Teoría de Lenguajes
Segunda parte
Teórica 9: Gramáticas LL y LR

Verónica Becher

Primer cuatrimestre 2020

Bibliografía para esta clase:

[1] A. V. Aho, J. D. Ullman, The Theory of Parsing, Translation, and Compiling, Vol. 1 , Parsing. Prentice Hall, 1972. `https://www-2.dc.uba.ar/staff/becher/Aho-Ullman-Parsing-V1.pdf` Capítulos 5.1 y 5.2.

Gramáticas LL y LR

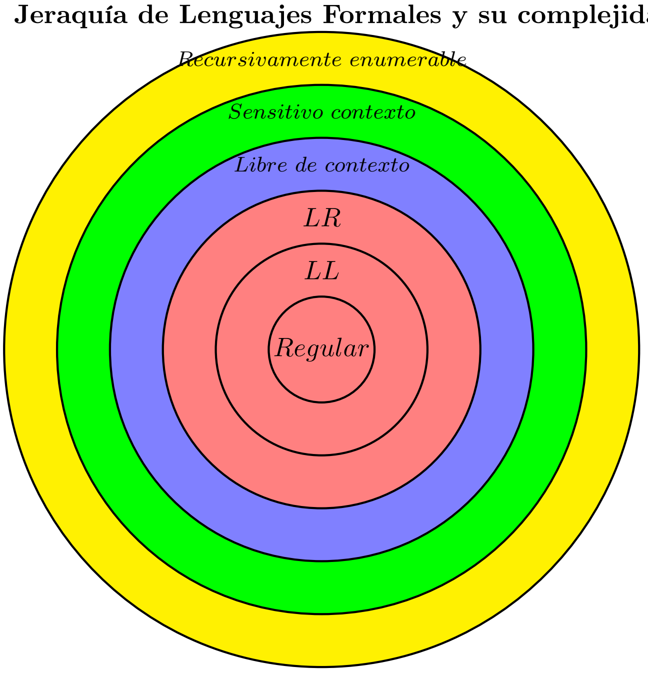
Hay gramáticas libres de contexto que generan lenguajes que se pueden analizar sintácticamente de manera determinística en tiempo lineal en el tamaño de la entrada, leyendo la entrada de izquierda a derecha.

Son las gramáticas LL y LR, los nombres vienen de:

L: left-to right parsing L: leftmost derivation

L: left-to right parsing R: rightmost derivation

Jeraquía de Lenguajes Formales y su complejidad de parsing



Complejdad
Rosa: lineal
Azul: cúbica
Verde: PSPACE
Amarillo: PSPACE

Gramáticas $LL(k)$

Una gramática es LL si es $LL(k)$, para algún número k entero mayor o igual que 1.

Las gramáticas $LL(k)$ son gramaticas libres de contexto no ambiguas para las cuales la derivación más a la izquierda está determinada por los símbolos ya leídos, y k símbolos más.

El parsing asociado es "top down", es decir, va encontrando una tras otra, las producciones que conforman la derivación que empieza con el símblo de inicio S hasta la cadena. Cada paso de la derivación se resuelve esencialmente en tiempo constante, y se demuestra que el tiempo total es lineal en el tamaño de la entrada.

Korenjack y Hopcroft definieron las gramaticas $LL(1)$ en 1966.

Lewis y Stearns (1968), Knuth (1967) y otros, hicieron la teoría y los algoritmos para $LL(k)$.

Gramáticas LL(k)

Sea $G = (T, N, P, S)$ libre de contexto y no ambigua, y sea $w = a_1 a_2 \dots a_n$ una cadena de $L(G)$. Entonces hay una única secuencia $\alpha_0, \alpha_1, \dots, \alpha_m$ de formas sentenciales tal que

$$S = \alpha_0, \quad \alpha_i \underset{L}{\Rightarrow} \alpha_{i+1}, \text{ para } i = 0, 1, \dots, m-1, \text{ y } \alpha_m = w$$

El parsing a izquierda de w es la secuencia de las m producciones usadas en la derivación de w.

Las gramáticas $LL(k)$ son tales que si $\alpha_i = a_1 \dots a_j A\beta$ entonces α_{i+1} es determinable conociendo solamente $a_1 \dots a_j$ y k símbolos más del input $a_{j+1} \dots a_{j+k}$ para el k que hemos fijado.

