4.10. Forma Normal de Chomsky (FNC)

Una GIC G está en **Forma Normal de Chomsky** (FNC) si satisface:

1. G no tiene variables inútiles.
2. G no tiene producciones λ (excepto posiblemente $S \to \lambda$).
3. Todas las producciones son de la forma: $A \to a$ (producciones simples) ó $A \to BC$ (producciones binarias).

En particular, una gramática en FNC no tiene producciones unitarias.

4.10.1 Teorema (Procedimiento de conversión a FNC). *Toda GIC G es equivalente a una gramática en Forma Normal de Chomsky.*

Demostración: Podemos transformar G en una gramática en FNC, equivalente a G, ejecutando los algoritmos de las secciones anteriores en el siguiente orden:

1. Eliminar las variables no terminales.
2. Eliminar las variables no alcanzables.
3. Eliminar las producciones λ (excepto, posiblemente, $S \to \lambda$).
4. Eliminar las producciones unitarias.
5. Las producciones resultantes (diferentes de $S \to \lambda$) son de la forma: $A \to a$ ó $A \to w$, donde $|w| \geq 2$. Estas últimas se pueden simular con producciones de la forma $A \to BC$ o $A \to a$. Se introduce primero, para cada $a \in \Sigma$, una variable nueva T_a cuya única producción es $T_a \to a$. A continuación, se introducen nuevas variables, con producciones binarias, para simular las producciones deseadas. □

La parte 5 del procedimiento anterior se ilustra en los dos siguientes ejemplos.

Ejemplo Simular la producción $A \to abBaC$ con producciones simples y binarias.

Solución: Introducimos las variables T_a y T_b, y las producciones $T_a \to a$ y $T_b \to b$. Entonces $A \to abBaC$ se simula con:

$$\begin{cases} A \to T_a T_b B T_a C \\ T_a \to a \\ T_b \to b \end{cases}$$

Ahora introducimos nuevas variables T_1, T_2, T_3 y las producciones binarias necesarias. Las únicas producciones de estas nuevas variables son las mostradas:

$$\begin{cases} A \to T_aT_1 \\ T_1 \to T_bT_2 \\ T_2 \to BT_3 \\ T_3 \to T_aC \\ T_a \to a \\ T_b \to b \end{cases}$$

Ejemplo Simular la producción $A \to BAaCbb$ con producciones simples y binarias.

Solución: Introducimos las variables T_a y T_b, y las producciones $T_a \to a$ y $T_a \to b$. Entonces $A \to BAaCbb$ se simula con:

$$\begin{cases} A \to BAT_aCT_bT_b \\ T_a \to a \\ T_b \to b \end{cases}$$

Ahora introducimos nuevas variables T_1, T_2, T_3, T_4 y las producciones binarias necesarias. Las únicas producciones de estas nuevas variables son las mostradas:

$$\begin{cases} A \to BT_1 \\ T_1 \to AT_2 \\ T_2 \to T_aT_3 \\ T_3 \to CT_4 \\ T_4 \to T_bT_b \\ T_a \to a \\ T_b \to b \end{cases}$$

En los siguientes ejemplos se ilustra el procedimiento completo para convertir una gramática dada a la Forma Normal de Chomsky (FNC).

Ejemplo Encontrar una GIC en FNC equivalente a la siguiente a la gramática:

$$G: \quad \begin{cases} S \to AB \mid aBC \mid SBS \\ A \to aA \mid C \\ B \to bbB \mid b \\ C \to cC \mid \lambda \end{cases}$$

<u>Solución</u>: El conjunto de variables terminables es

$$\textbf{TERM} = \{B, C, S, A\},$$

y el conjunto de variables alcanzables es

$$\textbf{ALC} = \{S, A, B, C\}.$$

Es decir, la gramática no tiene variables inútiles. El conjunto de variables anulables es

$$\textbf{ANUL} = \{C, A\}.$$

Al eliminar las producciones λ de G (la única es $C \to \lambda$) se obtiene la gramática equivalente G_1:

$$G_1 : \quad \begin{cases} S \to AB \mid aBC \mid SBS \mid B \mid aB \\ A \to aA \mid C \mid a \\ B \to bbB \mid b \\ C \to cC \mid c \end{cases}$$

A continuación encontramos los conjuntos unitarios de todas las variables:

$$\begin{aligned} &\textbf{UNIT}(S) = \{S, B\}. \\ &\textbf{UNIT}(A) = \{A, C\}. \\ &\textbf{UNIT}(B) = \{B\}. \\ &\textbf{UNIT}(C) = \{C\}. \end{aligned}$$

