

**Instituto Tecnológico y de Estudios Superiores de Monterrey**

Campus Monterrey



**Proyecto Final – Diseño de Compiladores TC3048**

**AGRO**

**Semestre febrero-junio 2021**

**Docentes:**

**Elda Guadalupe Quiroga González**

**Héctor Gibrán Ceballos Cancino**

Aglahir Jiménez Flórez A01364026

Roberto Ramírez Monroy                  A01366943

Índice

[**Descripción del proyecto** 4](#_Toc73453656)

[Propósito y alcance del proyecto 4](#_Toc73453657)

[Análisis de requerimientos y descripción de los principales *Test Cases* 4](#_Toc73453658)

[Descripción del PROCESO general seguido para el desarrollo del proyecto 4](#_Toc73453659)

[Reflexiones personales 5](#_Toc73453660)

[**Descripción del lenguaje** 5](#_Toc73453661)

[Nombre del lenguaje 5](#_Toc73453662)

[Descripción genérica de las principales características del lenguaje 6](#_Toc73453663)

[Listado de los errores que pueden ocurrir 6](#_Toc73453664)

[En compilación 6](#_Toc73453665)

[En ejecución 7](#_Toc73453666)

[**Descripción del compilador** 7](#_Toc73453667)

[Equipo de cómputo, lenguaje y utilerías especiales usadas 7](#_Toc73453668)

[Descripción del análisis léxico 7](#_Toc73453669)

[Patrones de construcción (expresados con expresiones regulares) de los elementos principales 7](#_Toc73453670)

[Enumeración de los “tokens” del lenguaje y su código asociado 7](#_Toc73453671)

[Descripción del análisis de sintaxis 8](#_Toc73453672)

[Gramática formal empleada para representar las estructuras sintácticas (sin “codificar”) 8](#_Toc73453673)

[Descripción de generación de código intermedio y análisis semántico 9](#_Toc73453674)

[Código de operación y direcciones virtuales asociadas a los elementos del código 9](#_Toc73453675)

[Diagramas de sintaxis con las acciones correspondientes marcadas sobre ellos 9](#_Toc73453676)

[Tabla de consideraciones semánticas (combinaciones factibles y errores de tipo) 18](#_Toc73453677)

[Descripción detallada del proceso de Administración de memoria en la compilación 19](#_Toc73453678)

[Especificación gráfica de CADA estructura de datos usada 19](#_Toc73453679)

[**Descripción de la máquina virtual** 21](#_Toc73453680)

[Equipo de cómputo, lenguaje y utilerías especiales usadas ( en caso de ser diferente que el compilador) 21](#_Toc73453681)

[Equipo de cómputo 21](#_Toc73453682)

[Lenguaje 21](#_Toc73453683)

[Utilerías de Python 3 usadas 21](#_Toc73453684)

[Descripción detallada del proceso de Administración de memoria en ejecución (arquitectura) 21](#_Toc73453685)

[Especificación gráfica de CADA estructura de datos usada para el manejo de scopes (Memoria local, global, etc.…) 22](#_Toc73453686)

[Asociación hecha entre las direcciones virtuales (compilación) y reales (ejecución) 22](#_Toc73453687)

[**Pruebas del funcionamiento del lenguaje** 22](#_Toc73453688)

[Incluir pruebas que “comprueben” el funcionamiento del proyecto 22](#_Toc73453689)

[Codificación de la prueba (en AGRO) 22](#_Toc73453690)

[Resultados arrojados por la generación de código intermedio y por la ejecución. 22](#_Toc73453691)

[**Documentación del código del proyecto** 22](#_Toc73453692)

# **Descripción del proyecto**

## Propósito y alcance del proyecto

Este proyecto es desarrollado con el propósito de acreditar la materia TC3048 (Diseño de compiladores) y conocer más a fondo el comportamiento verdadero de los lenguajes de programación que usamos en el día a día de nuestras vidas como desarrolladores de software.

El producto es un lenguaje orientado a objetos, con las características básicas de un lenguaje multiparadigma con conceptos también procedurales, al estilo de C++.

## Análisis de requerimientos y descripción de los principales *Test Cases*

**Requerimientos:**

* Lenguaje de programación siguiendo el paradigma de programación procedural.
* El lenguaje debe poder hacer declaración de variables.
* El lenguaje debe poder hacer asignación a variables.
* El lenguaje debe poder declarar funciones.
* El lenguaje debe poder distinguir contextos de memorias.
* El lenguaje debe poder interactuar con el usuario por medio de inputs y outputs.
* El lenguaje debe poder utilizar condicionales if-else.
* El lenguaje debe poder utilizar ciclos while y for.
* El lenguaje debe poder declarar variables dimensionadas (Arreglos y matrices).
* El lenguaje debe poder declarar clases con atributos, métodos y herencia simple.
* El lenguaje debe poder validar casos y mostrar errores al usuario.

## Descripción del PROCESO general seguido para el desarrollo del proyecto

Para el desarrollo de este proyecto usamos Git como herramienta de control de versiones ([rbtote/AGRO (github.com)](https://github.com/rbtote/AGRO)) teniendo en la herramienta un listado completo de los “*commits*” al proyecto, pero a continuación mostramos una lista de los que consideramos más importantes:

|  |  |
| --- | --- |
| **Fecha** | **Avance** |
| Apr 15 2021 | First iteration of the syntax of the language |
| Apr 16 2021 | This version includes changes in grammar according to a basic variable, function and object declaration. We have created a SymbolTable class that implements an infinite recursive linear scope, in which every scope has a pointer to its parent scope. |
| Apr 23 2021 | This version includes 3 new classes to handle our first code generation behaviors, we have set the neuralgic spots in the grammar syntax, for the linear statutes (variables, arithmetic assignments) |
| Apr 25 2021 | This version fixes the quad creation bug |
| May 1 2021 | This version includes a new set of instructions for Jump handling (GOTO). Added While and For loops code generation. Added input, print code generation. Added IF ELSE code generation. |
| May 8 2021 | This version includes the modular functionality for functions (declaration and call) and the needed new operation codes (era, gosub, endfunc and return [and their classes]) for the correct creation of all the needed quads. |
| May 15 2021 | This version includes the handling of return values in functions that are not voids, as well as the functionality for them in hyper\_expressions (used for recursion). |
| May 18 2021 | This code changes the quad representation to memory addresses, and fixed some bugs on parameter numbers.  Documentation 1st iteration. |
| May 24 2021 | In this version, we added all the array logic to the quad generation, changed back the operators to string, in order to help our VM |
| May 25 2021 | In this version, we added all the array logic to the quad generation, changed back the operators to string, in order to help our VM, that is already in development. Also, we validated some errors in compilation (use of non-declared vars) |
| May 28 2021 | Class and object creation finalized, just checking method logic. |
| May 31 2021 | Methods and inheritance working. |

## Reflexiones personales

Aglahir Jiménez Flórez

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Aglahir Jiménez Flórez

A01364026

Roberto Ramírez Monroy:

El desarrollo de este proyecto fue en realidad un reto muy interesante y disfrutable, aunque en el inicio del semestre quizás no teníamos idea de cómo se iban a realizar las etapas avanzadas del proyecto, como la máquina virtual y la asignación de elementos como arreglos u objetos en memoria. Entonces quizás llegamos a poner el alcance un poco más allá de lo que íbamos a poder hacer en el tiempo que teníamos. De cualquier manera, aprendí muchas cosas que pueden mejorar la manera en la que codificaré día a día en mi carrera, ya que no siempre conocemos más a fondo el cómo sirve la parte "escondida” en los lenguajes y compiladores no sabemos con exactitud la complejidad temporal o espacial de todo el código que estamos ejecutando, pero con los conocimientos que adquirí con el desarrollo del proyecto y las clases puedo hacer un mejor análisis del código que escriba. Y en general del producto final siento que lo hicimos de una manera muy buena y estoy orgulloso del resultado que pudimos generar.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Roberto Ramírez Monroy

A01366943

# **Descripción del lenguaje**

## Nombre del lenguaje

El lenguaje que desarrollamos tiene el nombre de AGRO, por las dos iniciales de los nombres de los integrantes del equipo, Aglahir y Roberto. Y haciendo un poco referencia al Pokémon llamado Aggron por la similitud fonética, entonces decidimos usarlo como “mascota” del proyecto.