Gramáticas $LL(k)$

Definición (PRIMEROS$_k(\alpha)$)**.** *Sea $k \geq 1$. Dada una gramática libre de contexto $G = \langle N, T, P, S\rangle$ y una cadena $\alpha \in (T \cup N)^*$,*

$$\text{PRIMEROS}_k(\alpha) = \left\{ z \in T^* : \begin{array}{l} |z| < k, \alpha \overset{*}{\underset{L}{\Rightarrow}} z, \text{ ó} \\ |z| = k, \alpha \overset{*}{\underset{L}{\Rightarrow}} zw \text{ para alguna } w \in T^* \end{array} \right\}$$

En particular para $k = 1$,

$$\text{PRIMEROS}(\alpha) = \left\{ z \in T \cup \{\lambda\} : \begin{array}{l} z = \lambda, \alpha \overset{*}{\underset{L}{\Rightarrow}} \lambda, \text{ ó} \\ z = a \in T, \alpha \overset{*}{\underset{L}{\Rightarrow}} aw \text{ para alguna } w \in T^* \end{array} \right\}$$

Gramática $LL(k)$

Definición (Gramática $LL(k)$)**.** *Una gramática libre de contexto $G = \langle N, T, P, S\rangle$ es $LL(k)$ si, cuando hay dos derivaciones a izquierda*

$$\begin{array}{ll} S & \overset{*}{\underset{L}{\Rightarrow}} wA\alpha \underset{L}{\Rightarrow} w\beta\alpha \overset{*}{\underset{L}{\Rightarrow}} wx \\ S & \overset{*}{\underset{L}{\Rightarrow}} wA\alpha \underset{L}{\Rightarrow} w\gamma\alpha \overset{*}{\underset{L}{\Rightarrow}} wy \end{array}$$

en las que $PRIMEROS_k(x) = PRIMEROS_k(y)$,

se cumple que $\beta = \gamma$.

Ejemplo de gramática $LL(1)$

$$\begin{aligned} S &\rightarrow aAS|b \\ A &\rightarrow a|bSA \end{aligned}$$

Supongamos

$$\begin{aligned} S &\overset{*}{\underset{L}{\Rightarrow}} wS\alpha \underset{L}{\Rightarrow} w\beta\alpha \overset{*}{\underset{L}{\Rightarrow}} wx \\ S &\overset{*}{\underset{L}{\Rightarrow}} wS\alpha \underset{L}{\Rightarrow} w\gamma\alpha \overset{*}{\underset{L}{\Rightarrow}} wy \end{aligned}$$

con PRIMEROS(x) = PRIMEROS(y).

Hay dos posibilidades. Una es PRIMEROS(x) = PRIMEROS(y) $= a$ entonces hay una única producción que lo produjo, $S \rightarrow aAS$. Por lo tanto, $\beta = \gamma = aAS$.

La otra posibilidad es PRIMEROS(x) = PRIMEROS(y) $= b$ entonces hay una única producción que lo produjo, $S \rightarrow b$, Por lo tanto, $\beta = \gamma = b$.

Notar que PRIMEROS(x) = PRIMEROS(y) $= \lambda$ es imposible porque S no deriva λ.

Con un argumento similar se demuestra lo mismo para

$$\begin{aligned} S &\overset{*}{\underset{L}{\Rightarrow}} wA\alpha \underset{L}{\Rightarrow} w\beta\alpha \overset{*}{\underset{L}{\Rightarrow}} wx \\ S &\overset{*}{\underset{L}{\Rightarrow}} wA\alpha \underset{L}{\Rightarrow} w\gamma\alpha \overset{*}{\underset{L}{\Rightarrow}} wy \end{aligned}$$

Ejemplo de gramática $LL(2)$

Consideremos esta gramática:

$$\begin{aligned} S &\rightarrow cAd \\ A &\rightarrow ab|a \end{aligned}$$

No es $LL(1)$ porque

$$\begin{aligned} S &\underset{L}{\Rightarrow} cAd \underset{L}{\Rightarrow} cabd \overset{*}{\underset{L}{\Rightarrow}} cabd \\ S &\underset{L}{\Rightarrow} cAd \underset{L}{\Rightarrow} cad \overset{*}{\underset{L}{\Rightarrow}} cad \end{aligned}$$

Tenemos que PRIMEROS(abd) = PRIMEROS(ad) $= a$ pero $ab \neq a$.

Sin embargo, $\text{PRIMEROS}_2(abd) \neq \text{PRIMEROS}_2(ad)$
Por lo tanto la gramática es $LL(2)$.

Otro ejemplo de gramática $LL(2)$

$$\begin{aligned} S &\rightarrow \lambda|abA \\ A &\rightarrow Saa|b \end{aligned}$$

No es $LL(1)$ porque

$$\begin{aligned} S &\overset{*}{\underset{L}{\Rightarrow}} abSaa \underset{L}{\Rightarrow} ab\beta aa \overset{*}{\underset{L}{\Rightarrow}} ababAaa \\ S &\overset{*}{\underset{L}{\Rightarrow}} abSaa \underset{L}{\Rightarrow} ab\gamma aa \overset{*}{\underset{L}{\Rightarrow}} abaa \end{aligned}$$

PRIMEROS($abAaa$) = PRIMEROS(aa) $= a$ pero $\beta = abA$, $\gamma = \lambda$ y $\beta \neq \lambda$.

