



Instituto Politécnico Nacional
Escuela Superior de Cómputo



Compiladores

Profr. Edgardo Adrián Franco Martínez

13 Lenguajes y gramáticas I

Definición de lenguaje y gramática

efranco.docencia@gmail.com



<http://computacion.cs.cinvestav.mx/~efranco>



Contenido

- Gramática
 - *Definición formal*
 - *Ejemplos*
 - *Definiciones*
- Lenguaje
 - *Definición formal*
- Importancia de las gramáticas
- Ejemplo: Palíndroma (PAL)
 - Definición recursiva de PAL
 - Gramática que define a PAL
- Observaciones del remplazo
- Explicación de los elementos de una gramática
- Lenguaje generado por una gramática (Derivaciones)
- Ejemplos



Gramática

- La **Gramática** es el **estudio de las reglas y principios** que regulan el uso de las lenguas y la organización de las palabras dentro de una oración. También se denomina así al conjunto de reglas y principios que gobiernan el uso de un lenguaje muy determinado; así, cada lenguaje tiene su propia gramática.





Gramática (Definición formal)

- La gramática es un ente formal para especificar, de una manera finita, el conjunto de cadenas de símbolos que constituyen un lenguaje.
- Una gramática es una cuádrupla :

$$G = (VT , VN , S , P)$$

donde :

- $VT = \{ \text{conjunto finito de símbolos terminales} \}$
- $VN = \{ \text{conjunto finito de símbolos no terminales} \}$
- S es el *símbolo inicial* y pertenece a VN .
- $P = \{ \text{conjunto de producciones o de reglas de derivación} \}$



Gramática (Definiciones)

- **Todas las cadenas** del lenguaje definido por la gramática están formadas con símbolos del **vocabulario terminal VT**.
 - El vocabulario terminal se define por enumeración de los símbolos terminales.
- El **vocabulario no terminal VN** es el conjunto de **símbolos introducidos como elementos auxiliares para la definición de la gramática**, y que no figuran en las sentencias del lenguaje.
 - El vocabulario no terminal se define por enumeración de los símbolos no terminales.



Gramática (Definiciones)

- La intersección entre el vocabulario terminal y no terminal es el conjunto vacío:

$$\{VN\} \cap \{VT\} = \{\emptyset\}$$

- La unión entre el vocabulario terminal y no terminal es el *vocabulario* :

$$\{VN\} \cup \{VT\} = \{V\}$$



Gramática (Definiciones)

- En ocasiones es importante distinguir si un determinado vocabulario incluye o no la cadena vacía, indicándose con superíndice + o superíndice *.

$$V^+ = V - \{\lambda\}$$

$$V^* = V + \{\lambda\}$$



Gramática (Definiciones)

- Las producciones **P** son las reglas que se aplican desde el símbolo inicial para obtener las cadenas del lenguaje.
- El conjunto de producciones **P** se define por medio de la enumeración de las distintas producciones, en forma de reglas o por medio de un metalenguaje.
 - *p.g. BNF (Backus Naur Form) o EBNF (Extended Backus Naur Form).*

Ejemplos

Ejemplo 1

Sea la gramática : $G = (VT, VN, S, P)$ donde $VT = \{a, b\}$, $VN = \{S\}$, y el conjunto de producciones es :

$$S \rightarrow ab$$

$$S \rightarrow aSb$$

Ejemplo 2

Sea la gramática $G = (VN, VT, S, P)$ donde :

$$VN = \{ \langle \text{número} \rangle, \langle \text{dígito} \rangle \}$$

$$VT = \{ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \}$$

$$S = \langle \text{número} \rangle$$

Las reglas de producción P son :

$$\langle \text{número} \rangle ::= \langle \text{dígito} \rangle \langle \text{número} \rangle$$

$$\langle \text{número} \rangle ::= \langle \text{dígito} \rangle$$

$$\langle \text{dígito} \rangle ::= 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$$



Lenguaje

- Se puede definir como un conjunto de palabras de un determinado alfabeto.
- Alguien puede pensar que los lenguajes se pueden definir por enumeración de las cadenas que pertenecen a dicho lenguaje, pero este método además de ineficiente, es en muchos casos imposible (*habitualmente un lenguaje tiene un número infinito de cadenas*).
- Así los lenguajes se definen por las propiedades que cumplen las cadenas del lenguaje.



