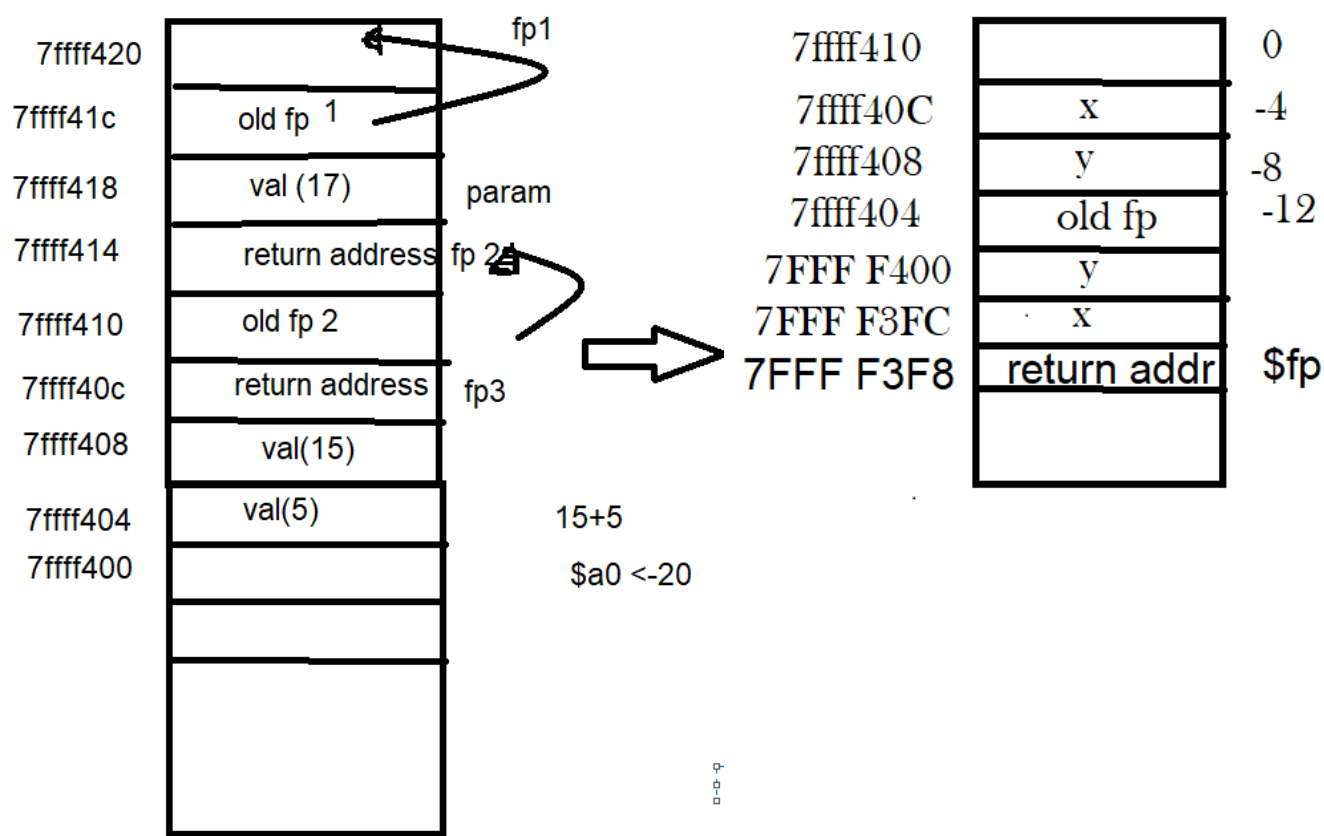
$\vdash foo(): Int \quad \vdash e2: Int$

$\vdash foo() + e2: Int$

Diagrama del Stack de Tablas de Símbolos usado



Estructura del Activation Record



Definición del Lenguaje

C- (C-minus)

Un lenguaje para un proyecto de compilador

(Lauden, 2004)

Es un subconjunto considerablemente restringido de C.

Contiene:

- Enteros
- Arreglos de enteros
- Funciones (con tipo o **void**)
- Declaraciones (estáticas) locales y globales
- Funciones recursivas (simples)
- Condicional **if-else**
- Ciclo **while**
- Función **input()** que lee desde el teclado
- Función **output()** que escribe a la pantalla

Un programa se compone de una secuencia de declaraciones de variables y funciones. Al final debe declararse una función **main**.