## Descripción genérica de las principales características del lenguaje

AGRO es un lenguaje de programación de alto nivel, con paradigma orientado a objetos (estilo Java, C++). Los datos primitivos que soporta son: int, float, char y string.

Permite creación de variables unidimensionales y bidimensionales, la creación de funciones y estatutos básicos como en C++ (condicionales, ciclos, imprimir, leer inputs, etc.)

## Listado de los errores que pueden ocurrir

Nuestro compilador, al constar de dos partes, tiene diferentes tipos de errores en cada una de las etapas de la corrida de un programa.

### En compilación

En compilación tenemos:

* Errores de sintaxis

Estos errores consisten en la incorrecta formación de los tokens contenidos en el lenguaje. (Los cuales están listados en los siguientes apartados de este documento). Ejemplo:

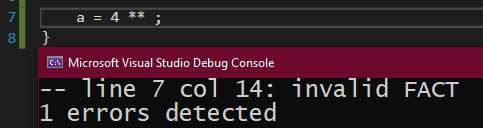
Una variable con nombre con caracteres especiales.

* Errores de gramática

Los errores de gramática consisten en el no cumplimiento de la estructura definida para los elementos.

* Incumplimiento regla

Ejemplo:

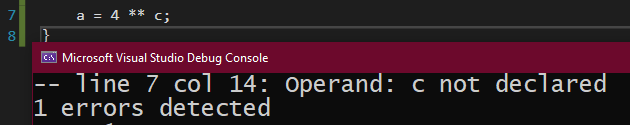


* Errores de semántica

Estos errores se presentan cuando no tiene lógica en compilación algo que se ha tecleado, aunque cumpla con la gramática y sintaxis definida.

* Variables no declaradas
* Variables haciendo operación no válida por tipos
* Falta de valor de retorno de funciones tipadas.
* Asignación de valor de funciones void

Ejemplo:



### En ejecución

# **Descripción del compilador**

## Equipo de cómputo, lenguaje y utilerías especiales usadas

Para el proyecto, utilizamos el generador de compiladores COCO/R para C#

## Descripción del análisis léxico

### Patrones de construcción (expresados con expresiones regulares) de los elementos principales

letter = 'A'..'Z' + 'a'..'z'.

digit = '0'..'9'.

tab = '\t'.

cr = '\r'.

lf = '\n'.

newLine = cr + lf.

notQuote = ANY - '"' - "\r\n".

id = letter { letter | digit }.

cte\_I = digit { digit }.

cte\_F = digit { digit } "." digit { digit }.

ctr\_Str = '"' {notQuote} '"'.

### Enumeración de los “tokens” del lenguaje y su código asociado

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Cbl** | { | **Add** | + | **Dot** | . | **Less** | < |
| **Cbr** | } | **Sub** | - | **Sadd** | += | **Greater** | > |
| **Bl** | [ | **Mul** | \* | **Ssub** | -= | **Lesseq** | <= |
| **br** | ] | **exponent** | \*\* | **sdiv** | /= | **greatereq** | >= |
| **Pl** | ( | **Div** | / | **Smul** | \*= | **Equaleq** | == |
| **Pr** | ) | **Intdiv** | // | **Increment** | ++ | **Different** | ¡= |
| **Comma** | , | **Module** | % | **Decrement** | -- | **And** | && |
| **Semicolon** | ; | **Equal** | = | **Colon** | : | **or** | || |

## Descripción del análisis de sintaxis

### Gramática formal empleada para representar las estructuras sintácticas (sin “codificar”)

**PROGRAM =** { DECLARATION } MAIN

**DECLARATION =** DEC\_VARS | DEC\_FUNC | DEC\_CLASS

**DEC\_VARS =** COMPOUND\_TYPE **id** {**, id**} **; |** SIMPLE\_TYPE **id** [ **[cte\_i]** [**[cte\_i]**]] **;**

**DEC\_FUNC =** TYPE\_FUNC **id (** [PARAMS\_FUNC] **) {** {DEC\_VARS | STATUTE} [RETURN]**}**

**DEC\_CLASS = class id** [: **id**] **{** CLASS\_DEF **}**

**COMPOUND\_TYPE = id**

**SIMPLE\_TYPE = int | float | char**

**MAIN = main {** {STATUTE | DEC\_VARS} **}**

**STATUTE =** FUNC\_CALL | METHOD\_CALL | INPUT | PRINT | CONDITIONAL | WHILE | FOR | ASSIGN

**TYPE\_FUNC = int | float | char | void**

**PARAMS\_FUNC =** SIMPLE\_TYPE **id** { **,** SIMPLE\_TYPE **id** }

**RETURN = return** HYPER\_EXP **;**

**EXP =** TERM { (**add | sub**) TERM}

**CLASS\_DEF =** ( **+ | -** ) (DEC\_VARS | DEC\_FUNC)

**ASSIGN =** VARIABLE\_ASSIGN **=** EXP **;** |

VARIABLE\_ASSIGN SHORT\_ASSIGN EXP**;** |

VARIABLE\_ASSIGN STEP**;**

**INPUT = input (**VARIABLE\_ASSIGN**);**

**PRINT = print (** {HYPER\_EXP} **) ;**

**CONDITIONAL = if (** HYPER\_EXP **)** BLOCK | **if (** HYPER\_EXP **)** BLOCK **else** BLOCK

**WHILE = while (** HYPER\_EXP **)** BLOCK

**FOR = for (** ASSIGN **;** HYPER\_EXP **;** ASSIGN**)** BLOCK

**FUNC\_CALL = id (); | id(** HYPER\_EXP {, HYPER\_EXP} **) ;**

**TERM =** FACT | FACT ( **\* | /** ) FACT

**SHORT\_ASSIGN = += | -= | \*= | /=**

**VARIABLE\_ASSIGN = id | id.id | id[**EXP**] | id[**EXP**] [**EXP**]**

**VARIABLE\_FACT = id | id() | id(**EXP {, EXP}**) | id.id | id.id(**EXP {,EXP}**) | id[**EXP**] | id[**EXP**] [**EXP**]**

