PRÁCTICA 6: *Lenguajes Formales y Gramáticas*

Pamela Viale Natalia Colussi Alejandro Hernández Valeria Perez Mogetta

Definición 1 Una gramática se dice:
- *regular* si cada producción es de la forma:

$$A \to a,\ A \to aB \ \text{ó}\ A \to \lambda \ \text{donde}\ A, B \in N \ \text{y}\ a \in T;$$

- *libre (o independiente) de contexto* si cada producción es de la forma:

$$A \to \delta \ \text{donde}\ A \in N \ \text{y}\ \delta \in (N \cup T)^*;$$

- *sensible al contexto* si cada producción es de la forma:

$$\alpha A \beta \to \alpha \delta \beta \ \text{donde}\ A \in N,\ \alpha, \beta \in (N \cup T)^* \ \text{y}\ \delta \in (N \cup T)^+ ;\ \text{con el agregado de la regla}$$
$$\sigma \to \lambda \ \text{siendo}\ \sigma \ \text{el estado inicial}$$

- *estructurada por frases o irrestricta* si no tiene restricciones sobre la forma de sus producciones, es decir si son de la forma:

$$\alpha \to \delta \ \text{donde}\ \alpha \in (N \cup T)^* - T^* \ \text{y}\ \delta \in (N \cup T)^*.$$

1. Clasifique cada una de las siguientes gramáticas (dando su tipo más restrictivo):

   (a) $T = \{a, b\}$, $N = \{\sigma, A\}$, símbolo inicial σ, y producciones:

   i. $\sigma \to b\sigma$
   ii. $\sigma \to aA$
   iii. $\sigma \to b$
   iv. $A \to a\sigma$
   v. $A \to bA$
   vi. $A \to a$

   (b) $T = \{a, b, c\}$, $N = \{\sigma, A, B\}$, símbolo inicial σ, y producciones:

   i. $\sigma \to AB$
   ii. $AB \to BA$
   iii. $A \to aA$
   iv. $B \to Bb$
   v. $A \to a$
   vi. $B \to b$

   (c) $T = \{a, b\}$, $N = \{\sigma, A, B\}$, símbolo inicial σ, y producciones:

i. $\sigma \to A$
ii. $\sigma \to AAB$
iii. $Aa \to ABa$
iv. $A \to aa$
v. $Bb \to ABb$
vi. $AB \to ABB$
vii. $B \to b$

(d) $T = \{a, b, c\}$, $N = \{\sigma, A, B\}$, símbolo inicial σ, y producciones:

i. $\sigma \to BAB$
ii. $\sigma \to ABA$
iii. $A \to AB$
iv. $B \to BA$
v. $A \to aA$
vi. $A \to ab$
vii. $B \to b$

2. Dé una derivación de las siguientes cadenas en las gramáticas especificadas:

   (a) $bbabbab$, gramática $1a$
   (b) $abab$, gramática $1b$
   (c) $aabaab$, gramática $1c$
   (d) $abbabb$, gramática $1d$

3. Muestre que la cadena $abbbabaaba$ no está en el lenguaje generado por la gramática $G = (N, T, P, \sigma)$, donde $N = \{\sigma, A, B\}$, $T = \{a, b\}$, símbolo inicial σ, y producciones:

$$\sigma \to aaBA,\ \sigma \to ABB,\ A \to aaB,\ A \to \lambda,\ aBa \to A,\ Aaa \to B,\ B \to AabaB,\ B \to bbb$$

   Sugerencia: Piense en toda derivación como una secuencia de cadenas sobre el alfabeto $N \cup T$ y halle un invariante, verificado por la cadena σ, que se preserve en la aplicación de todas las reglas de producción pero no sea verificado por la cadena propuesta.

4. Dé una gramática del tipo pedido que genere los siguientes lenguajes:

   - Gramática regular:
     (a) Cadenas sobre el alfabeto $\{a, b\}$ que comiencen con a.
     (b) Cadenas sobre el alfabeto $\{a, b\}$ que contengan exactamente una a y terminen con al menos una b.
     (c) Cadenas sobre el alfabeto $\{a, b\}$ que terminen con ba.
     (d) Cadenas sobre el alfabeto $\{a, b\}$ que contengan ba.
     (e) Cadenas sobre el alfabeto $\{a, b\}$ que no terminen con ab.