Veamos la gramática es $LL(2)$. Tenemos que $\text{PRIMEROS}_2(abAaa) \neq \text{PRIMEROS}_2(\lambda)$.
Y tenemos que

$$\begin{aligned} S &\overset{*}{\underset{L}{\Rightarrow}} abA \underset{L}{\Rightarrow} ab\beta \overset{*}{\underset{L}{\Rightarrow}} abb \\ S &\overset{*}{\underset{L}{\Rightarrow}} abA \underset{L}{\Rightarrow} abSaa \overset{*}{\underset{L}{\Rightarrow}} abaa \\ S &\overset{*}{\underset{L}{\Rightarrow}} abA \underset{L}{\Rightarrow} abSaa \overset{*}{\underset{L}{\Rightarrow}} ababAaa \end{aligned}$$

$\text{PRIMEROS}_2(b) \neq \text{PRIMEROS}_2(aa)$ y $\text{PRIMEROS}_2(b) \neq \text{PRIMEROS}_2(abAaa)$.

Una gramática que no es LL

$$
\begin{aligned}
S &\to A|B \\
A &\to aAb|0 \\
B &\to aBbb|1
\end{aligned}
$$

$L(G) = \{a^n 0 b^n : n \geq 1\} \cup \{a^n 1 b^{2n} : n \geq 1\}$

Notemos que para todo valor de $k \geq 1$,

$$
\begin{aligned}
S &\overset{*}{\underset{L}{\Rightarrow}} S \underset{L}{\Rightarrow} A \overset{*}{\underset{L}{\Rightarrow}} a^k 0 b^k \\
S &\overset{*}{\underset{L}{\Rightarrow}} S \underset{L}{\Rightarrow} B \overset{*}{\underset{L}{\Rightarrow}} a^k 1 b^{2k}
\end{aligned}
$$

$\mathrm{PRIMEROS}_k(a^k 0 b^k) = \mathrm{PRIMEROS}_k(a^k 1 b^{2k}) = a^k$, pero $A \neq B$.

Por lo tanto G no es $LL(k)$, para ningún k.

Teorema sobre gramáticas $LL(k)$

Teorema 1. *Una gramática libre de contexto $G = \langle N, T, P, S\rangle$ es $LL(k)$ si y solo si, para todos los $wA\alpha$ tales que $S \overset{*}{\underset{L}{\Rightarrow}} wA\alpha$ y para todo par de producciones $A \to \beta$ y $A \to \gamma$, con $\beta \neq \gamma$, $PRIMEROS_k(\beta\alpha) \cap PRIMEROS_k(\gamma\alpha) = \emptyset$,*

Demostración

Supongamos $S \overset{*}{\underset{L}{\Rightarrow}} wA\alpha$ y producciones $A \to \beta$, $A \to \gamma$ con $\beta \neq \gamma$ y

$$
\begin{aligned}
S &\overset{*}{\underset{L}{\Rightarrow}} wA\alpha \underset{L}{\Rightarrow} w\beta\alpha \overset{*}{\underset{L}{\Rightarrow}} wxy \\
S &\overset{*}{\underset{L}{\Rightarrow}} wA\alpha \underset{L}{\Rightarrow} w\gamma\alpha \overset{*}{\underset{L}{\Rightarrow}} wxz,
\end{aligned}
$$

con $x \in \mathrm{PRIMEROS}_k(\beta\alpha) \cap \mathrm{PRIMEROS}_k(\gamma\alpha)$ tal que si $|x| < k$ entonces $y = z = \lambda$. Como $\beta \neq \gamma$, G no es $LL(k)$.

Supongamos ahora que G no es $LL(k)$ entonces existen dos derivaciones

$$
\begin{aligned}
S &\overset{*}{\underset{L}{\Rightarrow}} wA\alpha \underset{L}{\Rightarrow} w\beta\alpha \overset{*}{\Rightarrow} wx \\
S &\overset{*}{\underset{L}{\Rightarrow}} wA\alpha \underset{L}{\Rightarrow} w\gamma\alpha \overset{*}{\Rightarrow} wy
\end{aligned}
$$

y existe $z \in T^*$, $|z| \leq k$ tal que $z = \mathrm{PRIMEROS}_k(x) = \mathrm{PRIMEROS}_k(y)$. Entonces, $z \in \mathrm{PRIMEROS}_k(\beta\alpha)$ y $z \in \mathrm{PRIMEROS}_k(\gamma\alpha)$.