**HYPER\_EXP =** SUPER\_EXP **|** SUPER\_EXP (REL\_EXP) SUPER\_EXP

**BLOCK = {** {STATUTE} **}**

**FACT = (** HYPER\_EXP **) |** (+ | -) (**cte\_i** | **cte\_f**) **|** VARIABLE\_FACT **| cte\_str**

**STEP = ++ | --**

**REL\_EXP = and | or | && | ||**

**SUPER\_EXP =** EXP **|** EXP (REL\_OP) EXP

**REL\_OP = < | > | <= | >= | == | !=**

## Descripción de generación de código intermedio y análisis semántico

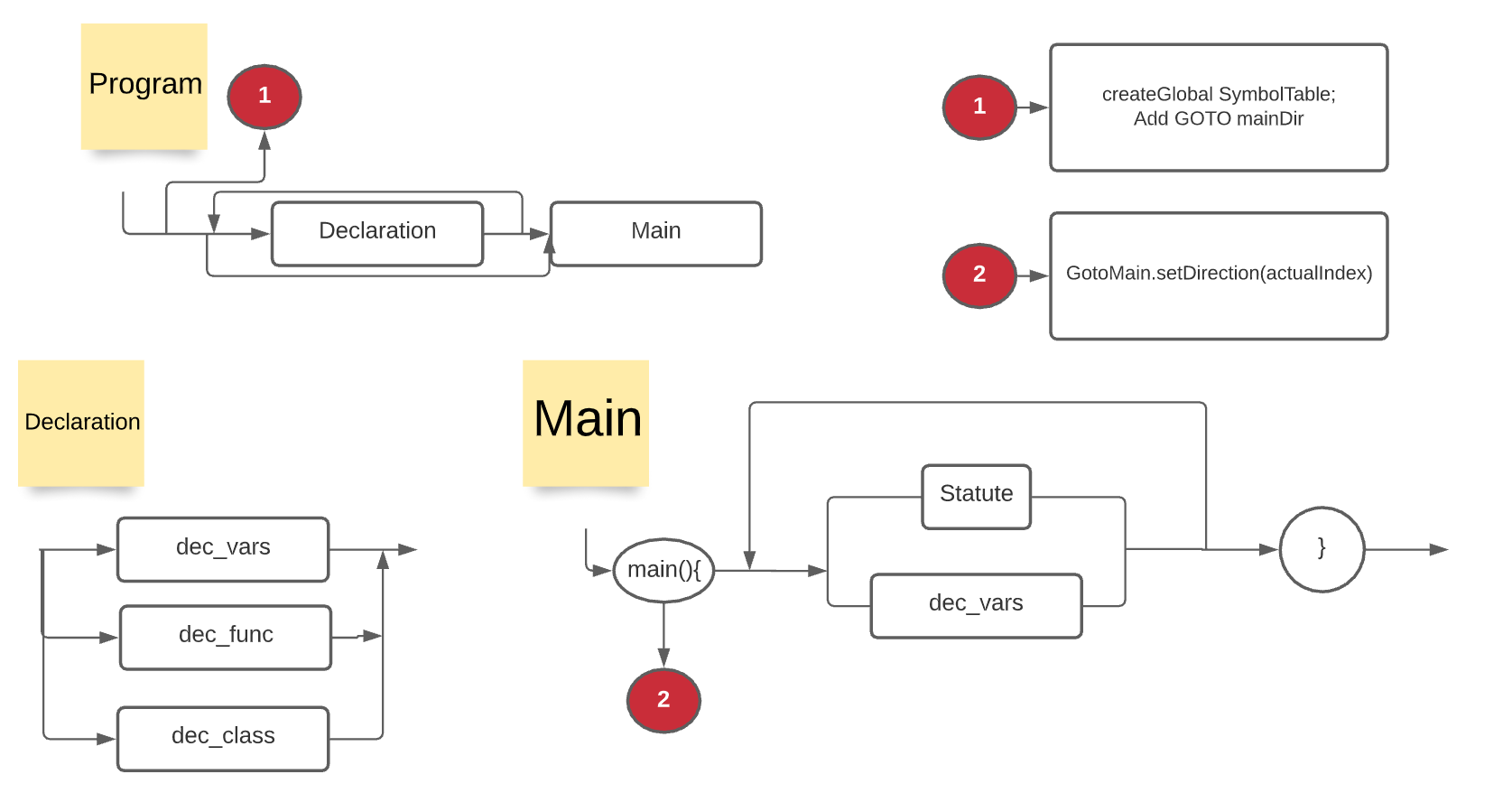
### Código de operación y direcciones virtuales asociadas a los elementos del código

