Sintaxis y Semántica de los lenguajes

Gramáticas libres de contexto

2021

Facultad Regional Delta,
Universidad Tecnológica Nacional

Prof. Juan Miguel Santos

Gramática libre de contexto.

Es una gramática del tipo 2 en la jerarquía de Chomsky.

Sus producciones son de la forma

$$A \rightarrow \alpha$$

donde $\alpha \in (V_T \cup V_N)^*$ y $A \in V_N$.

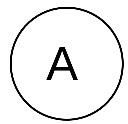
A las cadenas generadas por una gramática libre de contexto (GLC) se les puede asociar un árbol de derivación.

Arbol de derivación.

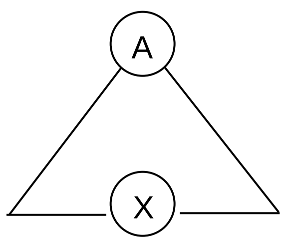
Se nota $\tau(A)$ donde A es un no terminal.

Un árbol de derivación es un grafo tal que

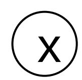
i) Si $A \in V_N$, $\tau(A)$ es un árbol de derivación y su representación es



ii) Si

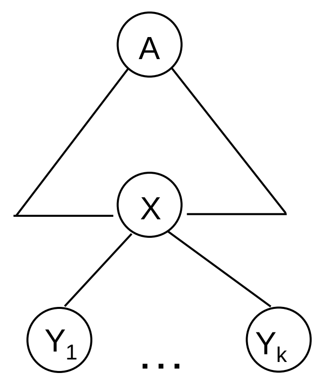


es un $\tau(A)$ y



es una hoja de (A) tal que

$X \in V_N \wedge X \rightarrow Y_1 Y_2 \ldots Y_k$ entonces



también es un árbol de derivación.

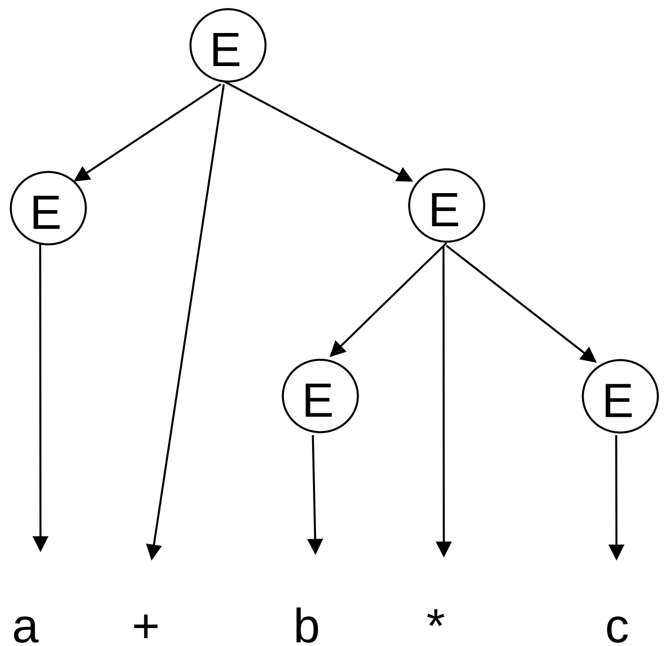
Gramática ambigua.

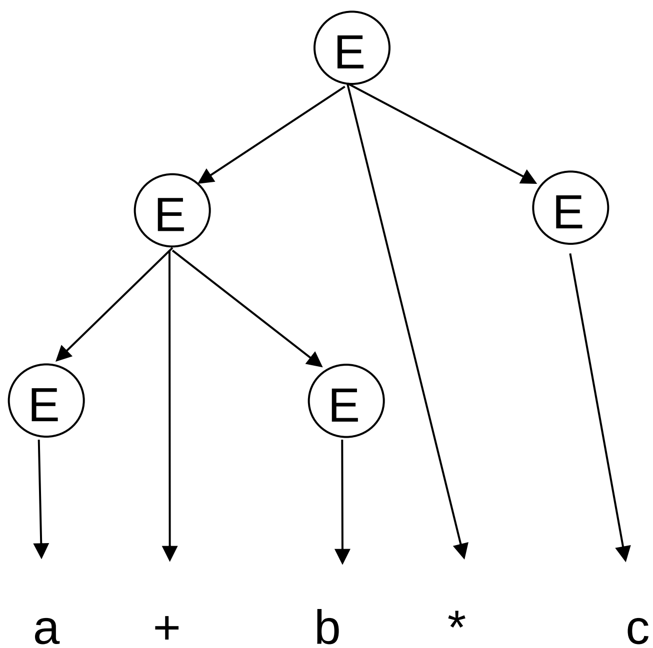
Una GLC G = < V_N, V_T, P, S > para la cual $\exists \alpha \in L(G)$ / $\exists\ \tau(S), \tau'(S) \wedge \tau(S) \neq \tau'(S)$ y α es base de ambos, se dice que la gramática es ambigua.

