

Práctico 5 Teoría de Lenguajes GRAMÁTICAS LIBRES DE CONTEXTO (GLC)

12 de mayo de 2020

Agenda

- 1 Breve repaso de conceptos básicos sobre GLC
- 2 Ejemplo 1 de GLC
- 3 Ejemplo 2 de GLC
- 4 Ejemplo de correctitud de una GLC

Componentes de una GLC

Sea un alfabeto Σ dado.

$G=(V,T,P,S)$ donde:

- V conjunto de **variables** de la gramática
- T conjunto de **terminales** de la gramática, son símbolos de Σ .
- P conjunto de reglas de la gramática, denominadas **producciones**.
- S variable inicial de la gramática, $S \in V$.

Las producciones de P son de la forma:

$$A \longrightarrow \alpha$$

$$A \in V \quad \alpha \in (V \cup T)^*$$

Derivación en una GLC

Derivación en un paso:

Sean $A \in V$ y $\alpha, \beta, \gamma \in (V \cup T)^*$.

Si $A \rightarrow \gamma$ es una producción de P , $\alpha A \beta \Rightarrow \alpha \gamma \beta$ indica que $\alpha \gamma \beta$ se deriva **en un paso** de $\alpha A \beta$ por aplicación de la regla de producción $A \rightarrow \gamma$. La derivación en un paso se denota por el metasímbolo $\Rightarrow$.

En particular, si $A \rightarrow \gamma \in P$, tomando $\alpha = \beta = \epsilon$ tenemos que $A \Rightarrow \gamma$ (toda regla de la gramática es una derivación en un paso).

Derivación en n pasos:

Se denota como $A \xRightarrow{(n)} \beta$ $A \in V, \beta \in (V \cup T)^*$

Derivación en un número no especificado de pasos:

Se denota como $A \xRightarrow{*} \beta$ $A \in V, \beta \in (V \cup T)^*$

Lenguaje generado por una GLC G

Sea G una GLC.

$$L(G) = \{x \in T^* / S \Rightarrow^* x\}$$

Se tiene que L es un **lenguaje libre de contexto** (LLC)

Ejemplo 1 - Planteo del problema

Definir una gramática libre de contexto que genere el siguiente lenguaje:

$$L_1 = \{x \in \{a, b\}^* / x = a^{2k+t}b^{k+t}a^p, \ k, t, p > 0\}$$

Tira más corta de L_1 : corresponde a tomar $k = t = p = 1$,
 $x = a^3b^2a = aaabba$

Idea para la solución

Reescribimos las tiras de L_1 como:

$$x = a^t a^{2k} b^k b^t a^p, \quad k, t, p > 0$$

Esto nos permite identificar 3 subgrupos dentro de una tira válida, que se pueden generar en forma independiente:

- 1 El bloque formado por las a^p finales
- 2 El bloque formado por las a^t y las b^t
- 3 El bloque formado por las $a^{2k} b^k$ centrales

La idea es definir y usar una variable distinta de la gramática para generar cada uno de los 3 bloques descritos.

Una posible gramática para L_1

$$x = a^t a^{2k} b^k b^t a^p, \quad k, t, p > 0$$

$$G_1 = (V, T, P, S)$$

$$V = \{S, A, B\}$$

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

$P :$

- ① $S \longrightarrow Sa$ // genera las a^p finales
- ② $S \longrightarrow Aa$ // pasa a generar las a^t y b^t asegurando $p > 0$
- ③ $A \longrightarrow aAb$ // genera las a^t y las b^t
- ④ $A \longrightarrow aBb$ // pasa a generar las $a^{2k} b^k$ asegurando $t > 0$
- ⑤ $B \longrightarrow aaBb$ // genera los $a^{2k} b^k$
- ⑥ $B \longrightarrow aab$ // asegura que $k > 0$

Derivación de la tira más corta de L_1 en la gramática G_1

Tenemos que la tira mas corta de L_1 es $x = a^3b^2a = aaabba$

Repetimos las producciones de P :

- ① $S \longrightarrow Sa$
- ② $S \longrightarrow Aa$
- ③ $A \longrightarrow aAb$
- ④ $A \longrightarrow aBb$
- ⑤ $B \longrightarrow aaBb$
- ⑥ $B \longrightarrow aab$

En la derivación que sigue, el número entre paréntesis debajo del símbolo de derivación indica qué regla de producción se usa.

$$S \xRightarrow{(2)} Aa \xRightarrow{(4)} aBba \xRightarrow{(6)} aaabba$$

La tira más corta de L_1 se deriva en 3 pasos en G_1 , es decir

$$S \xRightarrow{(3)} aaabba$$

Gramática alternativa para L_1

Se basa en escribir las tiras de L_1 en una nueva forma:

$$x = a^{2k} a^t b^t b^k a^p, \quad k, t, p > 0$$

Esto nos permite identificar los siguientes 3 subgrupos o bloques independientes dentro de una tira válida:

- ❶ El bloque formado por las a^p finales.
- ❷ El bloque formado por las a^{2k} del inicio y las b^k .
- ❸ El bloque formado por las $a^t b^t$.

