



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

Trabajo Práctico I

Grupo: Parsers para todos

Teoría de Lenguajes
Segundo Cuatrimestre de 2016

Integrante	LU	Correo electrónico
Federico De Rocco	408/13	fede.183@hotmail.com
José Massigoge	954/12	jmmassigoge@gmail.com
Leandro Anacondio	487/07	leandro.anacondio@gmail.com



Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359

<http://www.fcen.uba.ar>

Índice

1. Introducción	3
2. Gramática	4
3. Lexer	6
4. Parser	7
5. Test	9
5.1. Figura: Rectangle	9
5.2. Figura: Line	9
5.3. Figura: Circle	10
5.4. Figura: Ellipse	11
5.5. Figura: Polyline	12
5.6. Figura: Polygon	12
5.7. Errores Semánticos	14
6. Conclusiones	16

1. Introducción

En el presente Trabajo Práctico diseñamos una gramática para el lenguaje *Dibu* e implementamos su analizador léxico y su analizador sintáctico y semántico (parser). Utilizando estas herramientas creamos un programa que permite traducir de *Dibu* a *SVG* y compilar dicha traducción, en caso de que sea válida. A lo largo del presente informe vamos a describir la gramática para el lenguaje, el lexer y el parser que implementamos.

2. Gramática

En la siguiente sección vamos a mostrar y describir la gramática construida para el lenguaje *Dibu*, llamaremos a esta gramática G :

Producciones:
$$P \rightarrow S P \mid \lambda$$
$$S \rightarrow \text{id PARAMS}$$
$$\text{PARAMS} \rightarrow \text{id} = V, \text{PARAMS} \mid \text{id} = V$$
$$V \rightarrow \text{num} \mid \text{string} \mid (\text{num}, \text{num}) \mid [\text{ARRAY}]$$
$$\text{ARRAY} \rightarrow (\text{num}, \text{num}), \text{ARRAY} \mid (\text{num}, \text{num})$$
$$G = \{\{P, S, \text{PARAMS}, V, \text{ARRAY}\}, \{\text{num}, \text{string}, (, , ,), [,], =, \text{id}\}, \text{Descripto por la gramática}, P\}$$

Como se describe en el enunciado, el lenguaje *Dibu* es una serie de instrucciones de la forma:

IDENTIFICADOR PARAM1=V1, PARAM2=V2, ..., PARAMN=VN

Nuestra gramática consiste en cinco símbolos no terminales, que generan 11 producciones:

- Las producciones del no terminal ARRAY describen la secuencia de pares numericos que van dentro de un arreglo.
- Las producciones del no terminal V (valor) contiene las combinaciones de terminales que se pueden esperar como valores de los parámetros (num, string, point, array).
- PARAMS posee la secuencia compuesta por: parámetro = valor.
- S describe una figura, con su nombre y parámetros.
- P describe la serie compuesta de las instrucciones descriptas en el no terminal S.