Ejemplo: Sea la GLC G = < V_N, V_T, P, E > con

P={ E$\rightarrow$E+E / E*E / (E) / id / const } y

sea α = a+b*c existen dos arboles de derivación

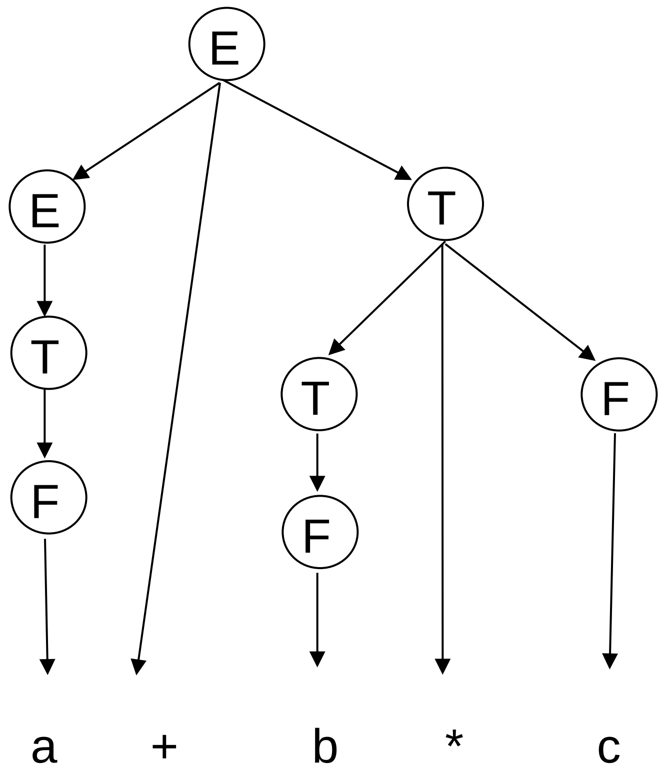




Existe una gramática de las expresiones que no es ambigua:

G = < { E, T, F }, { +, *, (,), id, const }, P, E >

con P = { E $\rightarrow$E+T / T, T $\rightarrow$T*F / F, F $\rightarrow$(E) / id / const }.



El último árbol de derivación se puede recorrer de varias maneras pero hay dos que nos interesan especialmente:

1)$E \rightarrow E+T \rightarrow T+T \rightarrow F+T \rightarrow a+T \rightarrow a+T*F \rightarrow a+F*F \rightarrow a+b*F \rightarrow a+b*c$

2)$E \rightarrow E+T \rightarrow E+T*F \rightarrow E+T*c \rightarrow E+F*c \rightarrow E+b*c \rightarrow T+b*c \rightarrow F+b*c \rightarrow a+b*c$

A la primer forma de derivar la denominamos

derivación mas a la izquierda

mientras que a la segunda la denominamos

derivación mas a la derecha.

Formalmente, $w\alpha\beta$ es una derivación mas a la izquierda si y sólo si

$wA\beta \rightarrow w\alpha\beta \wedge \exists A \rightarrow \alpha \in P \wedge w \in V_T^*$.

Análogamente, $w\alpha\beta$ es una derivación mas a la derecha si y sólo si $wA\beta \rightarrow w\alpha\beta \wedge \exists A \rightarrow \alpha \in P \wedge \beta \in V_T^*$.

Forma sentencial.

Una cadena α es una forma sentencial en G si α fue derivada a partir del símbolo distinguido de la gramática.

Forma sentencial izquierda.

Una cadena α es una forma sentencial izquierda en G si α fue derivada a partir del símbolo distinguido de la gramática usando siempre derivaciones más a la izquierda.

Forma sentencial derecha.

Una cadena α es una forma sentencial derecha en G si α fue derivada a partir del símbolo distinguido de la gramática usando siempre derivaciones más a la derecha.

Un lenguaje es **intrínsecamente ambiguo** si no existe ninguna gramática no ambigua que lo genere.

Lema del bombeo (para lenguajes del tipo 2)

Sea L un lenguaje del tipo 2, entonces existe $p > 0$ / $\forall\alpha \in L$, $|\alpha| \geq p$, $\exists$ r,x,y,z,s / $\alpha = rxyzs$, $|xyz| \leq p$, $(x\neq\lambda \vee z\neq\lambda)$ y $\forall i \geq 0$, $rx^iyz^is \in L$.

Definiciones previas:

Sea x un símbolo de una cadena que es base de un árbol $\tau(A)$, llamamos **camino** de x en $\tau(A)$ a la cantidad de arcos necesarios para llegar desde A a x.

