# Principios del Lenguajes de Programación

Clase 3 - Análisis Sintáctico

Sandra Roger

Facultad de Informática Universidad Nacional del Comahue

Primer Cuatrimestre

## Contenidos

Repaso

Introducción

Descripción de la sintaxis

Definición formal de lenguajes

Bibliografía

# ¿Qué vimos?

#### PROGRAMA ANALÍTICO:

**Unidad 1.** Introducción a los conceptos de los lenguajes de programación

Concepto del del lenguaje de programación. Buenos Lenguajes y Atributos. Historia y Evolución de los Lenguajes de Programación. Clasificación de lenguajes. Criterios de diseño y de implementación de un lenguaje de programación. Concepto de Paradigma: imperativo, funcional, lógico, orientado a objetos. Máquinas virtuales: Jerarquías. Sintaxis y semántica. Conceptos de Intérpretes y Compiladores. Técnicas formales de descripción sintáctica. Introducción a la Semántica Operacional, Axiomática y Denotacional. Atributos y Tiempos de Ligadura.

# ¿Qué vimos?

### Ortogonalidad:

Capacidad de combinar varias características de un lenguaje y que tengan significado.

- Más ortogonales entonces mayor cantidad de combinaciones de caracteres, por lo tanto menos excepciones y menos casos especiales que recorda
- Desventaja: Compilación sin errores con combinaciones lógicamente incoherente o ejecución ineficiente

# ¿Qué vimos?

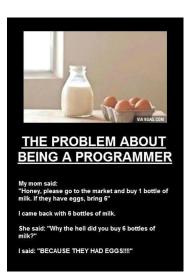
### Ortogonalidad:

Capacidad de combinar varias características de un lenguaje y que tengan significado.

- Más ortogonales entonces mayor cantidad de combinaciones de caracteres, por lo tanto menos excepciones y menos casos especiales que recordar
- Desventaja: Compilación sin errores con combinaciones lógicamente incoherente o ejecución ineficiente

MI VECINA ME VINO A RECLAMAR QUE MI PERRO PERSIGUE A SU HIJO EN BICICLETA. JAJAJAJA QUE TONTA, MI PERRO NI SABE ANDAR EN BICI.

MI VECINA ME VINO A RECLAMAR QUE MI PERRO PERSIGUE A SU HIJO EN BICICLETA. JAJAJAJA QUE TONTA, MI PERRO NI SABE ANDAR EN BICI.







## Lourdes no quiere a su tía porque es muy envidiosa.

¿Quién es envidiosa, Lourdes o su tía? El pez está listo para comer.

¿El pez está listo para ser comido o está listo para que le den de comer?

Estuve esperándote en el banco.

¿En qué banco? ¿Sentado en un banco o en una institución financiera? Se debe limpiar aquí.

Lourdes no quiere a su tía porque es muy envidiosa.

¿Quién es envidiosa, Lourdes o su tía? El pez está listo para comer.

¿El pez está listo para ser comido o está listo para que le den de comer?

Estuve esperándote en el banco.

¿En qué banco? ¿Sentado en un banco o en una institución financiera? Se debe limpiar aquí.

Lourdes no quiere a su tía porque es muy envidiosa.

¿Quién es envidiosa, Lourdes o su tía? El pez está listo para comer.

¿El pez está listo para ser comido o está listo para que le den de comer?

Estuve esperándote en el banco.

¿En qué banco? ¿Sentado en un banco o en una institución financiera? Se debe limpiar aquí.

Lourdes no quiere a su tía porque es muy envidiosa.

¿Quién es envidiosa, Lourdes o su tía?

El pez está listo para comer.

¿El pez está listo para ser comido o está listo para que le den de comer?

Estuve esperándote en el banco.

¿En qué banco? ¿Sentado en un banco o en una institución financiera?

Se debe limpiar aquí.

Lourdes no quiere a su tía porque es muy envidiosa.

¿Quién es envidiosa, Lourdes o su tía?