Por lo tanto, $\mathrm{PRIMEROS}_k(\beta\alpha) \cap \mathrm{PRIMEROS}_k(\gamma\alpha) \neq \emptyset$. □

Gramáticas $LL(1)$

Definición (SIGUIENTES(A))**.** *Dada una gramática $G = \langle N, T, P, S\rangle$ y un no-terminal A*

$$
SIGUIENTES(A) = \left\{ z \in T \cup \{\lambda\} : S \overset{*}{\Rightarrow} \alpha A \gamma,\ z \in PRIMEROS(\gamma) \right\}.
$$

Definición (Símbolos directrices)**.** *Definimos la función $SD : P \to \mathcal{P}(T)$ de la siguiente manera*

$$
SD(A \to \beta) = \begin{cases} PRIMEROS(\beta) & \text{si } \beta \text{ no es anulable} \\ PRIMEROS(\beta) \cup SIGUIENTES(A) & \text{si } \beta \text{ es anulable} \end{cases}
$$

Observación 2. *Una gramática $G = (N, T, P, S)$ libre de contexto es $LL(1)$ si y solo si para cada no terminal A y terminal a hay a lo sumo una única producción $A \to \beta$ tal que $a \in SD(A \to \beta)$.*

Gramáticas $LL(1)$

Observación 3. *Una gramática* $G = \langle N, T, P, S\rangle$ *es* $LL(1)$ *si, para todo par de producciones* $A \to \beta$ *y* $A \to \gamma$*, con* $\beta \neq \gamma$*,* $PRIMEROS(\beta SIGUIENTES(A)) \cap PRIMEROS(\gamma SIGUIENTES(A)) = \emptyset$.

Las gramáticas LL **no son recursivas a izquierda**

Teorema 4. *Toda gramática* $LL(1)$ *en la cual todos los no-terminales son alcanzables y activos no es recursiva a izquierda.*

Demostracion

Supongamos que la gramática G es recursiva a izquierda. Entonces, existe no terminal A y expresión α tal que

$$A \overset{+}{\Rightarrow} A\alpha$$

Hay dos casos.

Caso A genera una cadena no nula. Existe $t \in T^+$ tal que $A \underset{L}{\overset{+}{\Rightarrow}} t$,

$$A \underset{L}{\Rightarrow} B_1\beta_1 \underset{L}{\Rightarrow} \dots \underset{L}{\Rightarrow} B_i\beta_i \begin{cases} \underset{L}{\Rightarrow} B_{i+1}\delta\beta_i \underset{L}{\Rightarrow} \dots \underset{L}{\Rightarrow} A\alpha \underset{L}{\Rightarrow} \dots \underset{L}{\Rightarrow} t\alpha \\ \underset{L}{\Rightarrow} \gamma\beta_i \underset{L}{\Rightarrow} \dots \underset{L}{\Rightarrow} t \end{cases}$$

Las dos derivaciones desde A coinciden hasta $B_i\beta_i$. En la rama superior B_i se expande mediante $B_i \to B_{i+1}\delta$ (caso recursivo). En la rama inferior B_i se expande mediante $B_i \to \gamma$ (caso no-recursivo).

Entonces, si a es el primer símbolo de t,
$a \in SD\,(B_i \to B_{i+1}\delta)$ y $a \in SD\,(B_i \to \gamma)$.

Por lo tanto, G no es $LL(1)$.

Demostración, continuación

Caso A genera la cadena nula, $A \underset{L}{\overset{+}{\Rightarrow}} \lambda$.

$$A \underset{L}{\Rightarrow} B_1\beta_1 \underset{L}{\Rightarrow} \dots \underset{L}{\Rightarrow} B_i\beta_i \begin{cases} \underset{L}{\Rightarrow} B_{i+1}\delta\beta_i \underset{L}{\Rightarrow} \dots \underset{L}{\Rightarrow} A\alpha \underset{L}{\Rightarrow} \dots \underset{L}{\Rightarrow} \lambda \\ \underset{L}{\Rightarrow} \gamma\beta_i \underset{L}{\Rightarrow} \dots \underset{L}{\Rightarrow} \lambda \end{cases}$$

En la rama superior B_i se expande mediante $B_i \to B_{i+1}\delta$ (caso recursivo). En la rama inferior B_i se expande mediante $B_i \to \gamma$ (caso no-recursivo).