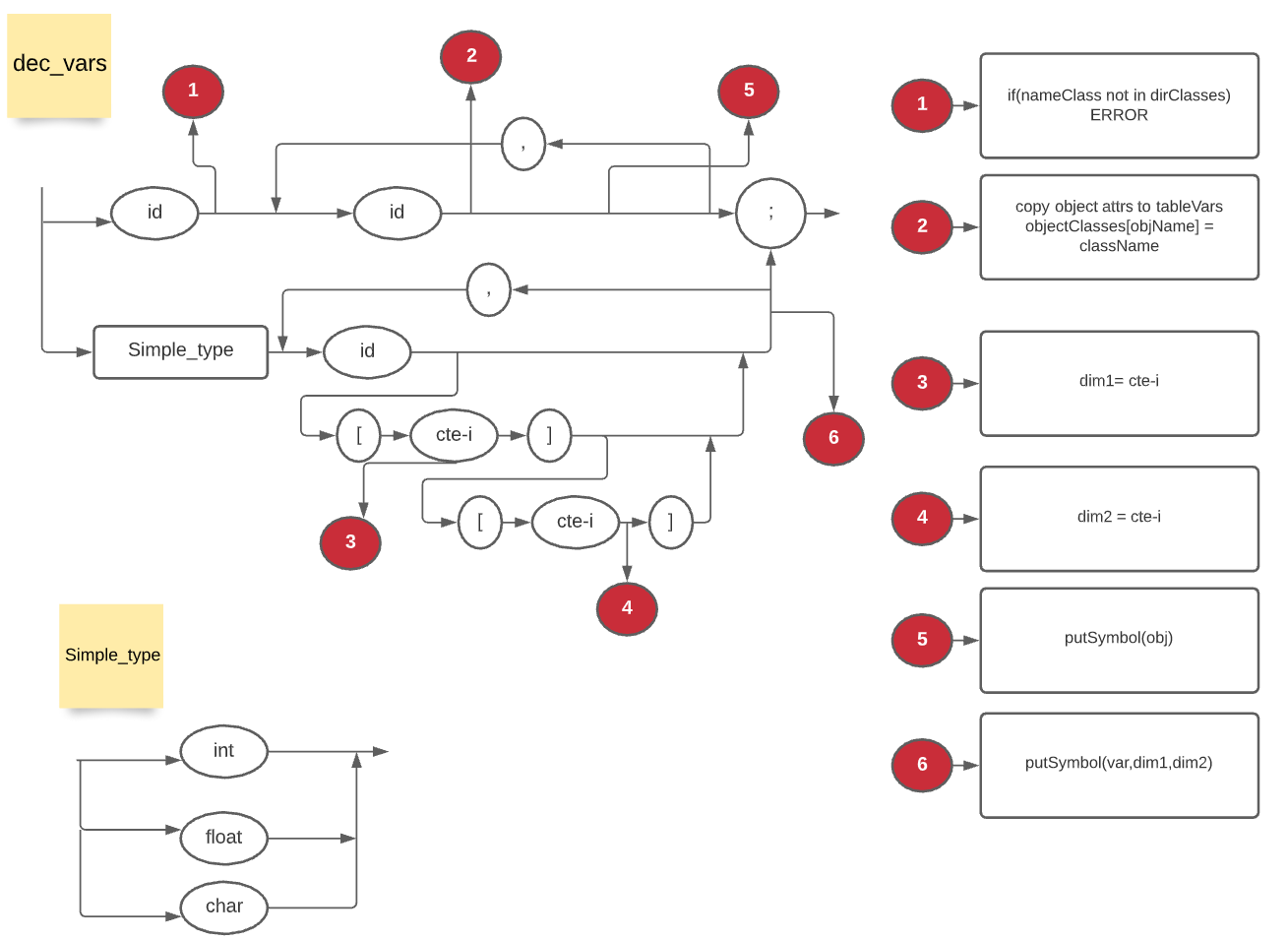
**Direcciones virtuales utilizadas:**

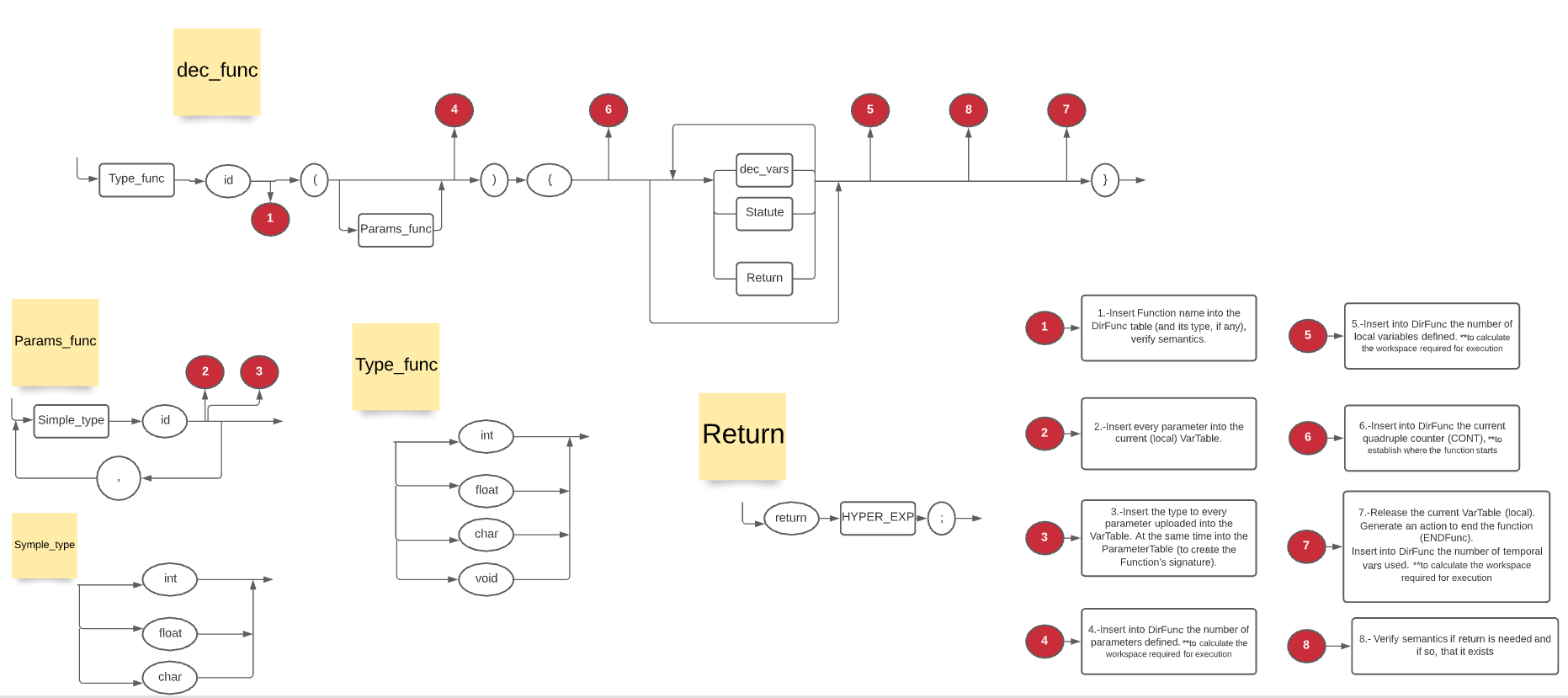
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **globalInt** | 1001 | **localInt** | 28001 | **constInt** | 42001 |
| **globalFloat** | 5001 | **localFloat** | 30001 | **constFloat** | 44001 |
| **globalChar** | 9001 | **localChar** | 32001 | **constChar** | 46001 |
| **globalTempInt** | 12001 | **localTempInt** | 34001 | **constString** | 48001 |
| **globalTempFloat** | 16001 | **localTempFloat** | 36001 | **Pointers** | 50001 |
| **globalTempChar** | 20001 | **localTempChar** | 38001 |  |  |

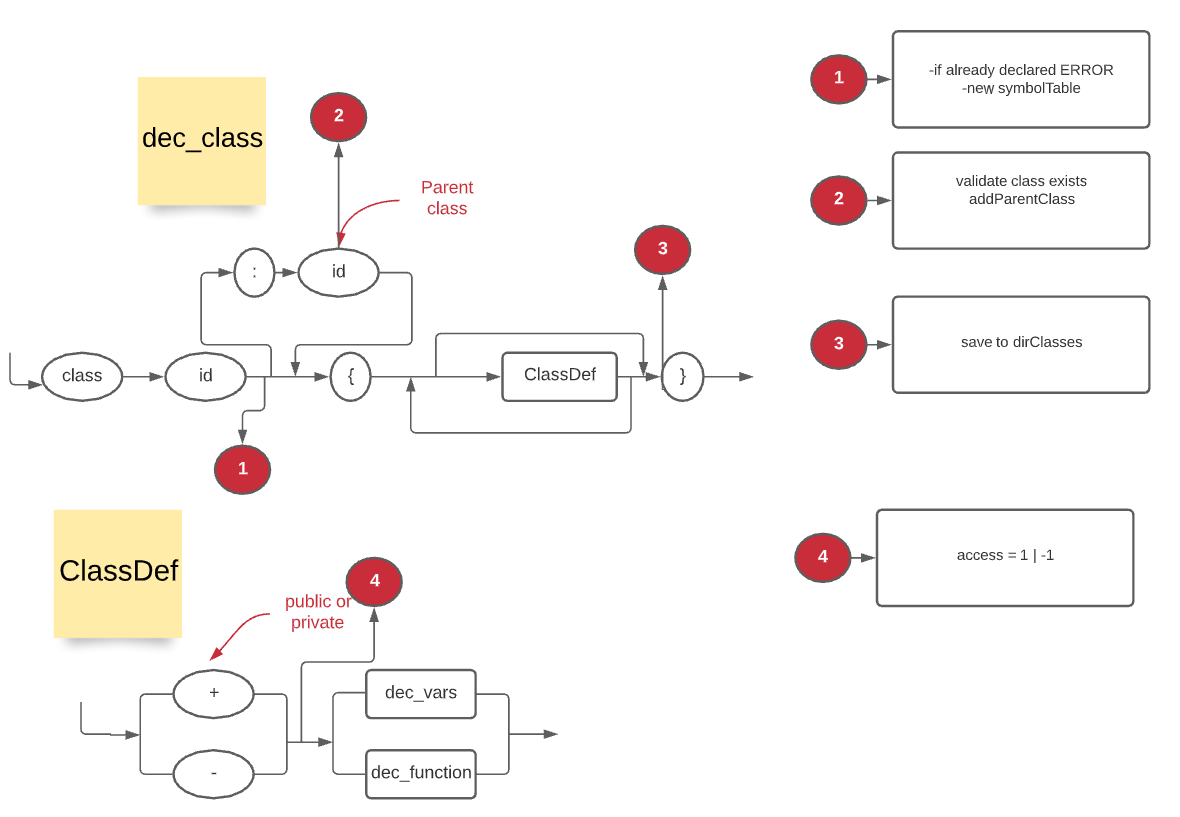
### Diagramas de sintaxis con las acciones correspondientes marcadas sobre ellos

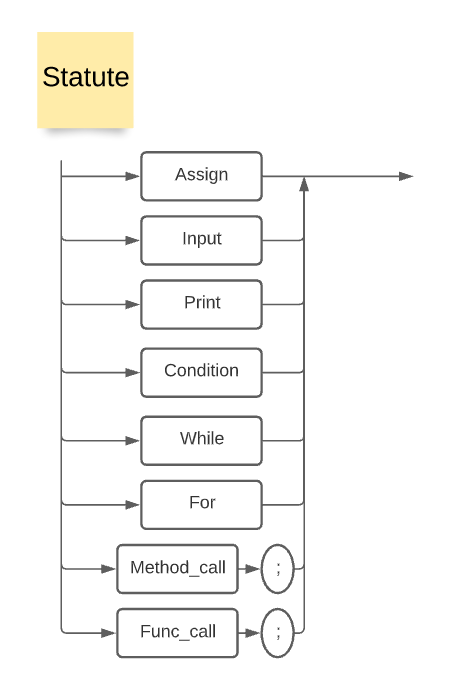
Link para lucidChart: [AGRO: Lucidchart](https://lucid.app/lucidchart/877200bf-d8c3-4432-bfdd-bdef3a8dfc60/view?page=XU9rVEFlgjx5) (diagramas en páginas)

