Laboratorio 7 – Teoría de la Computación

Problema 2: Simplificación de CFG y FNC

Indicaciones

Para cada gramática se aplican, en orden:

1. Eliminación de producciones ε.
2. Eliminación de producciones unitarias.
3. Eliminación de símbolos inútiles (no productores / no alcanzables).
4. Conversión a Forma Normal de Chomsky (FNC / CNF).

Cuando el lenguaje admite la cadena vacía, se introduce un nuevo símbolo inicial S_0 con $S_0 \to S \mid \varepsilon$.

Gramática 1

Original

$$\begin{aligned} &S \to 0A0 \mid 1B1 \mid BB \\ &A \to C \\ &B \to S \mid A \\ &C \to S \mid \varepsilon \end{aligned}$$

1) Eliminación de ε

Anulables: $C \Rightarrow \varepsilon \Rightarrow A$ es anulable ($A \Rightarrow C$), B es anulable ($B \Rightarrow A$ o $B \Rightarrow S$ con S anulable), y S es anulable ($S \Rightarrow BB$ con B anulable). Por tanto, $\{S, A, B, C\}$ son anulables.

Generando variantes (sin introducir el vacío como producción explícita):

$$\begin{aligned} &S \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB \mid B \\ &A \to C \\ &B \to S \mid A \\ &C \to S \end{aligned}$$

2) Eliminación de unitarias

Las unitarias: $A \to C$, $B \to S$, $B \to A$, $C \to S$, $S \to B$ (esta última por $S \to B$ ya presente). Por clausura unitaria, S, A, B, C heredan las *no unitarias* de S:

$$\{\, 0A0,\ 00,\ 1B1,\ 11,\ BB\,\}.$$

Eliminando unitarias, una forma consistente es:

$$\begin{aligned} &S \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB \\ &A \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB \\ &B \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB \\ &C \to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB \end{aligned}$$

3) Símbolos inútiles

Productores: hay producciones a terminales $(00, 11)$, por lo que todos pueden derivar terminales.
Alcanzables desde S: se alcanzan A y B por cuerpos; C no aparece ya en los cuerpos tras quitar unitarias $\Rightarrow$ se elimina C.

$$\begin{aligned} S &\to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB \\ A &\to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB \\ B &\to 0A0 \mid 00 \mid 1B1 \mid 11 \mid BB \end{aligned}$$

4) Conversión a FNC (CNF)

Reemplazamos terminales en cuerpos de longitud ≥ 2:

$$X_0 \to 0, \qquad X_1 \to 1.$$

Factorizamos pares terminal–no terminal:

$$Y_{A0} \to AX_0, \qquad Y_{B1} \to BX_1.$$

Una CNF válida (recordando que el lenguaje original admite ε; añadimos S_0):

$$\begin{aligned} S_0 &\to S \mid \varepsilon \\ S &\to X_0\,Y_{A0} \mid X_1\,Y_{B1} \mid B\,B \mid X_0\,X_0 \mid X_1\,X_1 \\ A &\to X_0\,Y_{A0} \mid X_1\,Y_{B1} \mid B\,B \mid X_0\,X_0 \mid X_1\,X_1 \\ B &\to X_0\,Y_{A0} \mid X_1\,Y_{B1} \mid B\,B \mid X_0\,X_0 \mid X_1\,X_1 \\ X_0 &\to 0, \quad X_1 \to 1, \quad Y_{A0} \to A\,X_0, \quad Y_{B1} \to B\,X_1 \end{aligned}$$

Gramática 2

Original

$$\begin{aligned} S &\to aAa \mid bBb \mid \varepsilon \\ A &\to C \mid a \\ B &\to C \mid b \\ C &\to CDE \mid \varepsilon \\ D &\to A \mid B \mid ab \end{aligned}$$

(E aparece pero no tiene producciones propias.)

1) Eliminación de ε

Anulables: $C \Rightarrow \varepsilon \Rightarrow A, B$ anulables ($A \Rightarrow C$, $B \Rightarrow C$). Además $D \Rightarrow A$ o B, por lo que D también puede anularse. S es anulable por $S \to \varepsilon$.

Quitamos ε explícitas y generamos variantes:

- $S \to aAa$ produce aAa y aa (si A anula).
  $S \to bBb$ produce bBb y bb (si B anula).
  $S \to \varepsilon$ se elimina.
- $A \to C \mid a$ se mantiene (sin ε directo en A).

- $B \to C \mid b$ se mantiene.
- $C \to CDE$ con C, D anulables y E no definido (no anulable): genera $CDE \mid DE \mid CE \mid E$. Se quita $C \to \varepsilon$.
- $D \to A \mid B \mid ab$ se mantiene.

$$
\begin{aligned}
S &\to aAa \mid aa \mid bBb \mid bb \\
A &\to C \mid a \\
B &\to C \mid b \\
C &\to CDE \mid DE \mid CE \mid E \\
D &\to A \mid B \mid ab
\end{aligned}
$$

2) Eliminación de unitarias

Unitarias: $A \to C$, $B \to C$, $D \to A$, $D \to B$, y dentro de C está $C \to E$ (unitaria).