Aquí $B_{i+1}\delta\beta_{i+1}$ y $\gamma\beta_i$ son anulables. Por lo tanto, $B_{i+1}\delta$ y γ también son anulables. Entonces,

$$\text{SIGUIENTES}(B_i) \subseteq SD(B_i \to B_{i+1}\delta) \text{ y}$$
$$\text{SIGUIENTES}(B_i) \subseteq SD(B_i \to \gamma),$$

por lo que $SD(B_i \to B_{i+1}\delta) \cap SD(B_i \to \gamma) \neq \emptyset$.

Concluímos que la gramática G no es $LL(1)$. $\square$

Las gramáticas LL **no son ambiguas**

Teorema 5. *Toda gramática* LL *es no ambigua.*

Demostración. Supongamos una gramática $LL(k)$ que es ambigua. Entonces hay una cadena w y dos derivaciones

$$S \underset{L}{\Rightarrow} \alpha_1 \underset{L}{\Rightarrow} \alpha_2 \ldots \underset{L}{\Rightarrow} \alpha_n \underset{L}{\Rightarrow} w$$
$$S \underset{L}{\Rightarrow} \beta_1 \underset{L}{\Rightarrow} \beta_2 \ldots \underset{L}{\Rightarrow} \beta_m \underset{L}{\Rightarrow} w$$

Consideremos el menor i tal que $\alpha_{n-i} \neq \beta_{m-i}$. Sea $y = \alpha_{n-(i-1)}$. Tenemos,

$$S \overset{*}{\underset{L}{\Rightarrow}} \alpha_{n-i} \underset{L}{\Rightarrow} y$$
$$S \overset{*}{\underset{L}{\Rightarrow}} \beta_{m-i} \underset{L}{\Rightarrow} y$$

Para todo sufijo z de w $\text{PRIMEROS}_k(z) = \text{PRIMEROS}_k(z)$ pero $\alpha_{n-i} \neq \beta_{m-i}$. Esto contradice que la gramática es $LL(k)$. $\square$