Al eliminar las producciones unitarias obtenemos la gramática equivalente G_2:

$$G_2 : \quad \begin{cases} S \to AB \mid aBC \mid SBS \mid aB \mid bbB \mid b \\ A \to aA \mid a \mid cC \mid c \\ B \to bbB \mid b \\ C \to cC \mid c \end{cases}$$

Luego introducimos las variables nuevas T_a, T_b y T_c, y las producciones $T_a \to a$, $T_b \to b$ y $T_c \to c$ con el propósito de que todas las producciones sean unitarias o de la forma $A \to w$, donde $|w| \geq 2$.

$$G_3 : \quad \begin{cases} S \to AB \mid T_aBC \mid SBS \mid T_aB \mid T_bT_bB \mid b \\ A \to T_aA \mid a \mid T_cC \mid c \\ B \to T_bT_bB \mid b \\ C \to T_cC \mid c \\ T_a \to a \\ T_b \to b \\ T_c \to c \end{cases}$$

Finalmente, se introducen nuevas variables, con producciones binarias, para simular las producciones de la forma $A \to w$, donde $|w| \geq 2$:

$$G_4: \quad \begin{cases} S \to AB \mid T_aT_1 \mid ST_2 \mid T_aB \mid T_bT_3 \mid b \\ A \to T_aA \mid T_CC \mid a \mid c \\ B \to T_bT_3 \mid b \\ C \to T_cC \mid c \\ T_1 \to BC \\ T_2 \to BS \\ T_3 \to T_bB \\ T_a \to a \\ T_b \to b \\ T_c \to c \end{cases}$$

Ejemplo Encontrar una GIC en FNC equivalente a la siguiente a la gramática:

$$G: \quad \begin{cases} S \to aS \mid aA \mid D \\ A \to aAa \mid aAD \mid \lambda \\ B \to aB \mid BC \\ C \to aBb \mid CC \mid \lambda \\ D \to aB \mid bA \mid aa \mid A \end{cases}$$

Solución: $\mathbf{TERM} = \{A, C, D, S\}$. Eliminando la variable no-terminable B obtenemos:

$$G_1: \quad \begin{cases} S \to aS \mid aA \mid D \\ A \to aAa \mid aAD \mid \lambda \\ C \to CC \mid \lambda \\ D \to bA \mid aa \mid A \end{cases}$$

El conjunto de las variables alcanzables de G_1 es $\mathbf{ALC} = \{S, A, D\}$. Eliminando la variable no-alcanzable C obtenemos:

$$G_2: \quad \begin{cases} S \to aS \mid aA \mid D \\ A \to aAa \mid aAD \mid \lambda \\ D \to bA \mid aa \mid A \end{cases}$$

El conjunto de variables anulables de G_2 es **ANUL** $= \{A, D, S\}$. Eliminando las producciones λ obtenemos:

$$G_3: \quad \begin{cases} S \to aS \mid aA \mid D \mid a \mid \lambda \\ A \to aAa \mid aAD \mid aa \mid aA \mid aD \mid a \\ D \to bA \mid aa \mid A \mid b \end{cases}$$

A continuación encontramos los conjuntos unitarios de todas las variables:

$$\begin{array}{l} \textbf{UNIT}(S) = \{S, D, A\}. \\ \textbf{UNIT}(A) = \{A\}. \\ \textbf{UNIT}(D) = \{D, A\}. \end{array}$$

Al eliminar las producciones unitarias obtenemos la gramática equivalente G_4:

$$G_4: \quad \begin{cases} S \to aS \mid aA \mid a \mid \lambda \mid aAa \mid aAD \mid aa \mid aD \mid bA \mid b \\ A \to aAa \mid aAD \mid aa \mid aA \mid aD \mid a \\ D \to bA \mid aa \mid b \mid aAa \mid aAD \mid aa \mid aA \mid aD \mid a \end{cases}$$

Finalmente, simulamos las producciones de G_4 con producciones unitarias y binarias:

$$G_5: \quad \begin{cases} S \to T_aS \mid T_aA \mid T_aT_1 \mid T_aT_2 \mid T_aT_a \mid T_aD \mid T_bA \mid a \mid b \mid \lambda \\ A \to T_aT_1 \mid T_aT_2 \mid T_aT_a \mid T_aA \mid T_aD \mid a \\ D \to T_bA \mid T_aT_a \mid b \mid T_aT_1 \mid T_aT_2 \mid T_aT_a \mid T_aA \mid T_aD \mid a \\ T_1 \to AT_a \\ T_2 \to AD \\ T_a \to a \\ T_b \to b \end{cases}$$

En algunas aplicaciones de la FNC es necesario exigir que la variable inicial S no aparezca en el cuerpo de ninguna producción. Si S aparece en el lado derecho de alguna producción se dice que S es **recursiva** ya que esto da lugar a derivaciones de la forma $S \stackrel{+}{\Longrightarrow} uSv$, con $u, v \in (V \cup \Sigma)^*$. El siguiente teorema es un resultado muy sencillo; establece que cualquier GIC se puede transformar en una GIC equivalente en la cual la variable inicial no es recursiva.