Las producciones de esta nueva gramática quedan:

$$S \longrightarrow Sa \mid Aa$$

$$A \longrightarrow aaAb \mid aaBb$$

$$B \longrightarrow aBb \mid ab$$

Observamos que desarrollamos dos gramáticas válidas para un mismo lenguaje, es decir que dado un lenguaje la gramática que lo genera no es necesariamente única.

Ejemplo 2 - Planteo del problema

Definir una gramática libre de contexto que genere el siguiente lenguaje:

$$L_2 = \{x \in \{a, b\}^* / x = b^p a^k b^r, r = \lfloor k/3 \rfloor + 1, k \geq 0, p > 0\}$$

Tira más corta de L_2 : $x = bb$

Corresponde a $k = 0, r = p = 1$,

Análisis de las tiras de L_2

Las tiras de L_2 son de la forma:

$$x = b^p a^k b^{\lfloor k/3 \rfloor} b, \quad k \geq 0, \quad p > 0$$

- El bloque o substring inicial b^p es independiente del resto, p no tiene relación con k . Esto sugiere que la gramática lo puede generar como un sublenguaje independiente.
- En el bloque central $a^k b^{\lfloor k/3 \rfloor}$, tenemos 3 casos:
 - 1 Si $k \bmod 3 = 0$ (k múltiplo de 3) $\implies |a| = 3|b|$.
 - 2 Si $k \bmod 3 = 1$ $\implies |a| = 3|b| + 1$.
 - 3 Si $k \bmod 3 = 2$ $\implies |a| = 3|b| + 2$.

La idea es definir y usar una variable distinta de la gramática para generar cada uno de los 2 bloques anteriores.

Una posible gramática para L_2

$$x = b^p a^k b^{\lfloor k/3 \rfloor} b, \quad k \geq 0, \quad p > 0$$

$$G_2 = (V, T, P, S)$$

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

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

P :

- ① $S \longrightarrow bS$ // genera las b^p iniciales
- ② $S \longrightarrow bAb$ // pasa a generar los $a^k b^{\lfloor k/3 \rfloor}$, con $p > 0$ y b final
- ③ $A \longrightarrow aaaAb$ // por cada 3 a's genera un b
- ④ $A \longrightarrow \epsilon$ // condicion de fin cuando $k \bmod 3 = 0$
- ⑤ $A \longrightarrow a$ // condicion de fin cuando $k \bmod 3 = 1$
- ⑥ $A \longrightarrow aa$ // condicion de fin cuando $k \bmod 3 = 2$

Ejemplo de correctitud de una GLC

Sea el siguiente lenguaje:

$$L_a = \{x \in \{a, b\}^* / x = a^n b^k, k \geq 1, n > k\}$$

El problema consiste en **demostrar** que la siguiente gramática G genera a L_a , es decir $L(G) = L_a$:

$$S \longrightarrow aSb \mid aUb$$

$$U \longrightarrow aU \mid a$$

Observar que las tiras de L_a pueden escribirse como $x = a^k a^m b^k$, $k \geq 1$, $m \geq 1$, lo que inspira la gramática G .

Vamos a definir también el lenguaje $L_b = \{w = a^m, m \geq 1\}$.

En el proceso de prueba se demostrará que L_b es el sublenguaje derivado de la variable U de la gramática G .

Idea de la prueba de correctitud

Para demostrar que $L(G) = L_a$ planteamos una prueba por doble inclusión:

- 1 $L_a \subseteq L(G)$. Esto nos asegura que toda tira del lenguaje es generada por la gramática (que no falten tiras). Se prueba por inducción completa en el largo de las tiras del lenguaje L_a .
- 2 $L(G) \subseteq L_a$. Esto nos asegura que la gramática no genera tiras de más (en exceso), que no pertenecen al lenguaje. Se prueba por inducción completa en el número de pasos de las derivaciones de las tiras de G .