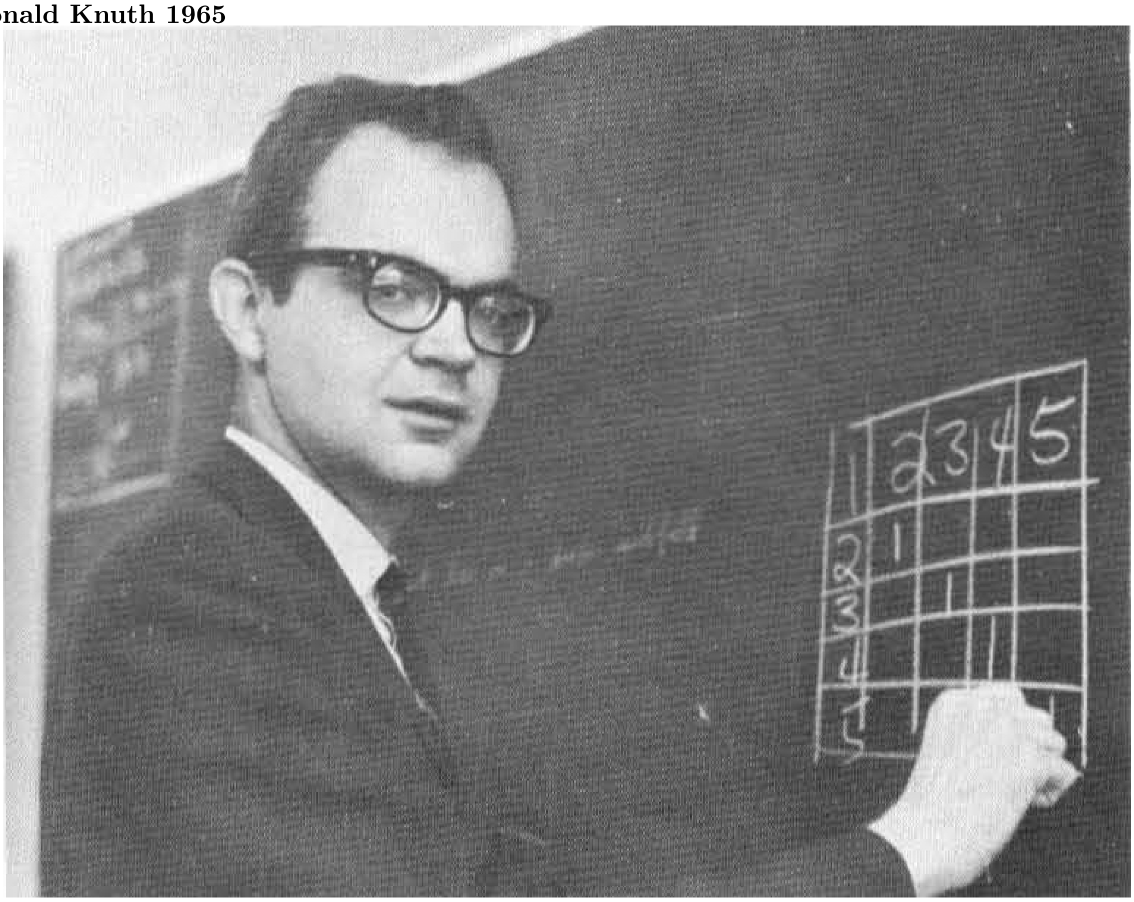
Gramáticas $LR(k)$

Las gramáticas $LR(k)$, con k un número entero mayor o igual que 0, son gramáticas libres de contexto no ambiguas para las cuales dada una expresión del lenguaje se puede encontrar su derivación más a la derecha de manera "bottom-up", de modo tal que en cada paso de la derivación está determinada por los símbolos ya leídos de la cadena de entrada y k símbolos más. De esta forma, cada paso de la derivación se resuelve esencialmente en tiempo constante.

La definición de gramática $LR(k)$se debe a Knuth (1965). Luego la técnica fue mejorada por muchos otros entre ellos De Remer (1969), Karenjack (196), Aho Ulmman (1971).

El concepto $LR(k)$ se extendio a gramáticas sensitivas al contexto Walters (1970).

Donald Knuth 1965



Gramáticas $LR(k)$

Sea $G = (T, N, P, S)$ libre de contexto y no ambigua. Consideremos la gramática extendida $G' = (T, N \cup \{S'\}, P \cup \{S' \to S\}, S')$.

Sea $w = a_1 a_2 \ldots a_m$ una cadena de $L(G)$. Dado que G es no ambigua hay una única secuencia de formas sentenciales $\alpha_0, \alpha_1, \ldots \alpha_m$ tal que

$$S' = \alpha_0, \qquad \alpha_i \underset{R}{\Rightarrow} \alpha_{i+1}, \text{ para } i = 0, 1, \ldots, m-1, \text{ y } \qquad \alpha_m = w,$$

donde S' es un nuevo símbolo de inicio y las derivaciones en G'.

El parsing a derecha de w es la secuencia de las m producciones usadas en la derivación de w.

Supongamos este parsing a derecha $\alpha_m, \alpha_{m-1}, \ldots \alpha_0$ con $\alpha_{i+1} = a_1 \ldots a_j A\beta$. Las gramáticas $LR(k)$ aseguran que cada α_i es determinable teniendo en cuenta solamente $a_1 \ldots a_j$ y k símbolos más del input $a_{j+1} \ldots a_{j+k}$ para el k que hemos fijado.

Gramáticas $LR(k)$

Dada una gramática libre de contexto $G = (N, T, P, S)$ definimos la gramática aumentada $G' = (N \cup \{S'\}, T, P \cup \{S' \to S\}, S')$

Definición. *Sea $G = (N, T, P, S)$ libre de contexto y G' su gramática aumentada. Decimos que G es $LR(k)$, con $k \geq 0$, si estas tres condiciones*

$$\begin{aligned} &S' \overset{*}{\underset{R}{\Rightarrow}} \alpha A w \underset{R}{\Rightarrow} \alpha\beta w \\ &S' \overset{*}{\underset{R}{\Rightarrow}} \gamma B x \underset{R}{\Rightarrow} \alpha\beta y \\ &PRIMEROS_k(w) = PRIMEROS_k(y) \end{aligned}$$

implican $\alpha A y = \gamma B x$.

Ejemplo de gramática $LR(0)$

$$\begin{aligned} S &\to D|E \\ D &\to 0D|d \\ E &\to 0E|e \end{aligned}$$

Cada derivación a derecha en G' es de la forma

$$\begin{aligned} S' \underset{R}{\Rightarrow} S \underset{R}{\Rightarrow} D \overset{i+1}{\underset{R}{\Rightarrow}} 0^i d, \text{ para } i \geq 0 \\ S' \underset{R}{\Rightarrow} S \underset{R}{\Rightarrow} E \overset{i+1}{\underset{R}{\Rightarrow}} 0^i e, \text{ para } i \geq 0 \end{aligned}$$

Entonces $L(G) = 0^* d | 0^* e$.

Para ver que G' es $LR(0)$ es suficiente ver que en las derivaciones

$$\begin{aligned} S' \overset{*}{\underset{R}{\Rightarrow}} \alpha A w \underset{R}{\Rightarrow} \alpha\beta w \\ S' \overset{*}{\underset{R}{\Rightarrow}} \gamma B x \underset{R}{\Rightarrow} \alpha\beta y \end{aligned}$$

$\alpha = \gamma$ y $A = B$, ya que necesariamente $w = x = \lambda$.