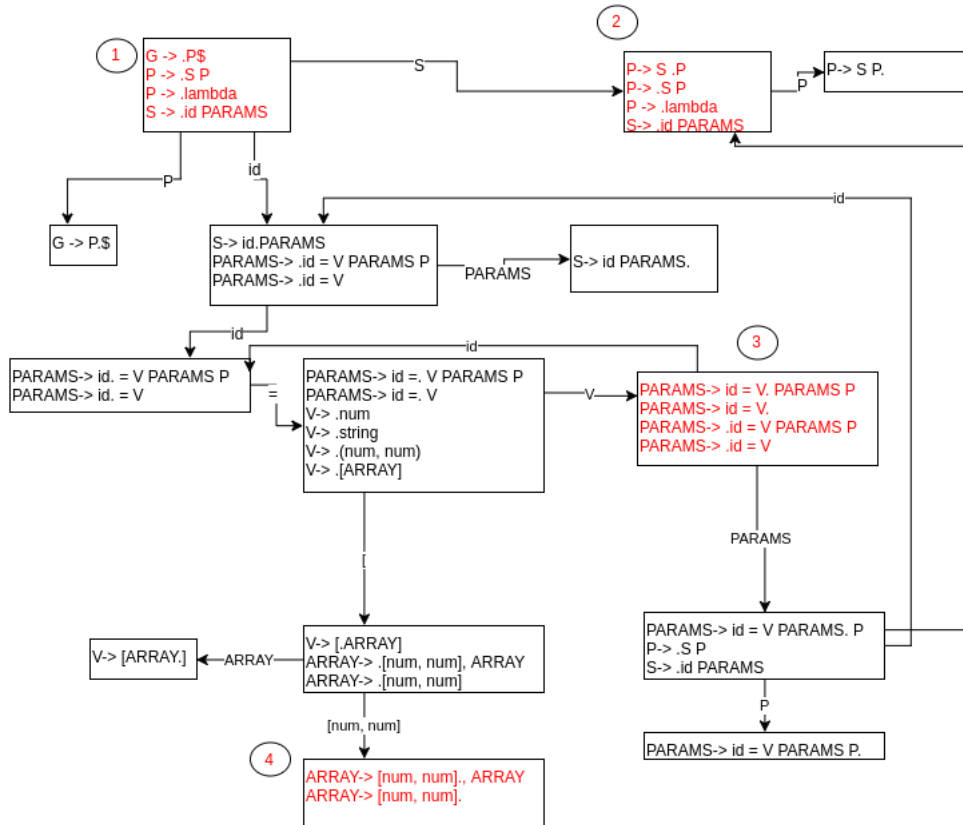
Con estas producciones se puede ver que nuestra gramática describe precisamente la serie de instrucciones de *Dibu*. Se debe aclarar que esta gramática no respeta todas las restricciones de *Dibu*, como por ejemplo que solo aparezca la instrucción size una vez. Estas cuestiones se tratarán en el lexer y el parser.

Por otro lado tenemos que definir que tipo de gramática es G . Para lograr este objetivo no realizaremos una demostración rigurosa aunque si analizaremos la misma descartando, en primera instancia, que la gramática sea ambigua. Posteriormente veremos a que tipos de gramáticas no pertenece G .

Esta gramática no es LL(1) porque en las producciones de PARAMS, por ejemplo, sucede que la intersección de símbolos distinguidos no es vacía:

$$SD(\text{PARAMS} \rightarrow \text{id} = V, \text{PARAMS}) \cap SD(\text{PARAMS} \rightarrow \text{id} = V) = \text{id}.$$

La gramática tampoco es LR(0). A continuación se mostrara el autómata correspondiente a la gramática, marcando los conflictos. Solamente se consideraron los estados del autómata que fueran interesantes por sus conflictos.



La gramática es SLR ya que conservando el mismo autómata vemos que los siguientes de P y ARRAY que son los implicados en las nodos que generan los conflictos son:

$$\text{Siguientes}(P) = \{\$ \} \quad \text{Siguientes}(\text{ARRAY}) = \{] \}$$

Con lo cual los nodos 1 y 2 que presentaban conflictos por el lambda quedan resueltos debido a que solamente se reducirá por \$ y no por id. Por último el nodo conflictivo 3 solo reducirá con] por lo que se eliminan todos los conflictos shift/reduce.

3. Lexer

En esta sección vamos a explicar el analizador léxico implementado para la gramática descrita anteriormente. Al aplicar un lexer sobre una cadena de caracteres este devolverá verdadero si la misma no contiene errores de carácter léxico como por ejemplo caracteres que el lenguaje no reconoce. Para llevar a cabo este analizador contamos con una serie de tokens de los cuales se da una descripción de sus posibles contenidos. Estos son:

NUMBER: Son todos los números compuestos por los caracteres numéricos del 0 al 9 y el punto. El lexer identifica el tipo del valor numérico siendo estos enteros y de punto flotante de acuerdo a si el número posee o no el punto. Esto se hace para el posterior control de errores.

Expresión Regular: $(0|..|9)^+|(0|..|9)^+?$

STRING: Cadenas de caracteres entre comillas compuestas por las letras de la a a la z, incluidas mayúsculas, los números del 0 al 9 y los símbolos +, -, y *, la coma y el espacio (este último lo representamos con la palabra espacio).

Expresión Regular: $"(a|..|z|A|..|Z| + | * | - |, |(espacio)?)(a|..|z|A|..|Z| + | * | - |, |0|..|9|(espacio)?)"$

ID: Igual que STRING pero sin las comillas.

Expresión Regular: $(a|..|z|A|..|Z| + | * | - |)^+(a|..|z|A|..|Z| + | * | - |0|..|9)^*$

LPAREN: El carácter [.