La ejecución inicia con una llamada a **main**.

Léxico de C-

1. Las palabras clave o reservadas del lenguaje son las siguientes:

```
else if int return void while
```

Todas las palabras reservadas o clave están reservadas, y deben ser escritas en minúsculas.

2. Los símbolos especiales son los siguientes:

```
+ - * / < <= > >= == != = ; , ( ) [ ] { } /* */
```

3. Otros tokens son *ID* y *NUM*, definidos mediante las siguientes expresiones regulares:

```
ID = letra letra*  
NUM = dígito dígito*  
letra = a|..|z|A|..|Z  
dígito = 0|..|9
```

Se distingue entre letras minúsculas y mayúsculas.

4. Los espacios en blanco se componen de blancos, retornos de línea y tabulaciones. El espacio en blanco es ignorado, excepto cuando deba separar *ID*, *NUM* y palabras reservadas.
5. Los comentarios están encerrados entre las anotaciones habituales del lenguaje C */*...*/*. Los comentarios se pueden colocar en cualquier lugar donde pueda aparecer un espacio en blanco (es decir, los comentarios no pueden ser colocados dentro de los token) y pueden incluir más de una línea. Los comentarios no pueden estar anidados.

Sintaxis de C-

Una gramática BNF para C- es como se describe a continuación:

1. $program \rightarrow declaration\text{-}list$
2. $declaration\text{-}list \rightarrow declaration\text{-}list\ declaration \mid declaration$
3. $declaration \rightarrow var\text{-}declaration \mid fun\text{-}declaration$
4. $var\text{-}declaration \rightarrow type\text{-}specifier\ ID \ ; \mid type\text{-}specifier\ ID \ [\ NUM \] \ ;$
5. $type\text{-}specifier \rightarrow int \mid void$
6. $fun\text{-}declaration \rightarrow type\text{-}specifier\ ID \ (\ params \) \ compound\text{-}stmt$
7. $params \rightarrow param\text{-}list \mid void$
8. $param\text{-}list \rightarrow param\text{-}list \ , \ param \mid param$
9. $param \rightarrow type\text{-}specifier\ ID \mid type\text{-}specifier\ ID \ [\]$
10. $compound\text{-}stmt \rightarrow \{ \ local\text{-}declarations \ statement\text{-}list \}$
11. $local\text{-}declarations \rightarrow local\text{-}declarations \ var\text{-}declaration \mid empty$
12. $statement\text{-}list \rightarrow statement\text{-}list \ statement \mid empty$
13. $statement \rightarrow expression\text{-}stmt \mid compound\text{-}stmt \mid selection\text{-}stmt \mid iteration\text{-}stmt \mid return\text{-}stmt$
14. $expression\text{-}stmt \rightarrow expression \ ; \ ;$
15. $selection\text{-}stmt \rightarrow if \ (\ expression \) \ statement \mid if \ (\ expression \) \ statement \ else \ statement$
16. $iteration\text{-}stmt \rightarrow while \ (\ expression \) \ statement$
17. $return\text{-}stmt \rightarrow return \ ; \mid return \ expression \ ;$
18. $expression \rightarrow var = expression \mid simple\text{-}expression$
19. $var \rightarrow ID \mid ID \ [\ expression \]$
20. $simple\text{-}expression \rightarrow additive\text{-}expression \ relop \ additive\text{-}expression \mid additive\text{-}expression$
21. $relop \rightarrow <= \mid < \mid > \mid >= \mid == \mid !=$
22. $additive\text{-}expression \rightarrow additive\text{-}expression \ addop \ term \mid term$
23. $addop \rightarrow + \mid -$
24. $term \rightarrow term \ mulop \ factor \mid factor$
25. $mulop \rightarrow * \mid /$
26. $factor \rightarrow (\ expression \) \mid var \mid call \mid NUM$
27. $call \rightarrow ID \ (\ args \)$
28. $args \rightarrow arg\text{-}list \mid empty$
29. $arg\text{-}list \rightarrow arg\text{-}list \ , \ expression \mid expression$

Semántica de C-

1. $program \rightarrow declaration\text{-}list$
2. $declaration\text{-}list \rightarrow declaration\text{-}list\ declaration \mid declaration$
3. $declaration \rightarrow var\text{-}declaration \mid fun\text{-}declaration$

Un programa (*program*) se compone de una lista (o secuencia) de declaraciones (*declaration-list*), las cuales pueden ser declaraciones de variable o función, en cualquier orden. Debe haber al menos una declaración. Las restricciones semánticas son como sigue (éstas no se presentan en C). Todas las variables y funciones deben ser declaradas antes de utilizarlas (esto evita las referencias de retroajuste). La última declaración en un programa debe ser una declaración de función con el nombre **main**. Advierta que C- carece de prototipos, de manera que no se hace una distinción entre declaraciones y definiciones (como en el lenguaje C).