Prueba de que $L_a \subseteq L(G)$

$$L_a = \{x = a^n b^k, k \geq 1, n > k\}$$

$$L_b = \{w = a^m, m \geq 1\}$$

$$S \longrightarrow aSb \mid aUb$$

$$U \longrightarrow aU \mid a$$

Se demostrarán por IC las siguientes propiedades:

$$P1) \text{ Si } x \in L_a \implies S \xRightarrow{*} x$$

$$P2) \text{ Si } w \in L_b \implies U \xRightarrow{*} w$$

La prueba se hace por IC **mutua** en el **largo de las tiras** de ambos lenguajes, es decir que se prueban en simultáneo ambas propiedades en una única IC, utilizando eventualmente para demostrar la tesis inductiva de una propiedad la hipótesis inductiva de la otra.

Paso base

$$\begin{aligned} L_a &= \{x = a^n b^k, k \geq 1, n > k\} & S &\longrightarrow aSb \mid aUb \\ L_b &= \{w = a^m, m \geq 1\} & U &\longrightarrow aU \mid a \end{aligned}$$

PB1) La tira más corta de L_a es $x = aab$ ($n = 2, k = 1$).

Derivación para esa x en G : $S \Longrightarrow aUb \Longrightarrow aab$

PB2) La tira más corta de L_b es $w = a$ ($m = 1$).

Derivación para esa w en G : $U \Longrightarrow a$

Paso Inductivo

$$L_a = \{x = a^n b^k, k \geq 1, n > k\}$$

$$L_b = \{w = a^m, m \geq 1\}$$

$$S \longrightarrow aSb \mid aUb$$

$$U \longrightarrow aU \mid a$$

Hipótesis inductivas:

$$HI1) \text{ Si } x \in L_a \text{ y } |x| \leq h \implies S \stackrel{*}{\Rightarrow} x$$

$$HI2) \text{ Si } w \in L_b \text{ y } |w| \leq h \implies U \stackrel{*}{\Rightarrow} w$$

Tesis inductivas:

$$TI1) \text{ Si } x \in L_a \text{ y } |x| = h + 1 \implies S \stackrel{*}{\Rightarrow} x$$

$$TI2) \text{ Si } w \in L_b \text{ y } |w| = h + 1 \implies U \stackrel{*}{\Rightarrow} w$$

Observamos que debe ser $h \geq 1$ pues $\epsilon \notin L_a, \epsilon \notin L_b$

Prueba de la TI1 - Caso 1) $k = 1$

$$\begin{aligned} L_a &= \{x = a^n b^k, k \geq 1, n \geq k\} & S &\longrightarrow aSb \mid aUb \\ L_b &= \{w = a^m, m \geq 1\} & U &\longrightarrow aU \mid a \end{aligned}$$

$$\text{HI2) Si } w \in L_b \text{ y } |w| \leq h \implies U \xRightarrow{*} w$$

$$\text{TI1) Si } x \in L_a \text{ y } |x| = h + 1 \implies S \xRightarrow{*} x$$

Sea $x = a^n b^k \in L_a$ con $|x| = n + k = h + 1$

Caso 1) $k=1$

$$\implies x = a^h b \text{ con } h > 1 \text{ pues } x \in L_a$$

$$x = aa^{h-1}b = ax_1b \text{ con } x_1 = a^{h-1} \text{ y } h-1 \geq 1$$

$$\xRightarrow{\text{definición de } L_b} x_1 \in L_b$$

$$x_1 \in L_b \text{ y } |x_1| = h-1 \leq h \xRightarrow{(\text{HI2})} U \xRightarrow{*} x_1$$

Tenemos entonces la siguiente derivación en G para x :

$$S \implies aUb \xRightarrow{*} ax_1b = a^h b = x \text{ entonces } S \xRightarrow{*} x$$

Prueba de la TI1 - Caso 2) $k > 1$

$$L_a = \{x = a^n b^k, k \geq 1, n \geq k\} \quad S \longrightarrow aSb \mid aUb$$

$$L_b = \{w = a^m, m \geq 1\} \quad U \longrightarrow aU \mid a$$

$$\text{HI1)} \text{ Si } x \in L_a \text{ y } |x| \leq h \implies S \xRightarrow{*} x$$

$$\text{TI1)} \text{ Si } x \in L_a \text{ y } |x| = h + 1 \implies S \xRightarrow{*} x$$

$$\text{Sea } x = a^n b^k \in L_a \text{ con } |x| = n + k = h + 1$$

Caso 2) $k > 1$

$$\implies x = aa^{n-1}b^{k-1}b = ax_1b$$

$$x_1 = a^{n-1}b^{k-1} \text{ con } k-1 \geq 1 \text{ por H) del caso y } n-1 \geq k-1 \text{ pues}$$

$$n \geq k \text{ (} x \in L_a \text{)} \xrightarrow{\text{definición de } L_a} x_1 \in L_a$$

$$x_1 \in L_a \text{ y } |x_1| = h - 1 \leq h \xrightarrow{(\text{HI1})} S \xRightarrow{*} x_1$$