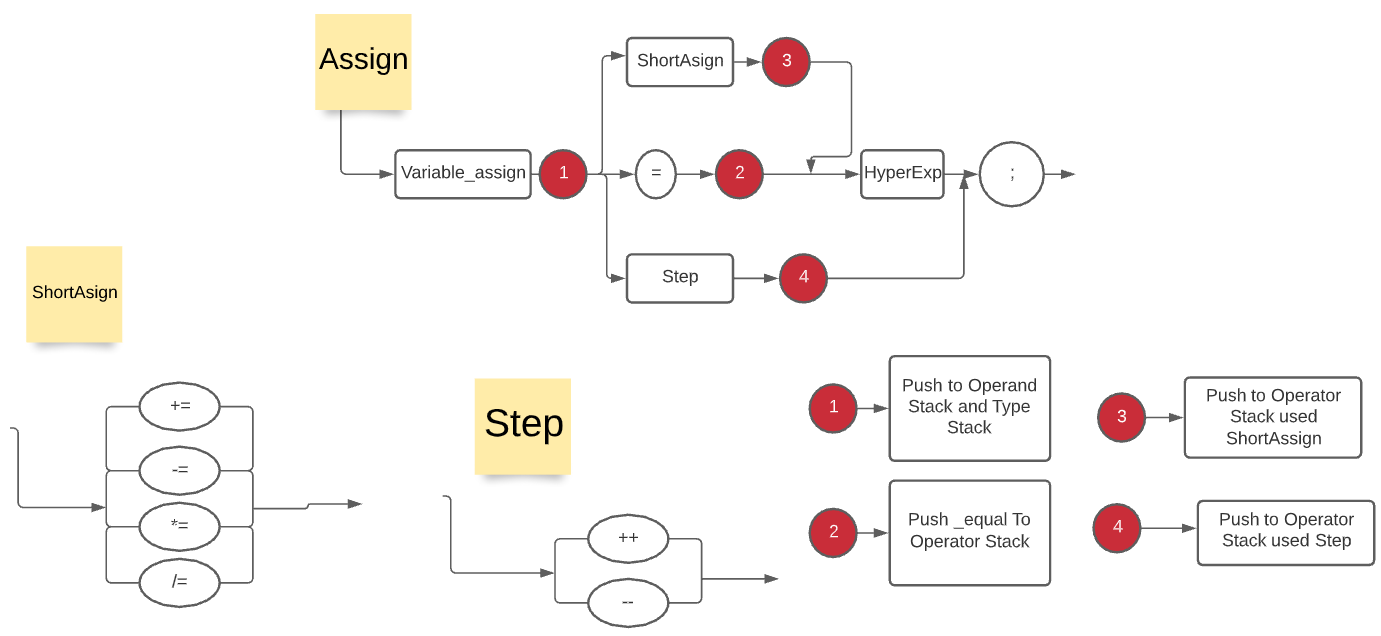


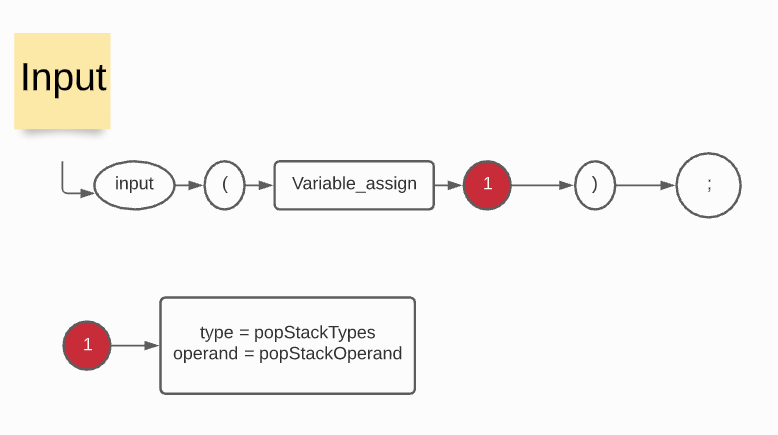


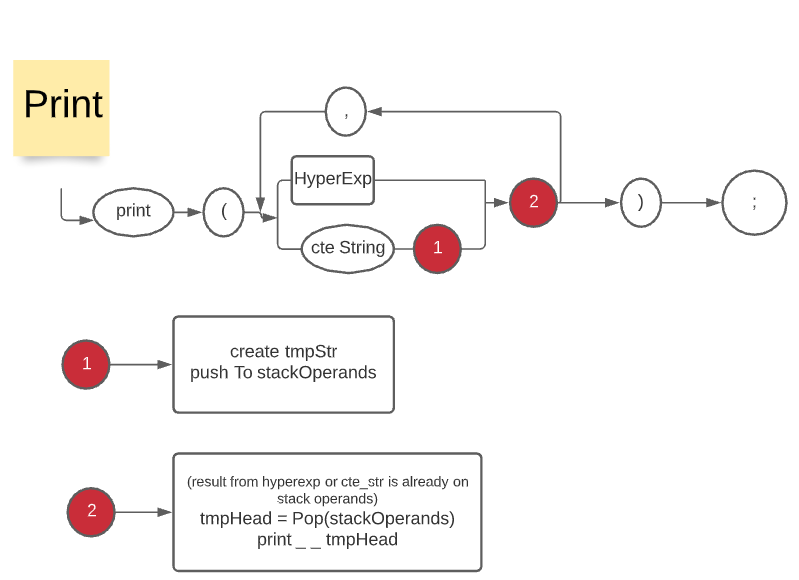


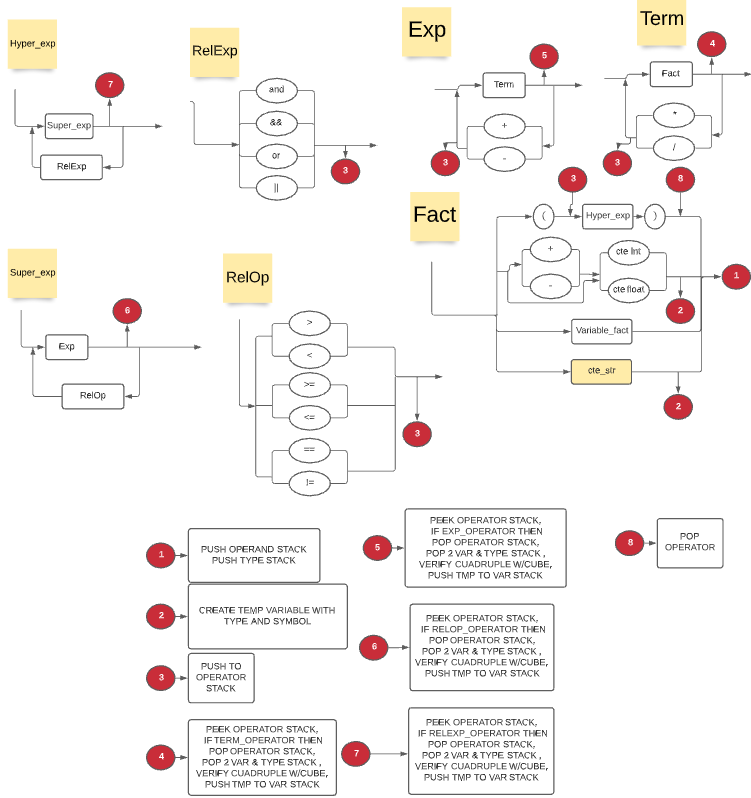


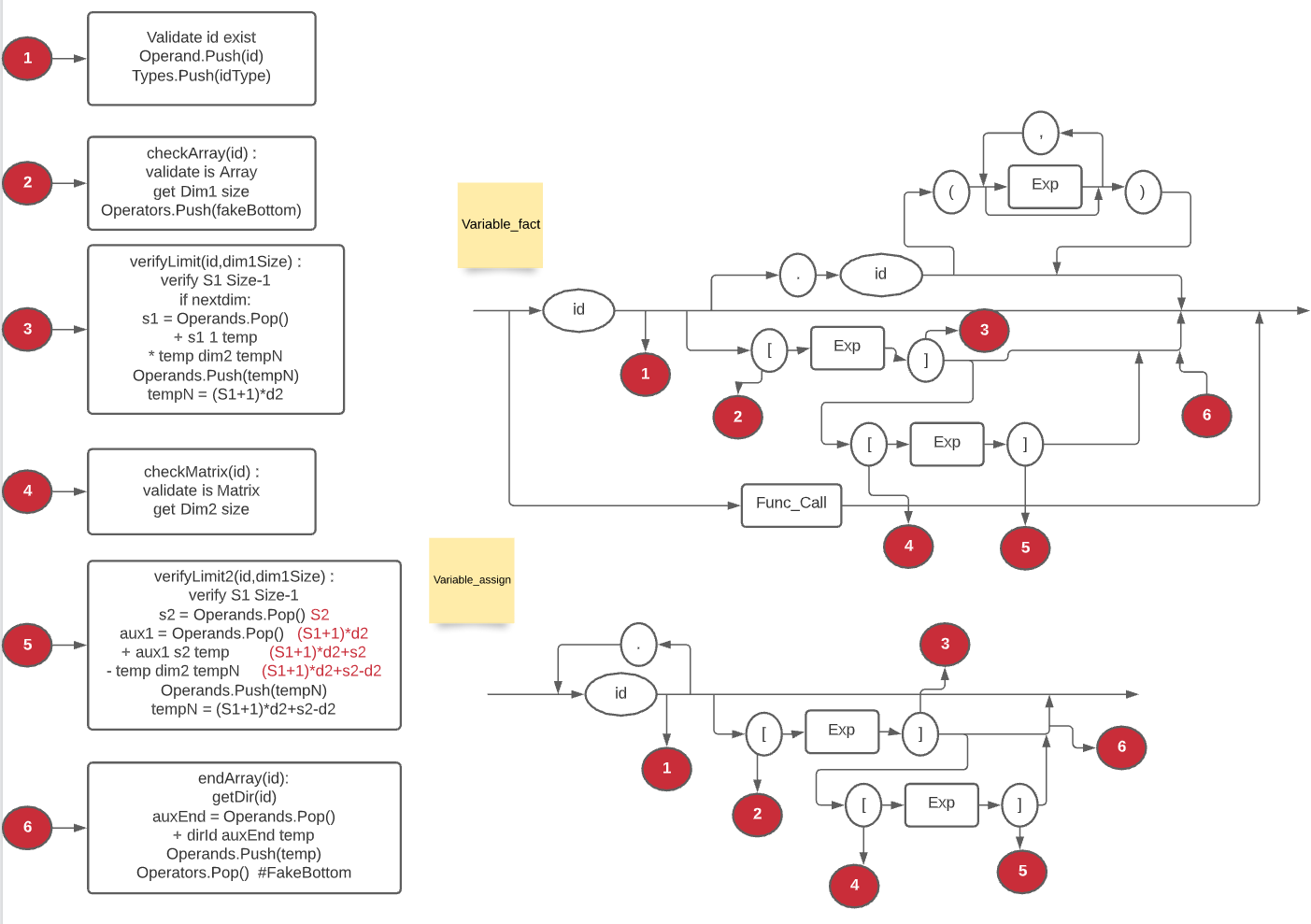


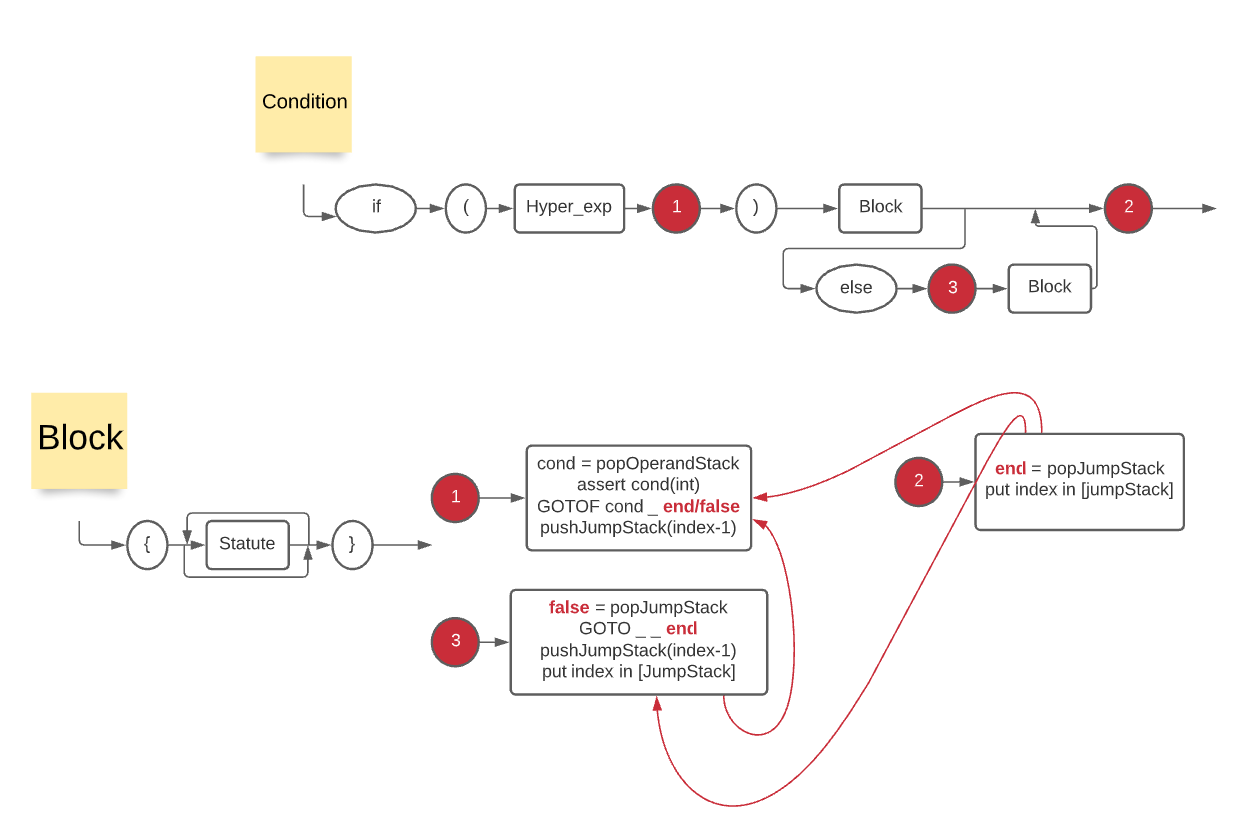


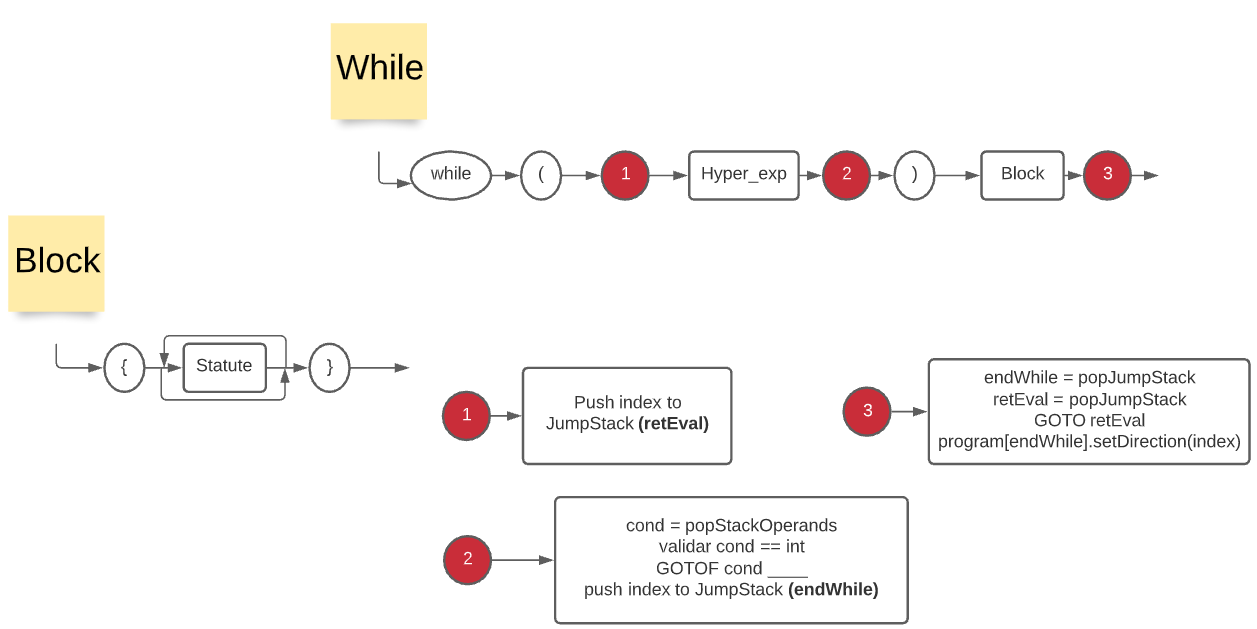


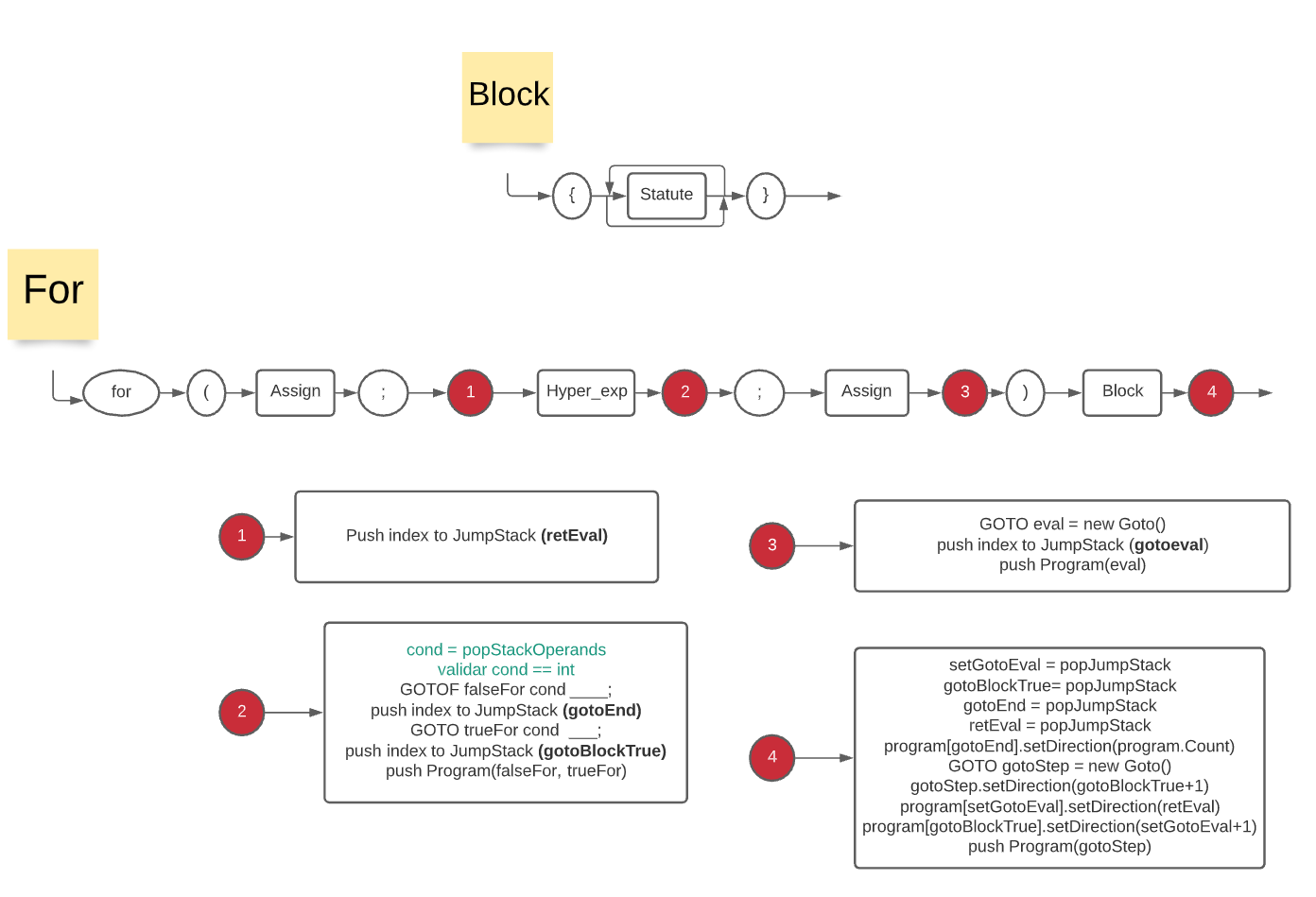


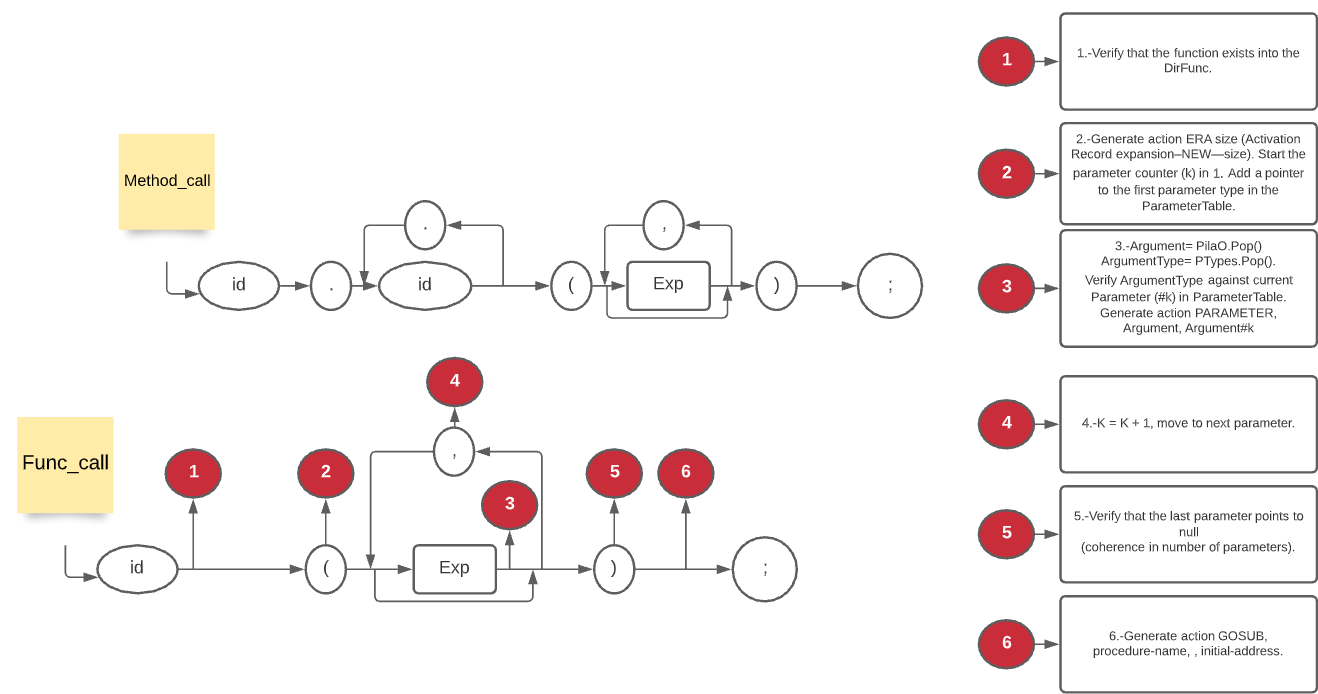












### Tabla de consideraciones semánticas (combinaciones factibles y errores de tipo)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **+, -, +=, -=** | **RightOper** | | | |
| **LeftOper** |  | **Int** | **Float** | **Char** |
| **Int** | int | int | int |
| **Float** | float | float | Error |
| **Char** | char | Error | char |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **\*, /, \*=, /=** | **RightOper** | | | |
| **LeftOper** |  | **Int** | **Float** | **Char** |
| **Int** | int | float | Error |
| **Float** | float | float | Error |
| **Char** | Error | Error | char |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **>, <, >=, <=, ==, ¡=** | **RightOper** | | | |
| **LeftOper** |  | **Int** | **Float** | **Char** |
| **Int** | int | int | Error |
| **Float** | int | int | Error |
| **Char** | Error | Error | int |
| **and, or** | **RightOper** | | | |
| **LeftOper** |  | **Int** | **Float** | **Char** |
| **Int** | int | Error | Error |
| **Float** | Error | Error | Error |
| **Char** | Error | Error | Error |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **\*\*** | **RightOper** | | | |
| **LeftOper** |  | **Int** | **Float** | **Char** |
| **Int** | int | float | Error |
| **Float** | float | float | Error |
