Gramáticas libres del contexto
(GLL, grado 2)

Tipos de Gramáticas

Tipo	Gramática	Restricciones a la forma de las Reglas	Lenguaje	Autómata
0	Irrestricta	Ninguna restricción $\mu_1\mu_2$ $\beta_1\beta_2$ β_3 con $\mu_i\beta_i \in (V \cup T)^*$	Recursivamente Enumerables	Máquina de Turing
1	Dependientes del Contexto o Sensibles de Contexto	La parte derecha contiene como mínimo los símbolos de la parte izquierda $A\mu B \rightarrow A\beta B$ con $A,B \in V$ y $\mu,\beta \in (V \cup T)^*$	Dependiente del Contexto	Autómata lineales infinitos
2	Independiente del Contexto	La parte izquierda solo puede tener un símbolo $A \rightarrow \beta$, con $A \in V$, y $\beta \in (V \cup T)^*$	Independiente del Contexto	Autómata de Pilas
3	Regulares	La regla solo puede tener 2 formas: $A \rightarrow aB$ y $A \rightarrow a$, con $A,B \in V$ y $a \in T$	Regulares	Autómata finito

Gramática libre de contexto

- En informática, una **gramática libre de contexto** es una gramática formal en la que cada regla de producción es de la forma:
  $\mathrm{V} \to w$

- Donde V es un símbolo no terminal y w es una cadena de terminales y/o no terminales. El término *libre de contexto* se refiere al hecho de que el no terminal V puede siempre ser sustituido por w sin tener en cuenta el contexto en el que ocurra. Un lenguaje formal es libre de contexto si hay una gramática libre de contexto que lo genera.

Gramática libre de contexto

- permiten describir la mayoría de los lenguajes de programación, de hecho, la síntaxis de la mayoría de lenguajes de programación está definida mediante gramáticas libres de contexto.

- son suficientemente simples como para permitir el diseño de eficientes algoritmos de análisis sintáctico que, para una cadena de caracteres dada determinen como puede ser generada desde la gramática.

Gramática libre de contexto

- Los analizadores LL tratan subconjuntos restringidos de gramáticas libres de contexto.

- La notación más frecuentemente utilizada para expresar gramáticas libres de contexto es la forma Backus-Naur.

Definición formal

- Una gramática libre de contexto puede ser definida mediante la 4-tupla:

$$G = (Vt, Vn, P, S) \text{ donde}$$

- V_t es un conjunto finito de terminales
- V_n es un conjunto finito de no terminales
- P es un conjunto finito de producciones
- $S \in V_n$ es el Símbolo Inicial
- los elementos de P son de la forma:

$$V_n \in (V_n \text{ U } V_t)^*$$

Gramáticas Libres del Contexto (Tipo 2)

Sea G = ({A}, {a, b}, S, P)   donde   P = { $S \rightarrow A$, $A \rightarrow aAb$, $A \rightarrow ab$ }

Algunas cadenas generadas por la gramática G:

$S \Rightarrow A \Rightarrow ab$

$S \Rightarrow A \Rightarrow aAb \Rightarrow aabb$

$S \Rightarrow A \Rightarrow aAb \Rightarrow aaAbb \Rightarrow aaabbb$

G es una gramática libre del contexto que genera el lenguaje:

$$L = \{w \in \Sigma^* / \ w = a^n b^n \text{ con } n > 0 \}$$

OTRA FORMA…

- Una gramática libre de contexto simple es:

$$S \rightarrow aSb \mid \varepsilon$$

- Donde | es un *o lógico* es usado para separar múltiples opciones para el mismo no terminal, ε indica una cadena vacía.

- Esta gramática genera el lenguaje no regular

$$L=\{w \in \Sigma^*/w=a^n b^n,\ n>0\}.$$

Ejemplo: Palíndromes

- Un palíndrome es una palabra w que cumple $\boldsymbol{w = w^{\mathrm{R}}}$, donde w^R es la misma palabra w sólo que escrita en orden inverso.