RPAREN: El carácter].

LBRACKET: El carácter (.

RBRACKET: El carácter).

COMMA: El carácter ,.

EQUALS: El carácter =.

IGNORE: El carácter vacío.

ERROR: Este aparece cuando se encuentra un token desconocido. En consecuencia se guarda la línea, valor, posición y tipo de valor para el error. Con este último cumplimos que si la cadena es inválida se detectara el error y se informará de él.

Como se puede ver con esta serie de tokens y sus respectivas implementaciones podemos describir todos los posibles no terminales y terminales necesarios para nuestra gramática. Quedan todavía restricciones como por ejemplo que ID solamente puede tener como valores a los identificadores de las instrucciones.

4. Parser

En esta sección vamos a explicar el analizador sintáctico o parser implementado. El análisis sintáctico convertirá al texto de entrada en el árbol de derivación pertinente. A partir de ello hacemos el análisis de las sentencias, utilizando estructuras auxiliares para cumplir posteriormente con el requerimiento de la generación del texto de salida (en formato svg).

La idea general fue definir una función que represente cada una de las producciones de nuestra gramática, y dentro de cada una ir manipulando la información para luego generar el output requerido. Como atributos utilizamos `lineno` y `lexpos` para detallar el número de línea y posición respectivamente, mientras que el atributo `value` nos devuelve el valor. Como estructuras auxiliares utilizamos un diccionario y una lista. El primero representa los parámetros obligatorios de cada una de las figuras, lo utilizamos dentro de la producción `STATE` para chequear que todos los parámetros obligatorios están dentro de la cadena de entrada para cada figura en particular. El segundo es simplemente una lista donde iremos acumulando las distintas figuras que se fueron generando, teniendo en cuenta que nuestro parser es bottom-up, las figuras se irán generando a partir de las hojas y una vez que se llegue a la producción inicial, tendremos la lista llena de las figuras que debemos imprimir. Agregamos una función que hiciera las veces de producción inicial solo a fines prácticos, para poder generar el lienzo final, e ir agregando las figuras que posteriormente dibujaremos. A continuación haremos un análisis de cada una de esas funciones para explicar cuál es su rol.

P.START: genera el lienzo llamando a la función `Scene` (definida en la clase `figuras.py`) y luego le agrega las figuras de la lista al mismo. También chequea que no se haya llamado a la función `size` mas de una vez.

P.PROGRAM.EMPTY: representa al programa que se genera a partir de la producción $P \mid \lambda$

P.PROGRAM.NONEMPTY: representa al programa que se genera a partir de la producción $P \rightarrow S P$

P.STATE: Es la función más compleja del parser. Primero chequea que el token sea `size`, si lo es, se revisa que tenga el `height` y `width` que son los parámetros requeridos. Si el token no es `size` estamos ante la producción para generar una figura, por ello, se guarda los parámetros recolectados por los nodos hijos en una variable, e inicializa la figura como objeto (correspondiente al nombre de la misma, dentro de una función auxiliar). Si la generación no lanza una excepción, va completando los atributos requeridos por esa figura y la agrega a la lista de resultados. Si no se genera bien la figura o si los parámetros son incorrectos se lanza la excepción correspondiente.

P.PARAMS_NONRECURSIVE: Representa la asignación de un valor a un parámetro sin recursión, es decir es el último de la lista o es único. Dentro de esta función se agrega el número de línea, posición y valor del parámetro al diccionario de parámetros.

P.PARAMS_RECURSIVE: Representa la seguidilla de parámetros separados por coma. Chequea que no haya repetidos y los va agregando a la lista de parámetros.

P.VALOR.NUMBER: Realiza la asignación de un valor a un parámetro de tipo numérico.

P.VALOR.STRING: Realiza la asignación de un valor a un parámetro de tipo string.

P.VALOR.POINT: Realiza la asignación de un valor a un parámetro de tipo punto, es decir dos números separados por coma.

P.VALOR.ARRAY: Realiza la asignación de un valor a un parámetro de tipo arreglo.

P.ARRAY.ELEMENT: Representa el array con un único elemento, o el elemento final de un arreglo.

P.ARRAY.RECURSIVE: Genera un nuevo elemento en un array y lo appendea a los elementos que siguen en la producción.