El pez está listo para comer.

¿El pez está listo para ser comido o está listo para que le den de comer?

Estuve esperándote en el banco.

¿En qué banco? ¿Sentado en un banco o en una institución financiera?

Se debe limpiar aquí.



- Informalidad Formalidad: Necesidad de contar con una descripción más precisa (formal) de un lenguaje de programación.
- Razonar matemáticamente sobre un programa lo cual requiere una verificación formal del comportamiento de un LdP.
- Estandarización: Independencia de la implementación y de la máquina.

La especificación precisa de un lenguaje de programación es esencial para describir el comportamiento computacional del mismo.

Sin una notación clara del efecto de los constructores del lenguaje es imposible determinar dicho comportamiento.

- Alfabeto: conjunto de cadenas de caracteres elegidas de una conjunto finito y fijo de símbolos.
- Frases o Constructores: Las cadenas que pertenecen al lenguaje

### Descripción del lenguaje

#### Sintáxis

Descripción del formato de los programas del lenguaje Descripción de cómo se forman las frases

#### Semántica

Estudia el comportamiento de los programas Manejo del "signifcado" de las frases

## Pragmática

Estudia aspectos relacionados con las técnicas empleadas para la construcción de programas.

Manejo del uso práctico de las frases

Grado de éxito con el que un programa cumple sus objetivos tanto en su fidelidad con el modelo de computación subyacente como su utilidad para los programadores Ej. El traductor necesita proveer opciones al usuario, para debugear, para interactuar con el sistema operativo y tal vez para interactuar con un entorno de desarrollo de software. Estas opciones conforman el pragmatismo del traductor (en general, estas facilidades forman parte del a definición del lenguaje)

#### Construcción Sintáctica:

## Ejemplo: Asignación de una expresión

```
< asig > \rightarrow < < variable > < signo\_asig > < expresin > ; < variable > \rightarrow < identificador > < signo\_asig > \rightarrow := | \dots | < expresion > \rightarrow \dots
```

#### Construcción Semántica:

## Ejemplo: Asignación de una expresión

$$a := b + 1$$

- ► Variable (lado izquierdo): lugar en la memoria.
- Expresión (lado derecho): computación de un valor basado en operandos (lugares en memoria, literales, constantes, etc.,) y operaciones
- Operación de Asignación: guardar el valor calculado de la expresión en el lugar de la memoria que representa el nombre de la variable

## Construcción Pragmática:

**Ejemplo: Asignación de una expresión** ¿Cuál es el "uso" de la asignación?:

- setear el valor de una expresión que necesito usar luego más de una vez.
- comunicar valores de un lugar a otro del programa
- modificar parte de la estructura de un objeto de datos.
- guardar valores acumulados en una iteración
- Etc.

- Los errores léxicos ocurren durante el análisis léxico:
  - ightharpoonup caracteres ilegales Ej. (en C) x@=2
  - errores ortográficos: Ej whille en lugar de while
- Los errores sintácticos incluyen tokens faltantes y expresiones mal organizadas
- Los errores semánticos pueden ser.
  - Estáticos: (es decir, detectados antes de la ejecución).
     Tipos incompatibles o variables no declaradas;
  - Dinámicos: (detectados durante la ejecución). Un subíndice fuera de rango o la división entre cero.

- Los errores léxicos ocurren durante el análisis léxico:
  - ightharpoonup caracteres ilegales Ej. (en C) x@=2
  - errores ortográficos: Ej whille en lugar de while
- Los errores sintácticos incluyen tokens faltantes y expresiones mal organizadas
- Los errores semánticos pueden ser.
  - Estáticos: (es decir, detectados antes de la ejecución).
     Tipos incompatibles o variables no declaradas;
  - Dinámicos: (detectados durante la ejecución). Un subíndice fuera de rango o la división entre cero.