Caso $A = S$ entonces $\beta = D$ o $\beta = E$. Luego $B = S$ y $\alpha = \gamma = \lambda$.

Caso $A = D$. Si $\beta = 0D$ entonces $B = D$. Y si $\beta = d$ también $B = D$.

El caso $A = E$ es similar. Concluimos que G es $LR(0)$.

Notar que G no es $LL(k)$ para ningun k.

Dado que $L(G) = 0^* d | 0^* e$ es regular, por lo que admite una gramática LL.

Ejemplo de gramática $LR(1)$

$$\begin{array}{lcl} S' & \to & S \\ S & \to & Sa \mid a \end{array}$$

Para decidir si la gramática G cumple la definición de $LR(1)$ debemos considerar todo par de derivaciones de uno o más pasos. Hay sólo 2 casos.

Primer Caso. Consideremos este par de derivaciones, para $j \geq 0$, $\ell \geq 1$,

$$\begin{array}{l} S' \overset{*}{\underset{R}{\Rightarrow}} Sa^j \quad \underset{R}{\Rightarrow} aa^j \\ S' \overset{*}{\underset{R}{\Rightarrow}} Sa^{j+\ell} \underset{R}{\Rightarrow} aa^{j+\ell} \end{array}$$

La única asignación posible para considerar la definición $LR(1)$ es:

$$\alpha = \gamma = \lambda, \quad A = B = S, \quad w = a^j, \quad x = a^{j+\ell}, \quad \beta = a, \quad y = a^{j+\ell}.$$

Notar que en este caso w es un prefijo propio de y. Si $j = 0$, $w = a^j = \lambda$, $y = a^{j+\ell}$, $\text{PRIMEROS}_1(w) \neq \text{PRIMEROS}_1(y)$, por lo tanto la implicación se cumple automáticamente. Si $j \geq 1$, $w = a^j$, $y = a^{j+\ell}$, $\text{PRIMEROS}_1(w) = \text{PRIMEROS}_1(y)$ y se cumple la implicación porque

$$\alpha Ay = Sa^{j+\ell} = \gamma Bx.$$

Concluimos que este par de derivaciones es consistente con la definición de $LR(1)$.

Segundo Caso. Consideremos este par de derivaciones,

$$\begin{array}{l} S' \overset{*}{\underset{R}{\Rightarrow}} S' \underset{R}{\Rightarrow} S \\ S' \overset{*}{\underset{R}{\Rightarrow}} S \underset{R}{\Rightarrow} Sa \end{array}$$

La única asignación posible para considerar la definición $LR(1)$ es:

$$\alpha = \gamma = \lambda, \quad A = S', \quad B = S, \quad w = \lambda, \quad x = \lambda, \quad \beta = S, \quad y = a.$$

Dado que $w = \lambda$, $y = a$, tenemos $\text{PRIMEROS}_1(w) \neq \text{PRIMEROS}_1(y)$. Entonces la implicación en la definición de $LR(1)$ se cumple automáticamente.

Concluímos que la gramática G es $LR(1)$. Esta es la afirmación más fuerte que podemos hacer, ya que G no es $LR(0)$, y toda gramática $LR(k)$ es también $LR(k+1)$.

Ejemplo de gramática que no es $LR(k)$ para ningun k

$$\begin{array}{lcl} S & \to & Ab \mid Bc \\ A & \to & Aa \mid \lambda \\ B & \to & Ba \mid \lambda \end{array}$$

Observemos que para todo $k \geq 0$ tenemos

$$\begin{array}{l} S' \overset{*}{\underset{R}{\Rightarrow}} S \overset{*}{\underset{R}{\Rightarrow}} Aa^k b \underset{R}{\Rightarrow} a^k b \\ S' \overset{*}{\underset{R}{\Rightarrow}} S \overset{*}{\underset{R}{\Rightarrow}} Ba^k c \underset{R}{\Rightarrow} a^k c \end{array}$$

$\text{PRIMEROS}_k(a^k b) = \text{PRIMEROS}_k(a^k c) = a^k$. Sin embargo $A \neq B$.

Notar que G no es $LL(k)$ para ningun k porque es recursiva a izquierda. Sin embargo $L(G) = a^*b|a^*c$ es regular por lo tanto admite una gramática LL y también admite una gramática LR.