Sea un árbol $\tau(A)$, llamamos **altura** de $\tau(A)$ a

$$\text{máx}\{|\alpha x| \ / \ x \text{ es hoja de } \tau(A)\}$$

donde αx son los caminos en $\tau(A)$ que van de A a x.

Lema

Sea la GLC G = $< V_N, V_T, P, S >$, $P \neq \phi$, si $\tau(S)$ es un árbol de derivación en G para α, a = max$\{|\beta| \,/\, A \rightarrow \beta \in P\}$ y h es la altura de $\tau(S)$, entonces

$\forall \alpha \,/\, S \rightarrow^* \alpha, \; |\, \alpha \,| \leq a^h$

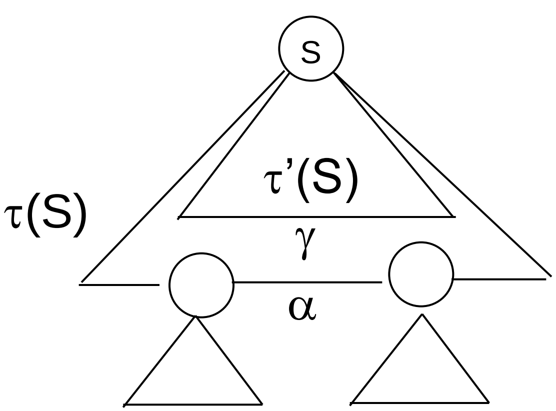
Demostración (por inducción en h):

<u>Caso base</u> (h = 0)

El único árbol es S y su altura es 0. Como $P \neq \phi$, $\exists\, A \rightarrow \beta \in P$ y la hipótesis es verdadera.

<u>Paso inductivo</u> (sea $\tau(S)$ de altura h+1):

Alguna o algunas de sus hojas tendrán caminos de longitud h+1.

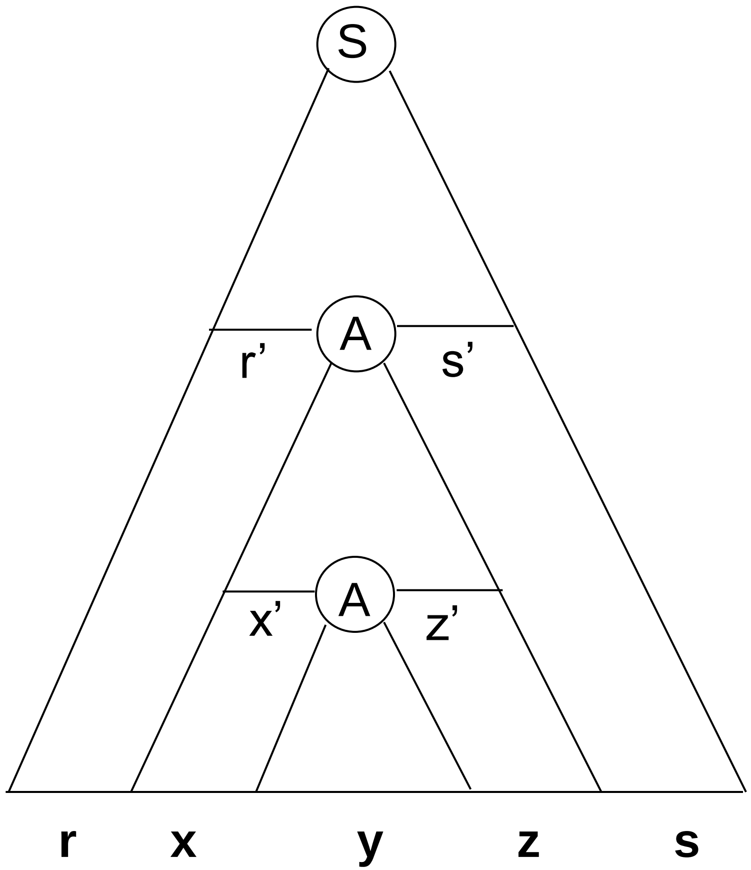


Si α es la base de $\tau(S)$ y γ es la base de $\tau'(S)$,
$|\, \alpha \,| \leq |\gamma|\, a \leq a^h.a = a^{h+1}$.

Lema del bombeo (para lenguajes del tipo 2)
$\forall$L LLC, $\exists$ p > 0 / ($\forall \alpha \in$ L, ($|\alpha| \geq$ p $\Rightarrow \exists$ r,x,y,z,s / α = rxyzs, $|xyz| \leq$ p, ($x \neq \lambda \vee z \neq \lambda$) y $\forall i \geq 0$, ($rx^iyz^is \in$ L))).