P.ERROR: Define los errores sintácticos que ocurrieron. A partir del token genera el mensaje correspondiente.

Nuestra implementación utiliza Scene, una clase pensada para construir una cadena de SVG a partir de los datos de los parámetros que tomamos de la cadena de *Dibu*. Después de declarar nuestra Scene se puede agregar líneas de figuras utilizando las funciones del mismo nombre que la figura y los parámetros. Al final, la cadena de Scene es una cadena SVG la cual podemos utilizar para generar la imagen pedida.

Nuestro parser actúa como una traducción dirigida por sintaxis que se encargara de analizar la cadena hasta encontrar algún error, como por ejemplo dos size en una cadena. Para esto se recurrirá a los siguientes atributos:

lineno: Es un número entero que indica la línea actual. Lo poseen num, string, y todos los no-terminales.
lexpos: Es un número entero que indica la posición en la línea. Lo poseen num, string, y todos los no-terminales.

value: Indica el valor según el contexto en que se encuentra. Lo poseen num, string y todos los no-terminales excepto por PARAMS.

parámetros: Es un diccionario que asocia a los nombres de los parámetros con su valor en una línea. Lo tiene PARAMS.

Además de estos atributos, tenemos dos diccionarios. El primero es diccSize que usamos para guardarnos los size's que aparecen en la cadena, en el caso de que no exista ninguno entonces se usará el Scene sin especificar un tamaño. Si hay mas de uno entonces estamos ante un error semántico que debemos avisar. El segundo de estos diccionarios es el de nombre idDicObligatory que sirve para poder guardar de forma estática los parámetros obligatorios de cada figura. Esto sirve para comprobar si en nuestra cadena todas las figuras tienen al menos todos sus parámetros obligatorios.

Por último tenemos a listaFiguras la cual guarda las figuras construidas para posteriormente agregarse al Scene.

Utilizando los atributos y las estructuras adicionales cubrir todos los casos en el lenguaje *Dibu* tales como la repetición de size, que si o si cada figura tenga sus parámetros obligatorios una vez por cada uno y sin orden, lo mismo para los opcionales pero sin que tengan que estar todos y además el poder construir la cadena SVG usando Scene.

El programa en si toma un archivo de texto que contiene una cadena del lenguaje *Dibu* y el nombre que va a tener la imagen resultante del análisis y traducción del primer parámetro. Tenemos un *try-catch*, el cual usaremos para poder capturar los errores propios de una cadena de entrada invalida y posteriormente mostrarlos por pantalla dando la posición, línea y el tipo de error(Sintáctico, Semántico y de Tokens). Una vez generada la cadena traducida de SVG mediante el parser, se la ejecutara generando la imagen y guardándola con el nombre pasado como parámetro.

5. Test

A continuación mostramos los resultados obtenidos de realizar test de las diversas figuras, especificando cual era el código *Dibu* de entrada, el código *SVG* de salida y la imagen generada.

5.1. Figura: Rectangle

Código *Dibu* de entrada:

```
size height=400, width= 1200
rectangle size=(398, 1198), upper_left=(1,1), fill="none", stroke="blue", stroke-width=2
rectangle upper_left=(400,100), fill="yellow", size=(200, 400), stroke-width=10, stroke="navy"
```

Código *SVG* de salida:

```
<?xml version="1.0"?>
<svg height="400" width="1200" >
  <g style="fill-opacity:1.0; stroke:black; stroke-width:1;">
    <rect x="1" y="1" height="398" width="1198" style="fill:none;stroke:blue;stroke-width:2" />
    <rect x="400" y="100" height="200" width="400" style="fill:yellow;stroke:navy;stroke-width:10" />
  </g>
</svg>
```