| **Char** | Error | Error | Error |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **//, %** | **RightOper** | | | |
| **LeftOper** |  | **Int** | **Float** | **Char** |
| **Int** | int | int | Error |
| **Float** | int | int | Error |
| **Char** | Error | Error | Error |

## Descripción detallada del proceso de Administración de memoria en la compilación

En nuestro compilador, usamos una administración de memoria con una arquitectura jerárquica, en la que iniciamos con una tabla de variables global y cada que entramos en un nuevo contexto (clase, función, main) creamos una nueva tabla de variables, que tiene un apuntador a la tabla anterior a la que se crea.

Las direcciones de variables, al ser parte de las tablas de variables, se van “limpiando” en cada contexto, siendo la única excepción las direcciones de constantes, las que nos aseguramos que sean insertadas en la tabla de variables global sin importar de en qué contexto se declaran.

Las direcciones de cada tabla se van modificando siempre que se use un nuevo símbolo de cualquier tipo en el contexto de esa tabla, de acuerdo al tamaño del símbolo que se agregue (en caso de arreglos o matrices).

Si se quiere usar un símbolo de la tabla de variables, se hace una búsqueda en la tabla del contexto actual, y si no se encuentra el símbolo, se hace una búsqueda en la tabla del contexto “padre”, siendo así posible usar variables globales en cualquier contexto.

Cuando se crea un objeto en un contexto, se consulta el directorio de clases, para crear todos los atributos de ese objeto en el contexto actual, pero considerando el acceso de los atributos.