- Los errores léxicos ocurren durante el análisis léxico:
  - ightharpoonup caracteres ilegales Ej. (en C) x@=2
  - errores ortográficos: Ej whille en lugar de while
- Los errores sintácticos incluyen tokens faltantes y expresiones mal organizadas
- Los errores semánticos pueden ser.
  - Estáticos: (es decir, detectados antes de la ejecución). Tipos incompatibles o variables no declaradas;
  - Dinámicos: (detectados durante la ejecución). Un subíndice fuera de rango o la división entre cero.

- Los errores léxicos ocurren durante el análisis léxico:
  - ightharpoonup caracteres ilegales Ej. (en C) x@=2
  - errores ortográficos: Ej whille en lugar de while
- Los errores sintácticos incluyen tokens faltantes y expresiones mal organizadas
- Los errores semánticos pueden ser.
  - Estáticos: (es decir, detectados antes de la ejecución).
     Tipos incompatibles o variables no declaradas;
  - Dinámicos: (detectados durante la ejecución). Un subíndice fuera de rango o la división entre cero.

- Los errores léxicos ocurren durante el análisis léxico:
  - caracteres ilegales Ej. (en C) x@ = 2
  - errores ortográficos: Ej whille en lugar de while
- Los errores sintácticos incluyen tokens faltantes y expresiones mal organizadas
- Los errores semánticos pueden ser.
  - Estáticos: (es decir, detectados antes de la ejecución).
     Tipos incompatibles o variables no declaradas;
  - Dinámicos: (detectados durante la ejecución). Un subíndice fuera de rango o la división entre cero.

#### Otros tipo de errores:

Los errores lógicos: Estos son aquellos que comete el programador y que hacen que el programa se comporte de una manera errónea o no deseable. Ej ciclos infitinitos si u y v valen 1

```
x = u;
y = v;
while ( y , != 0)
t = y;
y = x * y;
x = t;
```

- No hay violación en la fase semántica
- NO son errores de la traducción

# Definición formal de lenguajes

#### Reconocedores:

- Un dispositivo de reconocimiento lee cadenas de entrada de un lenguaje y decida si la entrada pertenece o no al lenguaje
- ► Ejemplo: analizar sintáctico de un compilador.

#### Generadores

- Un dispositivo que genera las oraciones (sentencias) de un lenguaje
- Uno puede determinar si la sintaxis de una sentencia particular es correcta, comparándolas con la generada por un generador
- ► Ejemplo: gramática

# Definición formal de lenguajes

#### Reconocedores:

- Un dispositivo de reconocimiento lee cadenas de entrada de un lenguaje y decida si la entrada pertenece o no al lenguaje
- ► Ejemplo: analizar sintáctico de un compilador.

#### Generadores:

- Un dispositivo que genera las oraciones (sentencias) de un lenguaje
- Uno puede determinar si la sintaxis de una sentencia particular es correcta, comparándolas con la generada por un generador
- ► Ejemplo: gramática

# Descripción de la sintaxis

### Terminología:

- Una sentencia (oración) es una cadena de caracteres pertenecientes a un alfabeto.
- Un lenguaje es un conjunto de sentencias
- Un lexema es la unidad sintáctica básica de un lenguaje (Ej, \*, suma, begin, etc.)
- Un token (componente léxico) es una categoría de lexemas (Ej, identificador, literal, etc.)

# Métodos formales para describir la sintaxis

Tres mecanismos describen el diseño e implementación de los lenguajes de programación:

- 1. Expresiones regulares
- Gramáticas formales
- 3. Gramáticas de atributos
- Las expresiones regulares nos permitirán definir los lexemas o tokens(trabajo del analizador léxico)
- Las gramáticas formales definirán la sintaxis y
- Las gramáticas de atributos, definirán la semántica estática.