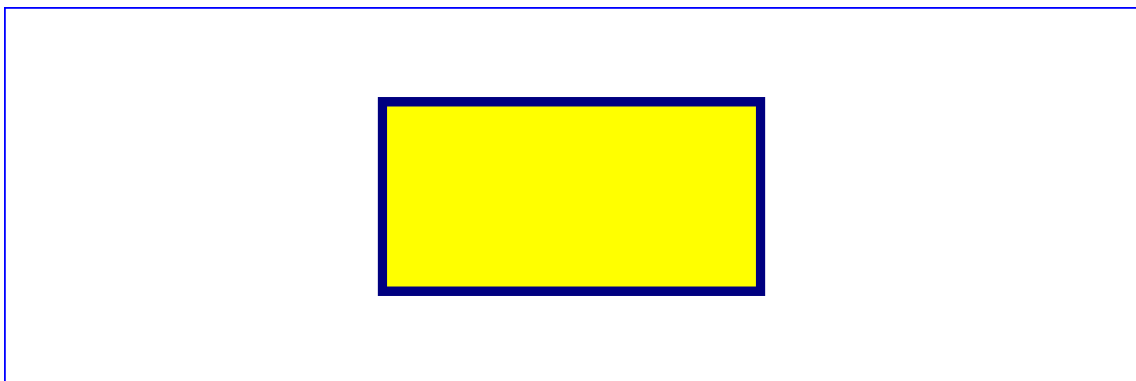


Figura 1: Imagen generada - Figura Rectangle

5.2. Figura: Line

Código *Dibu* de entrada:

```
size width=1200, height=400
rectangle size=(398, 1198), upper_left=(1,1), fill="none", stroke="blue", stroke-width=2
line from=(100,300), to=(300,100), stroke-width=5, stroke="green"
line from=(300,300), to=(500,100), stroke-width=10, stroke="green"
line from=(500,300), to=(700,100), stroke-width=15, stroke="green"
line from=(700,300), to=(900,100), stroke-width=20, stroke="green"
line from=(900,300), to=(1100,100), stroke-width=25, stroke="green"
```

Código *SVG* de salida:

```
<?xml version="1.0"?>
<svg height="400" width="1200" >
  <g style="fill-opacity:1.0; stroke:black;
stroke-width:1;">
    <rect x="1" y="1" height="398"
      width="1198" style="fill:none;stroke:blue;stroke-width:2" />
    <line x1="100" y1="300" x2="300" y2="100" style="fill:black;
      stroke:green;stroke-width:5"/>
    <line x1="300" y1="300" x2="500" y2="100" style="fill:black;
      stroke:green;stroke-width:10"/>
    <line x1="500" y1="300" x2="700" y2="100" style="fill:black;
      stroke:green;stroke-width:15"/>
    <line x1="700" y1="300" x2="900" y2="100" style="fill:black;
      stroke:green;stroke-width:20"/>
    <line x1="900" y1="300" x2="1100" y2="100" style="fill:black;
      stroke:green;stroke-width:25"/>
  </g>
</svg>
```

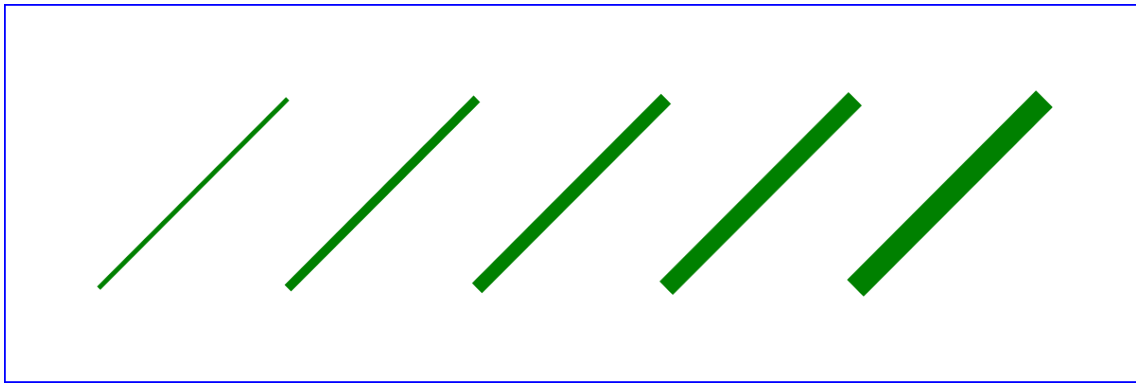


Figura 2: Imagen generada - Figura Linea

5.3. Figura: Circle

Código Dibu de entrada:

```
size width=1200, height=400
rectangle size=(398, 1198), upper_left=(1,1), fill="none", stroke="blue
", stroke-width=2
circle center=(600, 200), fill="red", stroke="blue", stroke-width=10,
radius=100
```