- Sea *PAL* el lenguaje de las palabras palíndromes sobre $\{a, b\}$, entonces:

$$PAL = \{w \in (a+b)^* \mid w = w^{\mathrm{R}}\}$$
$$= \{\varepsilon, a, \mathrm{b}, \mathrm{aa}, \mathrm{bb}, \mathrm{aaa}, \mathrm{bbb}, aba, \mathrm{bab}, \ldots\ldots\}$$

¿Es PAL un lenguaje regular?

Gramática que define *PAL*

1. $S \to \varepsilon \mid a \mid \quad b$
2. $S \to aSa,$
3. $S \to bSb.$

4. Cualquier palíndrome sobre $\{a, b\}$ debe poder ser obtenido aplicando un número finito de veces las reglas de reemplazo 1, 2 y 3.

Ejemplo: palabras no palíndromes

1. $S \rightarrow aSa \mid bSb \mid A$
2. $A \rightarrow aBb \mid bBa$
3. $B \rightarrow aB \mid bB \mid \quad \varepsilon$

Ejemplo: GLL para un palíndrome

1. $S \to \varepsilon, \;\; S \to a, \;\; S \to b.$

   o simplemente: $S \to \varepsilon \mid a \mid b$

2. $S \to aSa, \;\; S \to bSb.$

   o simplemente: $S \to aSa \mid bSb$

3. Cualquier palíndrome sobre $\{a, b\}$ debe poder ser obtenido aplicando un número finito de veces las reglas 1 y 2.

Ejemplo: palabras no palíndromes

1. $S \to aSa \mid bSb \mid A$
2. $A \to aBb \mid bBa$
3. $B \to aB \mid bB \mid \lambda$

4. *Esto define la negación del lenguaje L anteriormente descrito, es L*

5. *Notar que la negación de L no es igual que el complemento de L.*

Ejemplo

1. La GLL para el lenguaje de todas las cadenas que se pueden formar con las letras a y b, habiendo un número diferente de una que de otra, sería:
   1. $S \to U|V$
   2. $U \to TaU|TaT$
   3. $V \to TbV|TbT$
   4. $T \to aTbT|bTaT|\varepsilon$
2. T genera todas las cadenas con la misma cantidad de letras a que b, U genera todas las cadenas con más letras a, y V todas las cadenas con más letras b.

Ejemplo:

1. *La gramática sin restricciones*
    1. *V= {S, A, C}*
    2. *S= {a,b,c}*
    3. *S → aAbc | e*
    4. *A → aAbC | e*
    5. *Cb → bC*
    6. *Cc → cc*

2. *Con símbolo inicial S genera el lenguaje*
3. *L={ai bi ci | i>=0}*

EJEMPLO

$Sea\ L=\{w \in \Sigma^*/w=a^n b^n c^{n+m}, con\ \ n>0\ y\ m>0\}$

L NO es un lenguaje regular, y puede ser generado por la siguiente gramática libre de contexto:

1. $S \rightarrow aSc|B$
2. $B \rightarrow bBc|\varepsilon$

Árbol de Derivación

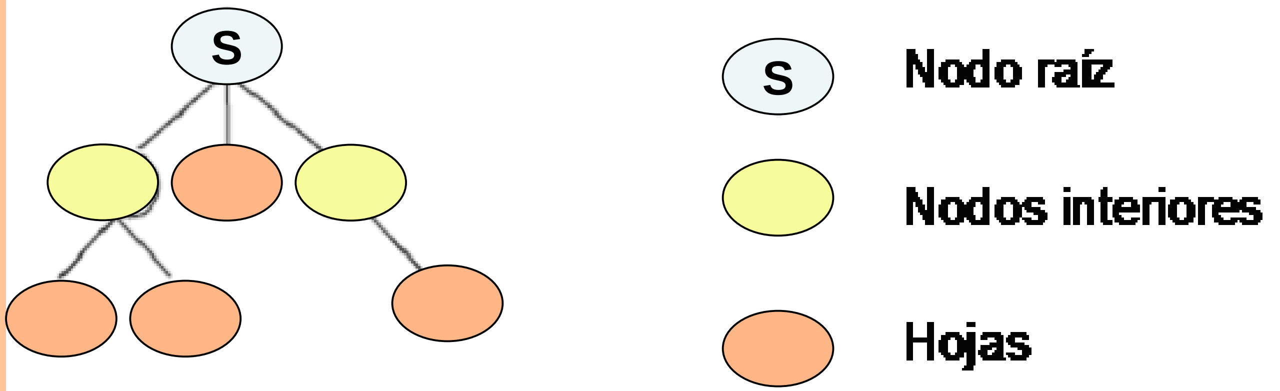
1. Un árbol de derivación (o árbol sintáctico) es una representación gráfica de como se deriva una forma sentencial a partir del símbolo no-terminal inicial.