Tenemos entonces la siguiente derivación en G para x :

$$S \implies aSb \xRightarrow{*} ax_1b = aa^{n-1}b^{k-1}b = a^n b^k = x \text{ entonces } S \xRightarrow{*} x$$

Prueba de la TI2

$$L_a = \{x = a^n b^k, k \geq 1, n > k\} \quad S \longrightarrow aSb \mid aUb$$

$$L_b = \{w = a^m, m \geq 1\} \quad U \longrightarrow aU \mid a$$

$$\text{HI2) Si } w \in L_b \text{ y } |w| \leq h \implies U \stackrel{*}{\Rightarrow} w$$

$$\text{TI2) Si } w \in L_b \text{ y } |w| = h + 1 \implies U \stackrel{*}{\Rightarrow} w$$

$$\text{Sea } w = a^m \in L_b \text{ con } |w| = m = h + 1$$

$$\implies w = aa^h = aw_1$$

$$w_1 = a^h \text{ con } h \geq 1 \xrightarrow{\text{definición de } L_b} w_1 \in L_b$$

$$w_1 \in L_b \text{ y } |w_1| = h \xrightarrow{(\text{HI2})} U \stackrel{*}{\Rightarrow} w_1$$

Tenemos entonces la siguiente derivación en G para w :

$$U \implies aU \stackrel{*}{\Rightarrow} aw_1 = aa^h = a^{h+1} = w \text{ entonces } U \stackrel{*}{\Rightarrow} w$$

Cierre de la Prueba de que $L_a \subseteq L(G)$

Al verificarse tanto los pasos base como las tesis inductivas,
Principio de IC $\implies$ quedan probadas las propiedades P1) y P2):

P1) Si $x \in L_a \implies S \overset{*}{\Rightarrow} x$

P2) Si $w \in L_b \implies U \overset{*}{\Rightarrow} w$

En particular la que nos interesa y queríamos demostrar es la propiedad P1), que establece que toda tira del lenguaje L_a es generada por la gramática G .

Prueba de que $L(G) \subseteq L_a$

$$L_a = \{x = a^n b^k, k \geq 1, n > k\}$$

$$L_b = \{w = a^m, m \geq 1\}$$

$$S \longrightarrow aSb \mid aUb$$

$$U \longrightarrow aU \mid a$$

Se demostrarán por IC las siguientes propiedades (recíprocas de las anteriores):

$$P1') \text{ Si } S \xRightarrow{*} x \implies x \in L_a$$

$$P2') \text{ Si } U \xRightarrow{*} w \implies w \in L_b$$

La prueba se hace por IC **mutua** en la **cantidad de derivaciones** necesarias para derivar una tira de terminales tanto en la gramática G con símbolo inicial S como en la de símbolo inicial U .

Paso base

$$L_a = \{x = a^n b^k, k \geq 1, n > k\}$$

$$L_b = \{w = a^m, m \geq 1\}$$

$$S \longrightarrow aSb \mid aUb$$

$$U \longrightarrow aU \mid a$$

PB1) La derivación más corta a partir de S es:

$$S \Longrightarrow aUb \Longrightarrow aab \in L_a \ (n = 2, k = 1).$$

PB2) La derivación más corta a partir de U es:

$$U \Longrightarrow a \in L_b \ (m = 1).$$

Paso Inductivo

$$L_a = \{x = a^n b^k, k \geq 1, n \geq k\}$$
$$L_b = \{w = a^m, m \geq 1\}$$

$$S \longrightarrow aSb \mid aUb$$
$$U \longrightarrow aU \mid a$$

Hipótesis inductivas:

$$HI1) \text{ Si } S \xRightarrow{k} x \text{ y } k \leq h \implies x \in L_a$$

$$HI2) \text{ Si } U \xRightarrow{k} w \text{ y } k \leq h \implies w \in L_b$$

Tesis inductivas:

$$TI1) \text{ Si } S \xRightarrow{h+1} x \implies x \in L_a$$

$$TI2) \text{ Si } U \xRightarrow{h+1} w \implies w \in L_b$$

Prueba de la TI1 - Caso 1) Primer paso $S \Rightarrow aSb$

$$L_a = \{x = a^n b^k, k \geq 1, n > k\}$$

$$L_b = \{w = a^m, m \geq 1\}$$

$$S \longrightarrow aSb \mid aUb$$

$$U \longrightarrow aU \mid a$$

$$\text{HI1) Si } S \xrightarrow{k} x \text{ y } k \leq h \implies x \in L_a$$

$$\text{HI2) Si } U \xrightarrow{k} w \text{ y } k \leq h \implies w \in L_b$$

$$\text{TI1) Si } S \xrightarrow{h+1} x \implies x \in L_a$$

Sea $S \xrightarrow{h+1} x$. Observamos que hay dos casos posibles para el primer paso de la derivación de x : $S \Rightarrow aSb$ o $S \Rightarrow aUb$.

Caso 1) Primer paso en la derivación $S \Rightarrow aSb$

$$\implies x = ax_1b \text{ con } S \xrightarrow{h} x_1 \xrightarrow{\text{(HI1)}} x_1 \in L_a$$

$$\xrightarrow{\text{definición de } L_a} x_1 = a^n b^k, k \geq 1, n > k \implies x = ax_1b = a^{n+1}b^{k+1}$$

$$\xrightarrow{\text{definición de } L_a} x \in L_a \text{ (} k+1 \geq 1, n+1 > k+1 \text{)}$$

Prueba de la TI1 - Caso 2) Primer paso $S \Rightarrow aUb$

$$L_a = \{x = a^n b^k, k \geq 1, n > k\}$$

$$L_b = \{w = a^m, m \geq 1\}$$

$$S \rightarrow aSb \mid aUb$$

$$U \rightarrow aU \mid a$$

$$\text{HI1) Si } S \xRightarrow{k} x \text{ y } k \leq h \Rightarrow x \in L_a$$

$$\text{HI2) Si } U \xRightarrow{k} w \text{ y } k \leq h \Rightarrow w \in L_b$$

$$\text{TI1) Si } S \xRightarrow{h+1} x \Rightarrow x \in L_a$$

$$\text{Sea } S \xRightarrow{h+1} x.$$

Caso 2) Primer paso en la derivación: $S \Rightarrow aUb$

$$\Rightarrow x = aw_1b \text{ con } U \xRightarrow{h} w_1 \xRightarrow{\text{(HI2)}} w_1 \in L_b$$

$$\xRightarrow{\text{definición de } L_b} w_1 = a^m, m \geq 1 \Rightarrow x = aw_1b = aa^m b = a^{m+1}b$$

$$\xRightarrow{\text{definición de } L_a} x \in L_a (k = 1, n = m + 1 > k)$$

Prueba de la TI2

$$L_a = \{x = a^n b^k, k \geq 1, n > k\}$$

$$L_b = \{w = a^m, m \geq 1\}$$

$$S \longrightarrow aSb \mid aUb$$

$$U \longrightarrow aU \mid a$$

$$\text{HI1) Si } S \xrightarrow{k} x \text{ y } k \leq h \implies x \in L_a$$

$$\text{HI2) Si } U \xrightarrow{k} w \text{ y } k \leq h \implies w \in L_b$$

$$\text{TI2) Si } U \xrightarrow{h+1} w \implies w \in L_b$$

$$\text{Sea } U \xrightarrow{h+1} w, h+1 > 1.$$

$\xrightarrow{\text{definición de G}}$ el primer paso en la derivación de w debe ser

$$U \implies aU.$$

$$\implies w = aw_1 \text{ con } U \xrightarrow{h} w_1 \xrightarrow{\text{(HI2)}} w_1 \in L_b.$$

$$\xrightarrow{\text{definición de } L_b} w_1 = a^m, m \geq 1.$$

$$\implies w = aw_1 = a^{m+1}, (m+1) \geq 1 \xrightarrow{\text{definición de } L_b} w \in L_b.$$

Cierre de la Prueba de que $L(G) \subseteq L_a$

Al verificarse tanto los pasos base como las tesis inductivas,
Principio de IC $\longrightarrow$ quedan probadas las propiedades P1') y P2'):

P1') Si $S \xRightarrow{*} x \implies x \in L_a$

P2') Si $U \xRightarrow{*} w \implies w \in L_b$

En particular la que nos interesa y queríamos demostrar es la propiedad P1'), que establece que toda tira de terminales generada por la gramática G es una tira del lenguaje L_a .

Cierre general de la Prueba de correctitud

Al haberse demostrado tanto que $L_a \subseteq L(G)$ como que $L(G) \subseteq L_a$, queda probada la correctitud de la gramática G para el lenguaje L_a , es decir $L(G) = L_a$.