En el caso de directorio de funciones y clases, solamente se tienen declarados de manera global, siendo así que, para la declaración de métodos de clases, estos se guardan en el mismo directorio de funciones global, solamente que el nombre se estructura como ***nombreClase+”.\_” +nombreMétodo***

### Especificación gráfica de CADA estructura de datos usada

|  |  |  |
| --- | --- | --- |
| Estructura | Estructura de datos |  |
| Tabla de variables | Int[] variableDirections  SymbolTable parentSymbolTable  Dictionary<string, int[]> symbols  Dictionary<string, Dictionary<string, int>> objects  Dictionary<string, string> objectClasses  Int id |  |
| Directorio funciones | Dictionary<string, Function> |  |
| Clase Function | Clase creada, con:  -int variableCount  -int tmpCount  -int quadIndex  -List<int> parameterTypes |
| Directorio de clases | Dictionary <string, Classes> |  |
| Clase Classes | Clase creada, con:  -int quadIndex  -int variableCount  -int methodCount  -Dictionary <string, int[]> |
| Cuádruplos | Clases:  Assign, Cuadruple, GoSub, Goto, InOut, Param, Return, Verify |  |
| Pilas (Operandos, Operadores, Tipos, Saltos) | *Stack<String>* **stackOperand**  *Stack<int>* **stackOperator**  *Stack<int>* **stackTypes**  *Stack<int>* **stackJumps** |  |
| Cubo Semántico | Dictionary<int, Dictionary<int, Dictionary<string, int>>> |  |
| Auxiliar trad operador | Dictionary<int, string> |  |
| Auxiliar trad tipos | Dictionary<int, string> |  |

# **Descripción de la máquina virtual**

## Equipo de cómputo, lenguaje y utilerías especiales usadas ( en caso de ser diferente que el compilador)

### Equipo de cómputo

Windows, Linux

### Lenguaje

Python 3

### Utilerías de Python 3 usadas

* re: Expresiones regulares
* operator: funciones con operadores disponibles en Python 3