4.10.2 Teorema. *Dada una GIC $G = (V, \Sigma, S, P)$ se puede construir una GIC $G' = (V', \Sigma, S', P')$ equivalente a G de tal manera que el símbolo inicial S' de G' no aparezca en lado derecho de las producciones de G'.*

Demostración: La nueva gramática G' tiene una variable más que G, la variable S', que actúa como la nueva variable inicial. Es decir, $V' = V \cup \{S'\}$. El conjunto de producciones P' está dado por $P' = P \cup \{S' \to S\}$. Es claro que $L(G) = L(G')$ y el símbolo inicial S' no aparece en el cuerpo de las producciones. □

Según este resultado, el papel de la variable inicial de la nueva gramática G' es únicamente iniciar las derivaciones.

Ejemplo Encontrar una GIC G' equivalente a la siguiente gramática G de tal manera que la variable inicial de G' no sea recursiva.

$$G: \quad \begin{cases} S \to ASB \mid BB \\ A \to aA \mid a \\ B \to bBS \mid \lambda \end{cases}$$

Solución: Según se indicó en la demostración del Teorema 4.10.2, la gramática pedida G' es

$$G': \quad \begin{cases} S' \to S \\ S \to ASB \mid BB \\ A \to aA \mid a \\ B \to bBS \mid \lambda \end{cases}$$

Nótese que S sigue siendo recursiva pero ya no es la variable inicial de la gramática.

Ejemplo Encontrar una GIC en FNC equivalente a la gramática G del ejemplo anterior, de tal manera que su variable inicial no sea recursiva.

Solución: Comenzamos transformando G en G', como se hizo en el ejemplo anterior. En G' todas las variable son útiles y $\textbf{ANUL} = \{B, S, S'\}$. Eliminando las producciones λ obtenemos:

$$G_1: \quad \begin{cases} S' \to S \mid \lambda \\ S \to ASB \mid BB \mid AB \mid AS \mid A \\ A \to aA \mid a \\ B \to bBS \mid bS \mid bB \mid b \end{cases}$$

Los conjuntos unitarios son:
$\textbf{UNIT}(S') = \{S', S\}$, $\textbf{UNIT}(S) = \{S, A\}$, $\textbf{UNIT}(A) = \{A\}$, $\textbf{UNIT}(B) = \{B\}$.

Eliminando las producciones unitarias se obtiene la gramática:

$$G_2 : \quad \begin{cases} S' \to ASB \mid BB \mid AB \mid AS \mid aA \mid a \mid \lambda \\ S \to ASB \mid BB \mid AB \mid AS \mid aA \mid a \\ A \to aA \mid a \\ B \to bBS \mid bS \mid bB \mid b \end{cases}$$

Simulando las producciones de G_2 con producciones unitarias y binarias se obtiene:

$$G_3 : \quad \begin{cases} S' \to AT_1 \mid BB \mid AB \mid AS \mid T_aA \mid a \mid \lambda \\ S \to AT_1 \mid BB \mid AB \mid AS \mid T_aA \mid a \\ A \to T_aA \mid a \\ B \to T_bT_2 \mid T_bS \mid T_bB \mid b \\ T_a \to a \\ T_b \to b \\ T_1 \to SB \\ T_2 \to BS \end{cases}$$

Ejercicios de la sección 4.10

1. Encontrar una gramática en FNC equivalente a la siguiente GIC:

$$G : \quad \begin{cases} S \to ABC \mid BaC \mid aB \\ A \to Aa \mid a \\ B \to BAB \mid bab \\ C \to cC \mid c \end{cases}$$

2. Encontrar una gramática en FNC equivalente a la siguiente GIC:

$$G : \quad \begin{cases} S \to aASb \mid BAb \\ A \to Aa \mid a \mid \lambda \\ B \to BAB \mid bAb \\ C \to cCS \mid \lambda \end{cases}$$

3. Para la gramática del ejercicio 2 anterior encontrar una GIC equivalente en FNC, de tal manera que su variable inicial no sea recursiva.