Lenguaje (Definición formal)

- El lenguaje $L(G)$ generado por una gramática G es el conjunto de todas las sentencias que puede generar G . Es decir expresado formalmente :

$$L(G) = \{\eta \in VT^*/S \rightarrow \eta\}$$

- Una sentencia pertenece a $L(G)$ si :
 - Está compuesta de símbolos terminales
 - La sentencia puede derivarse del símbolo inicial S aplicando las reglas de producción de la gramática.



Importancia de las gramáticas

**Son una herramienta
muy poderosa para
describir y analizar
lenguajes.**



Ejemplo: Palíndroma (PAL)

- Un palíndroma (PAL) es una palabra w que cumple $w = w^R$, donde w^R es la misma palabra w sólo que escrita en orden inverso.
- Sea $PAL = \{w \in (a + b)^* \mid w = w^R\}$, es decir, PAL es el lenguaje que consiste de las palabras palíndromas sobre $\{a, b\}$.
- Ejemplos de palíndromas sobre $\{a, b\}$
 - $a, aba, aabbaa, babaabab$.



Definición recursiva de *PAL*

1. $\lambda, a, b \in PAL$
2. Si $S \in PAL$ entonces aSa y $bSb \in PAL$
3. Ninguna cadena pertenece a *PAL* a menos que pueda ser obtenida utilizando las reglas 1 y 2 un número finito de veces. Podemos considerar a S como una variable que representa genéricamente un elemento de *PAL* cuyo valor queremos calcular. Informalmente, las reglas que definen *PAL* pueden reescribirse como:
 - S puede tomar el valor λ, a o b .
 - S puede tomar el valor aSa o bSb , donde la nueva S debe ser calculada. Y si en lugar de escribir “*puede tomar el valor*” escribimos el símbolo \rightarrow , entonces escribiríamos ...



Gramática que define *PAL*

1. $S \rightarrow \lambda, S \rightarrow a, S \rightarrow b.$
Abreviando: $S \rightarrow \lambda \mid a \mid b.$
2. $S \rightarrow aSa, S \rightarrow bSb.$
Abreviando: $S \rightarrow aSa \mid bSb.$
3. Cualquier palíndroma sobre $\{a, b\}$ debe poder ser obtenido aplicando un número finito de veces las reglas de reemplazo 1 y 2.



Ejemplo: palabras que no son palíndromas

- $S \rightarrow A$
- $A \rightarrow aBb \mid bBa$
- $B \rightarrow aB \mid bB \mid \lambda$



Observaciones del remplazo

- Por lo general las reglas se escriben $\alpha \rightarrow \beta$ en lugar de (α, β) .
- Aplicar la regla $\alpha \rightarrow \beta$ a una palabra $a\alpha b$ produce la palabra $a\beta b$, por lo que las reglas pueden ser vistas como reglas de remplazo.

Explicación de los elementos de una gramática

- **Símbolos terminales:** son elementos del alfabeto que no se pueden transformar, por eso se llaman terminales. Normalmente se denotan por letras minúsculas.
- **Variables o símbolos no terminales:** son elementos auxiliares que permiten poner restricciones sintácticas a un lenguaje. Las variables sí se pueden transformar, utilizando las reglas, en una cadena de variables y/o terminales. Por lo general se denotan por letras mayúsculas o por la notación *<variable>*.
- **Reglas:** permiten reemplazar variables para generar oraciones válidas de un lenguaje. Puede haber varias reglas para una misma variable, en algunos casos, las distintas opciones se colocan en una sola regla con los distintos reemplazos separados por “|”.
P.g. $\alpha \rightarrow \beta \mid \gamma \mid \delta$ abrevia las tres reglas

$$\alpha \rightarrow \beta$$

$$\alpha \rightarrow \gamma$$

$$\alpha \rightarrow \delta$$

- **Símbolo inicial:** es el símbolo a partir del cual se generan todas las palabras válidas.