4. $var\text{-}declaration \rightarrow type\text{-}specifier\ ID\ ; \mid type\text{-}specifier\ ID\ [\text{NUM}]\ ;$
5. $type\text{-}specifier \rightarrow int \mid void$

Una declaración de variable declara una variable simple de tipo entero o una variable de arreglo cuyo tipo base es entero, y cuyos índices abarcan desde 0 . . **NUM** - 1. Observe que en C- los únicos tipos básicos son entero y vacío ("void"). En una declaración de variable sólo se puede utilizar el especificador de tipo **int**. **Void** es para declaraciones de función (véase más adelante). Advierta también que sólo se puede declarar una variable por cada declaración.

6. $fun\text{-}declaration \rightarrow type\text{-}specifier\ ID\ (\text{params})\ compound\text{-}stmt$
7. $params \rightarrow param\text{-}list \mid void$
8. $param\text{-}list \rightarrow param\text{-}list\ ,\ param \mid param$
9. $param \rightarrow type\text{-}specifier\ ID \mid type\text{-}specifier\ ID\ [\]$

Una declaración de función consta de un especificador de tipo (*type-specifier*) de retorno, un identificador y una lista de parámetros separados por comas dentro de paréntesis, seguida por una sentencia compuesta con el código para la función. Si el tipo de retorno de la función es **void**, entonces la función no devuelve valor alguno (es decir, es un procedimiento). Los parámetros de una función pueden ser **void** (es decir, sin parámetros) o una lista que representa los parámetros de la función. Los parámetros seguidos por corchetes son parámetros de arreglo cuyo tamaño puede variar. Los parámetros enteros simples son pasados por valor. Los parámetros de arreglo son pasados por referencia (es decir, como apun­tadores) y deben ser igualados mediante una variable de arreglo durante una llamada. Advierta que no hay parámetros de tipo "función". Los parámetros de una función tienen un ámbito igual a la sentencia compuesta de la declaración de función, y cada invocación de una función tiene un conjunto separado de parámetros. Las funciones pueden ser recursivas (hasta el punto en que la declaración antes del uso lo permita).

10. $compound\text{-}stmt \rightarrow \{ \text{local-declarations statement-list} \}$

Una sentencia compuesta se compone de llaves que encierran un conjunto de declaraciones y sentencias. Una sentencia compuesta se realiza al ejecutar la secuencia de sentencias

11. *local-declarations* \rightarrow *local-declarations* *var-declaration* | *empty*
 12. *statement-list* \rightarrow *statement-list* *statement* | *empty*

13. *statement* \rightarrow *expression-stmt*
 compound-stmt
 selection-stmt
 iteration-stmt
 return-stmt

15. *selection-stmt* \rightarrow **if** (*expression*) *statement*
 if (*expression*) *statement* **else** *statement*

Una expresión es una referencia de variable seguida por un símbolo de asignación (signo de igualdad) y una expresión, o solamente una expresión simple. La asignación tiene la semántica de almacenamiento habitual: se encuentra la localidad de la variable representada por *var*, luego se evalúa la subexpresión a la derecha de la asignación, y se almacena el valor de la subexpresión en la localidad dada. Este valor también es devuelto como el valor de la expresión completa. Una *var* es una variable (entera) simple o bien una variable de arreglo subíndizada. Un subíndice negativo provoca que el programa se detenga (a diferencia de C). Sin embargo, no se verifican los límites superiores de los subíndices.

Las variables representan una restricción adicional en C^- respecto a C . En C el objetivo de una asignación debe ser un **valor I**, y los valores I son direcciones que pueden ser obtenidas mediante muchas operaciones. En C^- los únicos valores I son aquellos dados por la sintaxis de *var*, y así esta categoría es verificada sintácticamente, en vez de hacerlo durante la verificación de tipo como en C . Por consiguiente, en C^- está prohibida la aritmética de apuntadores.