2. Un árbol es un grafo dirigido acíclico en el cual cada nodo se conecta con un nodo distinguido, llamado nodo raíz mediante un único camino.

Árbol de Derivación

3. Un nodo n1 se dice descendiente de otro nodo n2 si se puede llegar a n1 a partir de n1. El nodo raíz no es descendiente de ningún nodo, y los nodos que no tienen descendientes se denominan hojas. El resto de los nodos se denominan nodos interiores.

Árbol de Derivación



Un árbol de derivación tiene las siguientes propiedades:

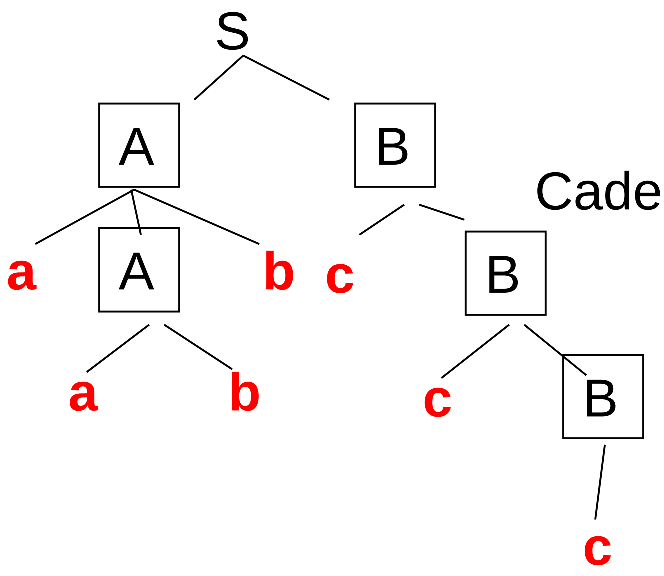
1. El nodo raíz está rotulado con el símbolo distinguido (inicial) de la gramática.
2. Cada hoja corresponde a un símbolo terminal o un símbolo no-terminal.
3. Cada nodo interior corresponde a un símbolo no-terminal.

Árbol de derivación

Sea G = ({A, B}, {a, b, c}, S, P) donde

P = { $S \rightarrow AB$, $A \rightarrow aAb$, $A \rightarrow ab$, $B \rightarrow cB$, $B \rightarrow c$ }

¿ La cadena aabbccc $\in$ L(G)?



Cadena derivada: *aabbccc*

Formas Normales de Chomsky

1. Es un modelo o forma normal para las producciones. Se dice que una GLC está en Forma Normal de Chomsky, si no contiene ε producciones y si todas las producciones son de la forma:

2. A→a, para a Є Σ
3. A→BC, con B y C no terminales

4. Toda GLC puede ser transformada en una GLC en Forma Normal de Chomsky.

5. Sea G una GLC tal que ε∉L(G)
6. Sea G una GLC tal que ε Є L(G)

Formas Normales

Forma Normal de Greibach

Una gramática libre de contexto G = < V, Σ, S, P> esta en la Forma Normal de Greibach si cada regla tiene alguna de las siguientes formas:

a. $A \rightarrow a\, A_1 A_2 \ldots A_n$

b. $A \rightarrow a$

c. $S \rightarrow \varepsilon$

En donde $a \in \Sigma$ y $A_i \in V - \{S\}$ para $i = 1,2, \ldots n$

Teorema. Toda gramática libre de contexto puede ser convertida a una gramática equivalente en la forma normal de Greibach.