# Métodos formales para describir la sintaxis

- Gramáticas
- Forma de Backus-Naur (Backus-Naur Form) y Gramáticas Libres de Contexto
- Forma de Backus-Naur Extendida (Extended BNF)

# Métodos formales para describir la sintaxis

EXAMPLE 3.1	A Grammar for a Small Language
	<pre><pre><pre><pre><pre><pre><pre><pre></pre></pre></pre></pre></pre></pre></pre></pre>

## Forma de Backus-Naur (BNF)

- Backus-Naur Form (1959)
- Desarrollada por John Backus para describir al Algol 58
- Equivalente a las gramáticas libres de contexto
- Meta-lenguaje utilizado para describir otro lenguaje

```
EXAMPLE 3.5
                                   BNF and EBNF Versions of an Expression Grammar
                                   BNF:
                                         \langle expr \rangle \rightarrow \langle expr \rangle + \langle term \rangle
                                                          <expr> - <term>
                                                          <term>
                                         <term> → <term> * <factor>
                                                        | <term> / <factor>
                                                         <factor>
                                         <factor> \rightarrow <exp> ** <factor>
                                                         | <exp>
                                         \langle \exp \rangle \rightarrow (\langle \exp r \rangle)
                                                      1 id
                                   EBNF:
                                         \langle \expr \rangle \rightarrow \langle term \rangle \{(+ | -) \langle term \rangle \}
                                         <term> \rightarrow <factor> \{(* \mid \land) <factor>\}
                                         < factor > \rightarrow < exp > \{ ** < exp > \}
                                         \langle \exp \rangle \rightarrow (\langle \exp r \rangle)
                                                      | id
```

## Derivación

- Cada cadena de símbolos en una derivación está en forma sentencial
- Una sentencia (frase u oración) es una forma sentencial que sólo tiene símbolos terminales
- Una derivación izquierda es aquella en la cual se expande el no terminal más a la izquierda de cada forma sentencial
- Una derivación puede ser a izquierda o a derecha

## Derivación

- Cada cadena de símbolos en una derivación está en forma sentencial
- Una sentencia (frase u oración) es una forma sentencial que sólo tiene símbolos terminales
- Una derivación izquierda es aquella en la cual se expande el no terminal más a la izquierda de cada forma sentencial
- Una derivación puede ser a izquierda o a derecha

## Derivación

- Cada cadena de símbolos en una derivación está en forma sentencial
- Una sentencia (frase u oración) es una forma sentencial que sólo tiene símbolos terminales
- Una derivación izquierda es aquella en la cual se expande el no terminal más a la izquierda de cada forma sentencial
- Una derivación puede ser a izquierda o a derecha

### Derivación

- Cada cadena de símbolos en una derivación está en forma sentencial
- Una sentencia (frase u oración) es una forma sentencial que sólo tiene símbolos terminales
- Una derivación izquierda es aquella en la cual se expande el no terminal más a la izquierda de cada forma sentencial
- Una derivación puede ser a izquierda o a derecha

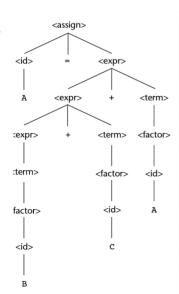
## Ejemplo de derivación

```
=> begin <stmt> ; <stmt list> end
         => begin <var> = <expression> ; <stmt_list> end
         => begin A = <expression> ; <stmt_list> end
         => begin A = <var> + <var> ; <stmt list> end
         => begin A = B + <var> ; <stmt list> end
         => begin A = B + C ; <stmt list> end
         => begin A=B+C; <stmt> end
         => begin A = B + C ; <var> = <expression> end
         => begin A = B + C ; B = <expression> end
         => begin A = B + C ; B = <var> end
         => begin A = B + C : B = C end
```

## Árbol sintáctico

Figure 3.4

A parse tree for A = B + C + A illustrating the associativity of addition

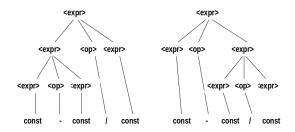


## Árbol sintáctico

$$<$$
 expr  $>$   $\rightarrow$   $<$  term  $><$  op  $><$  term  $>$   $|$  const  $<$  op  $>$   $\rightarrow$   $/|-$ 