20. $simple-expression \rightarrow additive-expression \text{ relop } additive-expression$
 $\quad \quad \quad | \quad additive-expression$

21. $relop \rightarrow <= \mid < \mid > \mid >= \mid == \mid !=$

Una expresión simple se compone de operadores relacionales que no se asocian (es decir, una expresión sin paréntesis puede tener solamente un operador relacional). El valor de una expresión simple es el valor de su expresión aditiva si no contiene operadores relacionales, o bien, 1 si el operador relacional se evalúa como verdadero, o 0 si se evalúa como falso.

22. $additive-expression \rightarrow additive-expression \text{ addop } term \mid term$

23. $addop \rightarrow + \mid -$

24. $term \rightarrow term \text{ mulop } factor \mid factor$

25. $mulop \rightarrow * \mid /$

Los términos y expresiones aditivas representan la asociatividad y precedencia típicas de los operadores aritméticos. El símbolo $/$ representa la división entera; es decir, cualquier residuo es truncado.

26. $factor \rightarrow (expression) \mid var \mid call \mid \text{NUM}$

Un factor es una expresión encerrada entre paréntesis, una variable, que evalúa el valor de su variable; una llamada de una función, que evalúa el valor devuelto de la función; o un NUM, cuyo valor es calculado por el analizador léxico. Una variable de arreglo debe estar sub-indizada, excepto en el caso de una expresión compuesta por una ID simple y empleada en una llamada de función con un parámetro de arreglo (véase a continuación).

27. $call \rightarrow ID \ (\ args)$

28. $args \rightarrow arg-list \mid empty$

29. $arg-list \rightarrow arg-list , expression \mid expression$

Una llamada de función consta de un **ID** (el nombre de la función), seguido por sus argumentos encerrados entre paréntesis. Los argumentos pueden estar vacíos o estar compuestos por una lista de expresiones separadas mediante comas, que representan los valores que se asignarán a los parámetros durante una llamada. Las funciones deben ser declaradas antes de llamarlas, y el número de parámetros en una declaración debe ser igual al número de argumentos en una llamada. Un parámetro de arreglo en una declaración de función debe coincidir con una expresión compuesta de un identificador simple que representa una variable de arreglo.

Finalmente, las reglas anteriores no proporcionan sentencia de entrada o salida. Debemos incluir tales funciones en la definición de C-, puesto que a diferencia del lenguaje C, C- no tiene facilidades de ligado o compilación por separado. Por lo tanto, consideraremos dos funciones por ser **predefinidas** en el ambiente global, como si tuvieran las declaraciones indicadas:

```
int input(void) { . . . }  
void output(int x) { . . . }
```

La función **input** no tiene parámetros y devuelve un valor entero desde el dispositivo de entrada estándar (por lo regular el teclado). La función **output** toma un parámetro entero, cuyo valor imprime a la salida estándar (por lo regular la pantalla), junto con un retorno de línea.

Ejemplos de programas en C-

El siguiente es un programa que introduce dos enteros, calcula su máximo común divisor y lo imprime:

```
/* Un programa para realizar el algoritmo
   de Euclides para calcular mcd. */

int gcd (int u, int v)
{ if (v == 0) return u ;
  else return gcd(v,u-u/v*v);
  /* u-u/v*v == u mod v */
}

void main(void)
{ int x; int y;
  x = input(); y = input();
  output(gcd(x,y));
}
```

A continuación tenemos un programa que introduce una lista de 10 enteros, los clasifica por orden de selección, y los exhibe otra vez:

```
/* Un programa para realizar ordenación por
   selección en un arreglo de 10 elementos. */

int x[10];

int minloc ( int a[], int low, int high )
{ int i; int x; int k;
  k = low;
  x = a[low];
  i = low + 1;
  while (i < high)
  { if (a[i] < x)
    { x = a[i];
      k = i; }
    i = i + 1;
  }
  return k;
}

void sort( int a[], int low, int high)
```

```

{ int i; int k;
  i = low;
  while (i < high-1)
  { int t;
    k = minloc(a,i,high);
    t = a[k];
    a[k] = a[i];
    a[i] = t;
    i = i + 1;
  }
}

```

```

void main(void)
{ int i;
  i = 0;
  while (i < 10)
  { x[i] = input();
    i = i + 1; }
  sort(x,0,10);
  i = 0;
  while (i < 10)
  { output(x[i]);
    i = i + 1; }
}

```