Demostración:
Sea a = max $\{|\beta| / A \rightarrow \beta \in P\}$ y tomamos $p=a^{|VN| +1}$.
Sea $\alpha \in$L, $|\alpha| \geq$ p y sea τ(S) un árbol para α donde h es la altura de τ(S), entonces por el lema anterior, $a^h \geq |\alpha| \geq p = a^{|VN| + 1}$ lo que implica
$h \geq |VN| + 1$.
Como h > |VN| entonces habrá algún no terminal que se repite mas de una vez en un camino de máxima longitud.

Consideremos entonces el árbol de derivación con un camino donde un cierto no terminal A se repite dos veces:



$S \to^* r'As'$

$r' \to^* r$

$s' \to^* s$

$S \to^* rAs$ (I)

$A \to^* x'Az'$
$x' \to^* x$
$z' \to^* z$

$A \to^* xAz$ (II)

$A \to^* y$ (III)

Tenemos que ver que $\forall i \geq 0$, $(rx^iyz^is \in L)$.

Hagamos inducción en i.

<u>Caso base</u> (i = 0):

Si i = 0, entonces rx^iyz^is es rys que se puede obtener por (I) y luego por (III).

<u>Paso inductivo</u> (la hipótesis vale para i):

Si la hipótesis es válida para i,

entonces se puede obtener a partir de $S \rightarrow rx^iyz^is$ para la cual antes tuvimos que derivar rx^iAz^is.

Luego, si a rx^iAz^i le aplico (II) y luego (III) obtengo rx^ixyzz^is que es $rx^{i+1}yz^{i+1}s$.

Ejemplo:

Sea $L = \{ a^n b^n c^n / n > 0 \}$, ¿es un LLC?

Supongamos que sí, luego se cumple el lema del bombeo para los LLC.

Así, $\exists$ p> 0 / para toda cadena α, $|\alpha| \geq p$ se cumple el lema. En particular para $\alpha = a^n b^n c^n$ con $n \geq p$.

Tenemos que ver ahora que realmente existe una descomposición rxyzs.

Supondremos $x \neq \lambda$ (es equivalente suponer $z \neq \lambda$).

Como $|xyz| \leq p$, x podrá estar formado por

1) Sólo por caracteres **a**. En este caso, al bombear x, aumenta la cantidad de caracteres **a**. Por otro lado, z ó contiene **a**s o contiene **b**s (no puede contener caracteres **c** pues $p \leq n$ y x ya contiene todos caracteres **a**). Luego, al bombear z no se equilibran las cantidades y la cadena no pertenece a L.

2) Sólo caracteres **b**. Razonamiento similar.

3) Sólo caracteres **c**. Aumenta sólo la cantidad de **c** y la cadena no pertenece a L.

Entonces ocurre un absurdo pues no se cumple el lema siendo que L es LLC y el absurdo proviene de suponer, precisamente, que L es LLC.

Símbolo no terminal alcanzable.

Sea una gramática $G = < V_N, V_T, P, S >$,

A $\in V_N$ es alcanzable si $\exists$ una forma sentencial de G que contenga a A.

Símbolo activo.

Sea una gramática $G = < V_N, V_T, P, S >$,

A $\in$ V_N es activo si y sólo si $\exists\ \sigma \in V_T^* \ /\ A \to^* \sigma$.

Eliminación de símbolos no activos.

1. $A \leftarrow \phi$

2. Para cada producción $B \rightarrow \beta$

    si $\beta \in V_T^*$ agregar B a A.

3. fin $\leftarrow$ falso

4. Mientras no fin

    fin $\leftarrow$ verdadero

    Para cada $B \rightarrow \beta$

        si $\beta \in (A \cup V_T)^*$

            agregar B a A

            fin $\leftarrow$ falso

Nota: el algortimo devuelve en A el conjunto de símbolos activos. Los símbolos del conjunto $(V_N - A)$ puede ser removidos junto a sus producciones.

Símbolo anulable.

Sea $G = < V_N, V_T, P, S >$, $A \in V_N$, A es anulable si $A \to^* \lambda$.

Se nota como Anul(A).

Construcción del conjunto de símbolos anulables de una gramática.

1. $A \leftarrow \phi$
2. Para cada producción $B \to \beta$
   si $\beta = \lambda$ agregar B a A.

3. Repetir

   Para cada producción $B \to \beta$

   si $\beta \in A^*$ agregar B a A.

   hasta que no haya cambios en A

Gramática reducida.

Una gramática es **reducida** si no posee símbolos inalcanzables ni símbolos no activos.

Gramática propia.

Un GLC $G = < V_N, V_T, P, S >$ es **propia** si y sólo si

$$\forall A \rightarrow \beta \in P, (\beta \neq \lambda \vee A = S).$$

Notación: si $\beta\gamma\alpha$ es una forma sentencial en la cual tenemos especial interés por γ se la puede notar $\ldots\gamma\ldots$