## Árbol sintáctico

$$<$$
 expr  $>$   $\rightarrow$   $<$  term  $><$  op  $><$  term  $>$   $|$  const  $<$  op  $>$   $\rightarrow$   $/|-$ 



## BNF y E-BNF

BNF

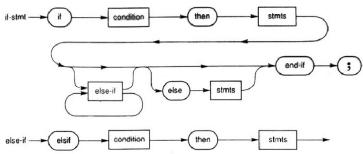
$$< expr> 
ightarrow < expr> + < term> \ | < expr> - < term> \ | < term> \ < term> \ < term> \int < term> \ | < term> \ | < factor> \ | < factor> \ | < factor>$$

E-BNF

$$<$$
 expr  $>$   $\rightarrow$   $<$  term  $>$   $(+|-)$   $<$  term  $>$   $<$  term  $>$   $\rightarrow$   $<$  factor  $>$   $(*|/)$   $<$  factor  $>$ 

## Diagramas de Vías

Diagramas sintácticos: muestran las cadenas válidas en forma gráficas



## Diagramas de Vías

Diagramas sintácticos: muestran las cadenas válidas en forma gráficas

Operación	BNF	Diagrama de Comway
Yuxtaposición	AB	→ <b>A</b> → <b>B</b> →
Opción	A B	<b>A</b>
	ε Β	B
Repetición	1 o más veces {B}	<b>B</b>
	0 o más veces [B]	

## Ejemplo en BNF

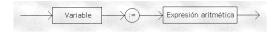
```
<comando asignación> ::= <variable> := <expresión aritmética>
<expresión aritmética> ::= <fermino> | <expresión aritmética> + <fermino> | <expresión aritmética> - <fermino>
<fermino> ::= <primario> | <fermino> * <primario> | <fermino> / <primario>
<primario> ::= <variable> | <número> | <(expresión aritmética>)
<variable> ::= <identificador> | <identificador> | subíndice> ::= <expresión aritmética>
```

## Ejemplo en BNF Extendido

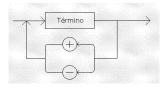
```
<comando asignación> ::= <variable> := <expresión aritmética>
<expresión aritmética> ::= <término> { [+|-] <término> }*
<término> ::= <primario> { [*|/] <primario> }
<primario> ::= <variable> | <número> | (<expresión aritmética>)
<variable> ::= <identificador> | <identificador> [lista subíndice>]
lista subíndice> ::= <expresión aritmética> { , <expresión aritmética> }*
```

# Ejemplo en Diagramas de Vías

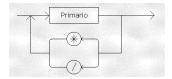
Comando asignación



#### Expresión aritmética.

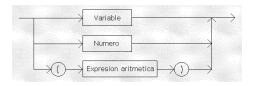


Término.

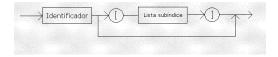


# Ejemplo en Diagramas de Vías

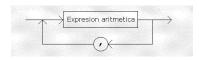
Primario.



Variable.



Lista subíndice.



## Ambigüedad sintáctica

Algunas características para determinar si una gramática es ambigüa:

- si la gramática genera una sentencia con más de una derivación a izquierda.
- si la gramática genera una sentencia con más de una derivación a derecha.

#### Semántica estática:

- Una gramática de atributos es una mecanismo utilizado para describir más de lo que se puede con la gramática libre de contexto pero que pueden ser evaluados en compilación
- Es una extensión de las gramáticas libres de contexto, permite la descripción de ciertas reglas del lenguaje como la compatibilidad de tipos.
- La idea es agregarle a la gramática tradicional, información semántica a lo largo del árbol de parsing.
- Gramática de Atributos se define como
  - una gramática libre de contexto a la cual se le han agregado atributos, funciones para la evaluación de los atributos y funciones predicado.