Las gramáticas LR son no ambiguas

Teorema 6. *Toda gramática LR es no ambigua.*

Demostración. Supongamos una gramática $LR(k)$ que es ambigua. Entonces hay una cadena w y dos derivaciones

$$S \underset{R}{\Rightarrow} \alpha_1 \underset{R}{\Rightarrow} \alpha_2 \ldots \underset{R}{\Rightarrow} \alpha_n \underset{R}{\Rightarrow} w$$
$$S \underset{R}{\Rightarrow} \beta_1 \underset{R}{\Rightarrow} \beta_2 \ldots \underset{R}{\Rightarrow} \beta_m \underset{R}{\Rightarrow} w$$

Consideremos el menor i tal que $\alpha_{n-i} \neq \beta_{m-i}$. Sea $y = \alpha_{n-(i-1)}$.

$$S \overset{*}{\underset{R}{\Rightarrow}} \alpha_{n-i} \underset{R}{\Rightarrow} y$$
$$S \overset{*}{\underset{R}{\Rightarrow}} \beta_{m-i} \underset{R}{\Rightarrow} y$$

Para todo sufijo z de y $\text{PRIMEROS}_k(z) = \text{PRIMEROS}_k(z)$ pero $\alpha_{n-i} \neq \beta_{m-i}$. Esto contradice que la gramática es $LR(k)$. $\square$

Teorema 7 (Theorem 8.1 Aho Ulman vol 2)**.** *Los lenguajes LL están incluidos en los lenguajes LR.*

Teorema 8 (Theorem 8.16 Aho Ullman vol 2)**.** *Los lenguajes libres de contexto reconocibles mediante autómatas de pila determinsísticos coinciden con los lenguajes* $LR(1)$.

Notar que una de las implicaciones resulta de que el algoritmo de parsing $LR(1)$ se implementa en un autómata de pila determinístico.

Teorema 9 (Theorem 8.1 Aho Ulman vol 2)**.** *Para toda gramática* $LR(k)$*,* $k \geq 0, G$*, hay una gramática* $LR(1)$ G' *tal que* $L(G') = L(G)$*.*

Algoritmo para PRIMEROS(X)
  if $X \in T$ **then**
    PRIMEROS$(X) = \{X\}$
  else
    PRIMEROS$(X) \leftarrow \emptyset$
    for todas las producciones $X \to Y_1Y_2 \ldots Y_k \in P$ **do**
      PRIMEROS$(X) \leftarrow$ PRIMEROS$(X) \cup$ PRIMEROS(Y_1)
      $i = 2$
      while $i \leq k \ \wedge \ Y_1Y_2 \ldots Y_{i-1} \overset{*}{\underset{G}{\Rightarrow}} \lambda$ **do**
        PRIMEROS$(X) =$ PRIMEROS$(X) \cup$ PRIMEROS(Y_i)
        $i = i + 1$
      end while
      if $X \overset{*}{\underset{G}{\Rightarrow}} \lambda$ **then**
        PRIMEROS$(X) =$ PRIMEROS$(X) \cup \{\lambda\}$
      end if
    end for
  end if

Algoritmo para PRIMEROS$(X_1X_2 \ldots X_n)$**, con** $n > 1$
  PRIMEROS$(X_1X_2 \ldots X_n) \leftarrow$ PRIMEROS$(X_1) - \{\lambda\}$
  i=2
  while $i \leq n \ \wedge \ X_1X_2 \ldots X_{i-1} \overset{*}{\underset{G}{\Rightarrow}} \lambda$ **do**
    PRIMEROS$(X_1X_2 \ldots X_n) \leftarrow$ PRIMEROS$(X_1X_2 \ldots X_n) \cup ($PRIMEROS$(X_i) - \{\lambda\})$

$i = i + 1$
end while
if $X_1 X_2 \ldots X_n \overset{*}{\underset{G}{\Rightarrow}} \lambda$ **then**
PRIMEROS($X_1 X_2 \ldots X_n$) = PRIMEROS($X_1 X_2 \ldots X_n$) $\cup \{\lambda\}$
end if

Algoritmo para SIGUIENTES(B)**, para todo** $B \in N$
SIGUIENTES(S) $\leftarrow \{\$\}$
for todos los $B \in N$ diferentes de S **do**
SIGUIENTES(B) $\leftarrow \emptyset$
end for
for todas las $A \to \alpha B \beta \in P$ **do**
SIGUIENTES(B) $\leftarrow$ SIGUIENTES(B) $\cup$ (PRIMEROS(β) $- \{\lambda\}$)
end for
repeat
for todos los $B \in N$ tales que $A \to \alpha B \beta \in P \ \wedge \ \beta \overset{*}{\underset{G}{\Rightarrow}} \lambda$ **do**
SIGUIENTES(B) $\leftarrow$ SIGUIENTES(B) $\cup$ SIGUIENTES(A)
end for
until ningún SIGUIENTES(B) cambie más