Código SVG de salida:

```
<?xml version="1.0"?>
<svg height="400" width="1200" >
  <g style="fill-opacity:1.0; stroke:black;
stroke-width:1;">
    <rect x="1" y="1" height="398"
      width="1198" style="fill:none;stroke:blue;stroke-width:2" />
    <circle cx="600" cy="200" r="100"
      style="fill:red;stroke:blue;stroke-width:10" />
  </g>
</svg>
```

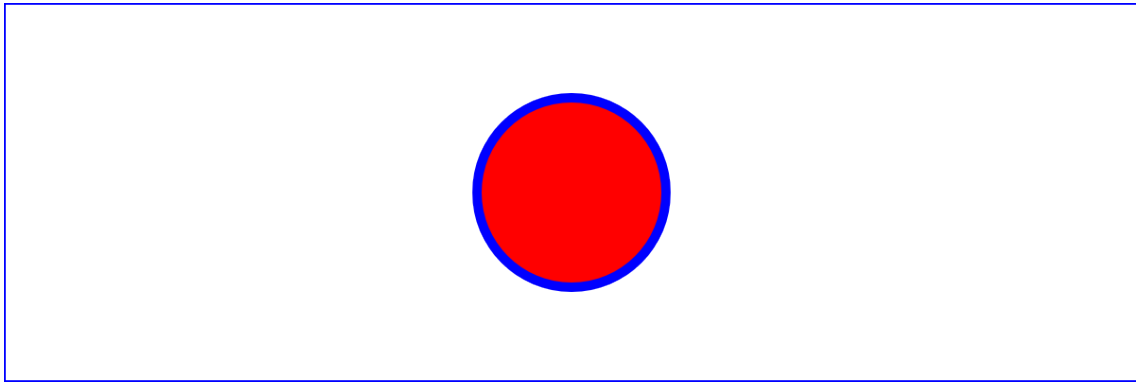


Figura 3: Imagen generada - Figura Circulo

5.4. Figura: Ellipse

Código *Dibu* de entrada:

```
size width=1200, height=400
rectangle size=(398, 1198), upper_left=(1,1), fill="none", stroke="blue
", stroke-width=2
ellipse center=(300, 200), fill="red", ry=100, rx=250
ellipse ry=100, fill="none", rx=250, center=(900, 200), stroke="blue",
stroke-width=20
```

Código SVG de salida:

```
<?xml version="1.0"?>
<svg height="400" width="1200" >
  <g style="fill-opacity:1.0; stroke:black;
stroke-width:1;">
    <rect x="1" y="1" height="398"
      width="1198" style="fill:none;stroke:blue;stroke-width:2" />
    <ellipse cx="300" cy="200" rx="250" ry="100"
      style="fill:red;stroke:none;stroke-width:1"/>
    <ellipse cx="900" cy="200" rx="250" ry="100"
      style="fill:none;stroke:blue;stroke-width:20"/>
  </g>
</svg>
```

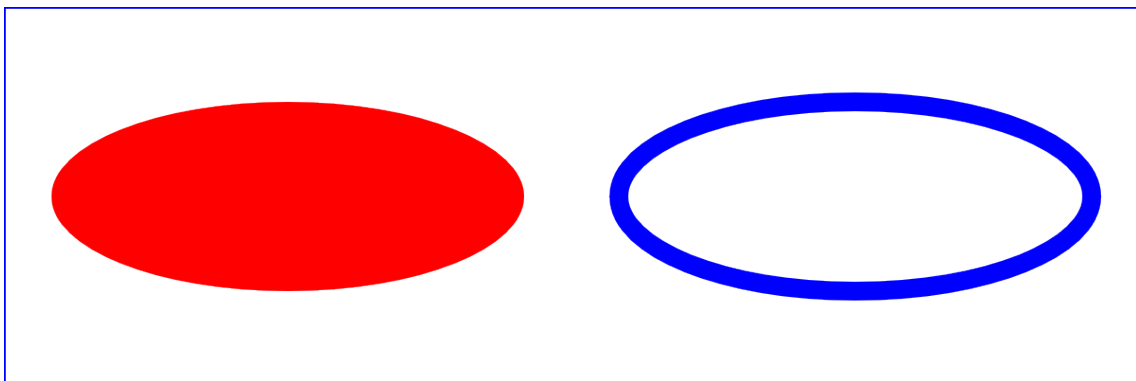


Figura 4: Imagen generada - Figura Elipse

5.5. Figura: Polyline

Código *Dibu* de entrada:

```
size width=1200, height=400
rectangle size=(398, 1198), upper_left=(1,1), fill="none", stroke="blue
", stroke-width=2
polyline points=[(50,375), (150,375), (150,325), (250,325), (250,375),
(350,375), (350,250), (450,250), (450,375), (550,375), (550,175),
(650,175), (650,375), (750,375), (750,100), (850,100), (850,375),
(950,375), (950,25), (1050,25), (1050,375), (1150,375)], stroke="blue
", stroke-width=10, fill="none"
```