#### Atributo:

- asociados a los símbolos de la gramática, son similares a las variables (en el sentido de que pueden tener valores asociados)
- Funciones para la evaluación de atributos:
  - Ilamadas funciones semánticas, están asociadas a las reglas de la gramática. Son utilizadas para indicar cómo los atributos especificados son computados.
- Funciones predicado:
  - determinan la regla semántica estática del lenguaje, están asociadas a reglas de la gramática.

#### Definición

Una gramática de atributos es una gramática libre de contexto G=(S,N,T,P) con los siguientes agregados:

- Para cada símbolo de la gramática X, hay un conjunto de atributos.
- ► El conjunto se divide en dos conjuntos disjuntos atributos sintetizados y heredados respectivamente
- Hay atributos intrínsecos en las hojas del árbol de derivación
- Cada regla tiene un conjunto de funciones que definen ciertos atributos de los no-terminales en la regla.
- Cada regla tiene un conjunto (posiblemente vacío) de predicados para chequear la consistencia de los atributos.

#### **EXAMPLE 3.6**

#### An Attribute Grammar for Simple Assignment Statements

- Syntax rule: <assign> → <var> = <expr>
   Semantic rule: <expr>.expected\_type ← <var>.actual\_type

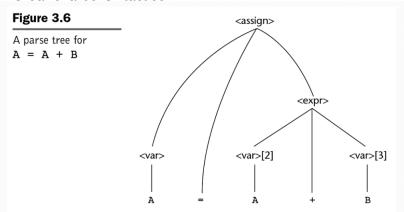
Predicate: <expr>.actual\_type == <expr>.expected\_type

- Syntax rule: <expr> → <var>
   Semantic rule: <expr>.actual\_type ← <var>.actual\_type
   Predicate: <expr>.actual\_type == <expr>.expected\_type
- Syntax rule: <var> → A | B | C
   Semantic rule: <var>.actual\_type ← look-up(<var>.string)

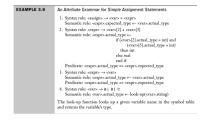
The look-up function looks up a given variable name in the symbol table and returns the variable's type.

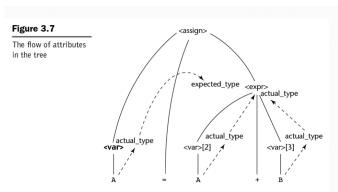
Ejemplo: A = A + B

1. Crear el árbol sintáctico

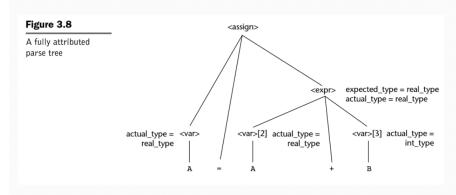


- 2. Agregar los atributos intrínsecos
- 3. Heredar atributos

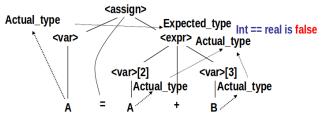




#### 4. Sintetizar atributos



#### 5. Determinar si es correcto



### Desafíos

Dada la siguiente gramática ambigua, dar otra que no lo sea.

```
< ifSent > \Rightarrow if < expLog > then < sent > | if < expLog > then < sent > else < sent > < sent > \Rightarrow < ifSent >
```

Dar una gramática ambigüa y otra que no lo es para la asignacion cuya operaciones pueden ser sumas y divisiones solamente. El orden de precedencia es primero la división y luego la suma. Las expresiones que se pueden sumar o dividir son identificador solamente.

# Bibliografía

- Sebesta, Robert, Concepts of Programming Languages, 9th Edition. 2009. Capítulo 1 y 3.
- Pratt, Terrance W., Programming Languages: Design and Implementation, 3rd Edition. Capítulo 3 (y Cap. 4 de la 4ta edición).
- Material de apoyo en pedco.
- (opcional) Louden, Kenneth, Programming Languages: principles and practices. 3rd Edition 2011. Cengage Learning. Capítulo 1 y 6.