Lenguaje generado por una gramática

- **Definición:** Decimos que la cadena w_1 *deriva en un paso* a la cadena w_2 ($w_1 \Rightarrow_G w_2$) si y solo si existen cadenas $x, y \in \mathbf{V}^*$ tales que $w_1 = xuy$ y $w_2 = xvy$ y además existe una regla $u \rightarrow v$ en \mathbf{P} .
 - Se acostumbra omitir el subíndice que indica la gramática G .
- **Definición:** una cadena $w \in \mathbf{V}^*$ es *derivable* a partir de la gramática G si y solo si existe una secuencia de derivación iniciando en el símbolo inicial y terminando en la cadena w : $S = w_1 \Rightarrow w_2 \Rightarrow w_3 \Rightarrow \dots \Rightarrow w_n = w$.
 - Escribimos $\alpha \Rightarrow_G \beta$ si α deriva a β en 0 o más pasos.
- **Definición:** el *lenguaje generado* por una gramática G , $L(G)$, es igual al conjunto de las palabras en \mathbf{VT} derivables a partir de G .
 - Una gramática describe las reglas sintácticas del lenguaje. Si una palabra no sigue las reglas, entonces no pertenecen al lenguaje generado por la gramática.

Ejemplo

- $G = (VT, VN, S, P)$
 - $VN = \{S, A, B\}$
 - $VT = \{a, b, c\}$
 - $P: S \rightarrow AccA \quad A \rightarrow BA \mid \lambda \quad B \rightarrow a \mid b \mid c$
- $w_1 = abcc \in L(G)$ y $w_2 = acb \notin L(G)$

Cadena	Regla	Derivación
S	$S \rightarrow AccA$	$S \Rightarrow AccA$
$AccA$	$A \rightarrow BA$	$\Rightarrow BAccA$
$BAccA$	$B \rightarrow a$	$\Rightarrow aAccA$
$aAccA$	$A \rightarrow BA$	$\Rightarrow aBAccA$
$aBAccA$	$B \rightarrow b$	$\Rightarrow abAccA$
$abAccA$	$A \rightarrow \lambda$	$\Rightarrow abccA$
$abccA$	$A \rightarrow \lambda$	$\Rightarrow abcc$



Derivaciones

- **Derivación por la izquierda:** las reglas de reemplazo son aplicadas a la primera variable de izquierda a derecha.
- **Derivación por la derecha:** las reglas de reemplazo son aplicadas a la última variable de izquierda a derecha.
- P.g:

$$G = (VN, VT, S, P)$$

$$VN = \{S, A\}$$

$$VT = \{a, b\}$$

$$P: S \rightarrow AA \quad A \rightarrow AAA \mid bA \mid Ab \mid a$$

- En esta gramática existen, por lo menos cuatro derivaciones de la palabra *ababaa*.

Derivaciones de *ababaa*

$S \Rightarrow AA$	$S \Rightarrow AA$	$S \Rightarrow AA$	$S \Rightarrow AA$
$\Rightarrow aA$	$\Rightarrow AAAA$	$\Rightarrow Aa$	$\Rightarrow aA$
$\Rightarrow aAAA$	$\Rightarrow aAAA$	$\Rightarrow AAAa$	$\Rightarrow aAAA$
$\Rightarrow abAAA$	$\Rightarrow abAAA$	$\Rightarrow AAbAa$	$\Rightarrow aAAa$
$\Rightarrow abaAA$	$\Rightarrow abaAA$	$\Rightarrow ABaa$	$\Rightarrow abAAa$
$\Rightarrow ababAA$	$\Rightarrow ababAA$	$\Rightarrow AbAbaa$	$\Rightarrow abAbAa$
$\Rightarrow ababaA$	$\Rightarrow ababaA$	$\Rightarrow Ababaa$	$\Rightarrow ababAa$
$\Rightarrow ababaa$	$\Rightarrow ababaa$	$\Rightarrow ababaa$	$\Rightarrow ababaa$