Código SVG de salida:

```
<?xml version="1.0"?>
<svg height="400" width="1200" >
  <g style="fill-opacity:1.0; stroke:black;
stroke-width:1;">
    <rect x="1" y="1" height="398"
      width="1198" style="fill:none;stroke:blue;stroke-width:2" />
    <polyline points=" 50,375 150,375 150,325 250,325 250,375 350,375
      350,250 450,250 450,375 550,375 550,175 650,175 650,375
      750,375 750,100 850,100 850,375 950,375 950,25 1050,25
      1050,375 1150,375"
      style="fill:none;stroke:blue;stroke-width:10"/>
  </g>
</svg>
```

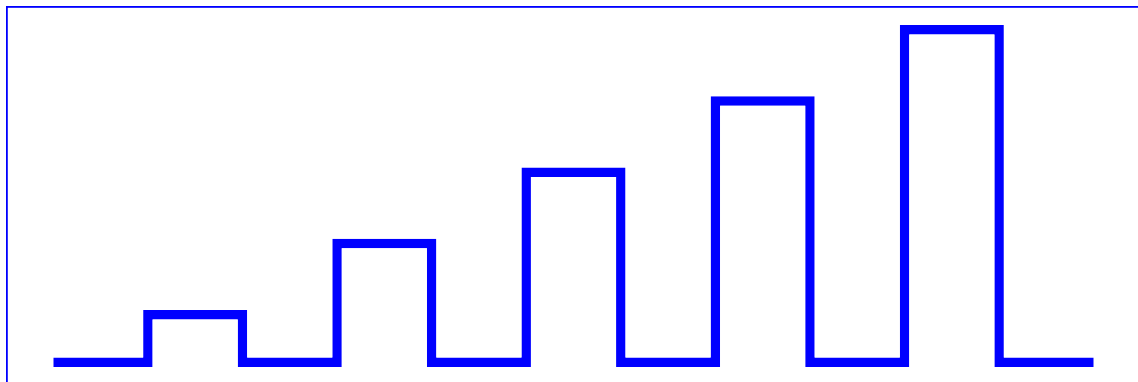


Figura 5: Imagen generada - Figura Polyline

5.6. Figura: Polygon

Código *Dibu* de entrada:

```
size width=1200, height=400
rectangle size=(398, 1198), upper_left=(1,1), fill="none", stroke="blue
", stroke-width=2
polygon fill="red", stroke="blue", stroke-width=10, points=[(350,75),
(379,161), (469,161), (397,215), (423,301), (350,250), (277,301),
(303,215), (231,161), (321,161)]
polygon fill="lime", stroke="blue", stroke-width=10, points=[(850,75),
(958,137.5), (958,262.5), (850,325), (742,262.6), (742,137.5)]
```

Código SVG de salida:

```
<?xml version="1.0"?>
<svg height="400" width="1200" >
  <g style="fill-opacity:1.0;stroke:black;
stroke-width:1;">
    <rect x="1" y="1" height="398"
      width="1198" style="fill:none;stroke:blue;stroke-width:2" />
    <polygon points="350,75 379,161 469,161 397,215 423,301 350,250 277,301 303,215 231,161 321,161"
      style="fill:red;stroke:blue;stroke-width:10"/>
    <polygon points="850,75 958,137 958,262 850,325 742,262 742,137"
      style="fill:lime;stroke:blue;stroke-width:10"/>
  </g>
</svg>
```