- A hereda no unitarias de C: CDE, DE, CE, E; y conserva a.
- B hereda lo mismo y conserva b.
- D hereda de A (a, y también C-expansiones) y de B (b, y C-expansiones), además de ab.
- En C, eliminamos la unitaria a E y conservamos CDE, DE, CE.

$$
\begin{aligned}
S &\to aAa \mid aa \mid bBb \mid bb \\
A &\to a \mid CDE \mid DE \mid CE \\
B &\to b \mid CDE \mid DE \mid CE \\
C &\to CDE \mid DE \mid CE \\
D &\to ab \mid a \mid b \mid CDE \mid DE \mid CE
\end{aligned}
$$

3) Símbolos inútiles

Productores: E no tiene producciones propias $\Rightarrow$ cualquier regla que contenga E no puede derivar sólo terminales. Por tanto, C (cuyas reglas requieren E en CDE, DE, CE) *no* produce terminales. Eliminamos E y C y todas las reglas que los mencionen.

$$
\Rightarrow \quad
\begin{aligned}
S &\to aAa \mid aa \mid bBb \mid bb \\
A &\to a \\
B &\to b \\
D &\to ab \mid a \mid b
\end{aligned}
$$

Alcanzables desde S: aparecen A y B en los cuerpos de S. D ya no es alcanzable (no aparece desde S) $\Rightarrow$ se elimina D.

$$
\boxed{
\begin{aligned}
S &\to aAa \mid aa \mid bBb \mid bb \\
A &\to a \\
B &\to b
\end{aligned}
}
$$

4) Conversión a FNC (CNF)

Introducimos variables para terminales en cuerpos de longitud ≥ 2:

$$T_a \to a, \qquad T_b \to b.$$

Factorizamos pares terminal–no terminal y triples:

$$Y_{Aa} \to A\,T_a, \qquad Y_{Bb} \to B\,T_b,$$

y descomponemos aAa y bBb en binarias:

$$\begin{aligned}
S &\to T_a\,Y_{Aa} \mid T_a\,T_a \mid T_b\,Y_{Bb} \mid T_b\,T_b \\
A &\to a \\
B &\to b \\
Y_{Aa} &\to A\,T_a, \qquad Y_{Bb} \to B\,T_b, \qquad T_a \to a,\ T_b \to b
\end{aligned}$$

Como la gramática original tenía $S \to \varepsilon$, si queremos preservar ε en CNF, añadimos S_0:

$$S_0 \to S \mid \varepsilon.$$

Gramática 3

Original

$$\begin{aligned}
S &\to ASA \mid aB \\
A &\to B \mid S \\
B &\to b \mid \varepsilon
\end{aligned}$$

1) Eliminación de ε

Anulables: $B \Rightarrow \varepsilon \Rightarrow A$ es anulable ($A \Rightarrow B$). S no es anulable (siempre produce al menos una a por aB, o mantiene S dentro de ASA).

Generamos variantes (sin vacío):

$$\begin{aligned}
S &\to ASA \mid SA \mid AS \mid aB \mid a \\
A &\to B \mid S \\
B &\to b
\end{aligned}$$

2) Eliminación de unitarias

Unitarias: $A \to B$ y $A \to S$. Sustituyendo:

$$A \Rightarrow \{\, b \,\} \cup \{\, ASA,\ SA,\ AS,\ aB,\ a \,\}.$$

$$\Rightarrow \quad \begin{aligned}
S &\to ASA \mid SA \mid AS \mid aB \mid a \\
A &\to ASA \mid SA \mid AS \mid aB \mid a \mid b \\
B &\to b
\end{aligned}$$

3) Símbolos inútiles

Productores: $B \to b$; A produce a o b; S produce a.
Alcanzables desde S: se alcanzan A y B. No hay símbolos inútiles.

4) Conversión a FNC (CNF)

Introducimos variable para a en cuerpos binarios y agrupamos pares:

$$T_a \to a, \qquad X_{AS} \to A\,S.$$

Convertimos ASA a binario en dos pasos ($ASA \Rightarrow (AS)A \Rightarrow X_{AS}\,A$):

$$\begin{aligned}
S &\to X_{AS}\,A \mid S\,A \mid A\,S \mid T_a\,B \mid a \\
A &\to X_{AS}\,A \mid S\,A \mid A\,S \mid T_a\,B \mid a \mid b \\
B &\to b \\
X_{AS} &\to A\,S, \qquad T_a \to a
\end{aligned}$$

Todas las producciones son del tipo variable→variable variable o variable→terminal. Aquí $\varepsilon \notin L$, así que no hace falta S_0.

Observación final

En FNC se permiten únicamente reglas $V \to V_1V_2$ (dos variables) o $V \to t$ (terminal), y opcionalmente $S_0 \to \varepsilon$ si la cadena vacía pertenece al lenguaje. Cuando un terminal aparece en un cuerpo de longitud ≥ 2, se introduce una variable que lo produzca (p.ej. $T_a \to a$) y se usa en su lugar.