* sys: biblioteca para leer parámetros al iniciar el programa en consola

## Descripción detallada del proceso de Administración de memoria en ejecución (arquitectura)

La memoria en la máquina virtual consta de múltiples contextos de memoria donde cada uno incluye sus propias estructuras para almacenar los tipos de variables más simples como int, float, char, string.

Los contextos disponibles en memoria son:

* global
* global temporales
* local
* local temporales
* constantes
* apuntadores

Cuando se llama un método se asigna el MemoryContext necesario según un directorio de funciones. Este MemoryContext se pone en la pila de localMemory y localTempMemory respectivamente. Al terminar la llamada de una función, se borran estos MemoryContext de la pila.

### Especificación gráfica de CADA estructura de datos usada para el manejo de scopes (Memoria local, global, etc.…)



### Asociación hecha entre las direcciones virtuales (compilación) y reales (ejecución)

Cada vez que se accede a una posición en memoria, se tienen definidos previamente los rangos numéricos en que una dirección pertenece a cada tipo de contexto en Memory y cada tipo de variable de MemoryContext (int, float, char, string).

En el caso de la memoria localMemory y localTempMemory, se accede al MemoryContext que esté hasta arriba de la pila de estas estructuras, ya que se refiere siempre al contexto local actual.

# **Pruebas del funcionamiento del lenguaje**

## Incluir pruebas que “comprueben” el funcionamiento del proyecto

### Codificación de la prueba (en AGRO)

### Resultados arrojados por la generación de código intermedio y por la ejecución.

# **Documentación del código del proyecto**