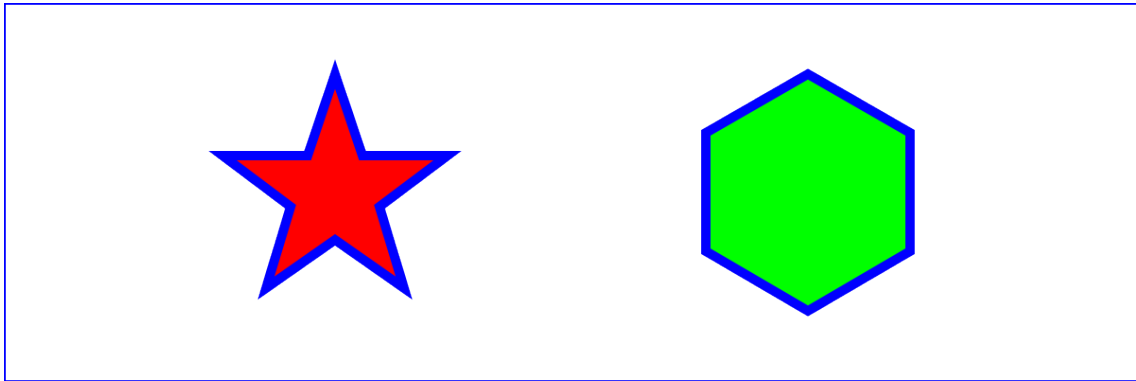


Figura 6: Imagen generada - Figura Poligono

5.7. Errores Semánticos

Otra de las cuestiones que nos pareció importante testear es la capacidad del parser de detectar errores semánticos. A partir del enunciado, concluimos que podrían darse los siguientes errores de esta índole:

1. **Dos o mas usos de la sentencia Size.**
2. **Nombre de figura invalida.**
3. **Parámetro invalido.**
4. **Ausencia de parámetros obligatorios.**
5. **Parámetros repetidos.**

Propusimos los siguientes test para testear esta característica:

1. **Dos o mas usos de la sentencia Size.**

Código *Dibu* de entrada:

```
size width=1200, height=400
rectangle upper_left=(1,1), fill="none", stroke="blue", stroke-
width=2, size=(200, 200)
size height=400, width=200
size height=800, width=800
rectangle upper_left=(50,50), fill="none", stroke="red", stroke-
width=2, size=(500, 800)
```

Texto emitido en el stdout:

```
Semantic error: Dos o Mas Size
line: 3
position: 116
```

2. **Nombre de figura invalida.**

Código *Dibu* de entrada:

```
cuadrado size=200, fill="red"
rectangle upper_left=(1,1), fill="none", stroke="blue", stroke-
width=2, size=(200, 200)
```

Texto emitido en el stdout:

```
Semantic error: Figura Invalida
line: 1
position: 9
```

3. **Parámetro invalido.**

Código *Dibu* de entrada:

```
line from=(100,300), to=(300,100), stroke-width=5, stroke="green"
line from=(300,300), to=(500,100), stroke-width=10, stroke="green"
, rotate=(90,90)
```

Texto emitido en el stdout:

```
Semantic error: Parametro Invalido  
line: 2  
position: 71
```

4. Ausencia de parámetros obligatorios.

Código *Dibu* de entrada:

```
line to=(300,100), stroke-width=5, stroke="green"  
line from=(300,300), to=(500,100), stroke-width=10, stroke="green"
```

Texto emitido en el stdout:

```
Semantic error: Faltan Parametros obligatorios  
line: 1  
position: 5
```

5. Parámetros repetidos.

Código *Dibu* de entrada:

```
line from=(100,100), to=(300,100), stroke-width=5, stroke="green",  
    stroke-width=3  
line from=(300,300), to=(500,100), stroke-width=10, stroke="green"
```

Texto emitido en el stdout:

```
Semantic error: Parametros repetidos  
line: 1  
position: 35
```

6. Conclusiones

El desafío de desarrollar un lenguaje desde su inicio, resulta interesante para comprender como se definen las reglas sintácticas y semánticas que conforman la estructura del mismo. A lo largo del trabajo tuvimos que repensar varias veces incluso los puntos que parecían más simples, como ser la definición de la gramática y el impacto que generaba cada decisión tomada. En el aspecto meramente implementativo también modificamos sobre la marcha las estructuras utilizadas y la forma en que utilizábamos los datos, en un proceso iterativo. Todo lo antes dicho lleva a pensar que incluso el desarrollo de lenguajes pequeños, conlleva un gran trabajo y cada decisión tomada impacta de forma directa en las etapas siguientes.