- Gramática independiente de contexto:
  - (f) Cadenas sobre el alfabeto $\{a, b\}$ de la forma $a^n b^n$ para $n \geq 0$.
  - (g) Cadenas sobre el alfabeto $\{true, p, q, \wedge, \vee, \neg, (,)\}$ definido inductivamente como el menor conjunto tal que:
    - i. $true \in \mathcal{L}$
    - ii. $p \in \mathcal{L}$
    - iii. $q \in \mathcal{L}$
    - iv. Si $A \in \mathcal{L}$ entonces $(\neg A) \in \mathcal{L}$.
    - v. Si $A, B \in \mathcal{L}$ entonces $(A \wedge B) \in \mathcal{L}$.
    - vi. Si $A, B \in \mathcal{L}$ entonces $(A \vee B) \in \mathcal{L}$.
  - (h) Cadenas palíndromes[1] sobre el alfabeto $\{a, b\}$.
  - (i) $\left\{a^n b^n c^k \mid n, k \in \mathbb{N}\right\}$
  - (j) Cadenas sobre el alfabeto $\{a, b\}$ que empiezan y terminan con el mismo símbolo.
  - (k) Cadenas sobre el alfabeto $\{a, b\}$ de longitud impar.
  - (l) Cadenas sobre el alfabeto $\{a, b, c\}$ de longitud impar y cuyo símbolo central es c.
  - (m) $\{a^n b^{2n} c^m \mid n, m \in \mathbb{N}_0\}$
  - (n) $\{a^n b^m c^m \mid n, m \in \mathbb{N}_0\}$

5. Sea $\mathcal{L}$ el conjunto de cadenas sobre $\{a, b\}$ que contienen la misma cantidad de símbolos a y b. Analice si cada una de las siguientes gramáticas genera $\mathcal{L}$. En caso negativo, dé un contraejemplo (es decir, una cadena que sea generada por la gramática pero no esté en $\mathcal{L}$, o una cadena que esté en $\mathcal{L}$ pero no sea generada por la gramática). En todas las gramáticas el símbolo inicial es S.
   - (a) $S \to aSb \mid bSa \mid \lambda$
   - (b) $S \to aSb \mid bSa \mid \lambda \mid SS$
   - (c) $S \to aB \mid bA \mid \lambda$, $B \to b \mid bA$, $A \to a \mid aB$
   - (d) $S \to aSb \mid baS \mid abS \mid bSa \mid \lambda$
   - (e) $S \to aB \mid bA$, $A \to a \mid Sa$, $B \to b \mid SB$

Definición 2 Una gramática se dice que está en *forma normal de Chomsky* sii toda producción es de la forma:

$$A \to BC,\ A \to a \ \text{ ó } \ S \to \lambda$$

donde $A, B, C \in N$, $a \in T$, S es el no terminal inicial y B y C son distintos de S.

[1] Una cadena $x_1 x_2 \dots x_n$ es palíndrome sii $x_1 x_2 \dots x_n = x_n \dots x_2 x_1$

6. Convierta a una gramática equivalente que esté en forma normal de Chomsky:

   - $S \to xSy$
   - $S \to wNz$
   - $N \to S$
   - $N \to \lambda$

7.

   (a) Se puede probar que el lenguaje $\mathcal{L} = \{a^n b^n c^n \mid n \in \mathbb{N}\}$ no es independiente de contexto. Usando este resultado, probar que el conjunto de los lenguajes independientes de contexto no es cerrado bajo la operación de intersección, hallando dos lenguajes que sean independientes de contexto cuya intersección sea $\mathcal{L}$.

   (b) Probar que el conjunto de los lenguajes independientes de contexto no es cerrado bajo la operación de complemento.

8. Se puede probar que la intersección de un lenguaje independiente del contexto y uno regular es independiente de contexto. Usar este resultado y el enunciado del ejercicio 7a para probar que el lenguaje $\{w \in \{a, b, c\}^* \mid N_a(w) = N_b(w) = N_c(w)\}$ no es independiente del contexto.

   Notación: $N_x(w)$ denota la cantidad de ocurrencias del símbolo x en la cadena w.

9. Probar que si $\mathcal{L}$ es un lenguaje independiente de contexto, entonces $\mathcal{L}^R$ también es independiente de contexto.

10. Sea la gramática con producciones:

$$S \to aS \mid Sb \mid a \mid b$$

   y símbolo inicial S. Pruebe, por inducción sobre el número de pasos en la derivación, que toda cadena generada por la gramática no contiene la subcadena ba.

Definición 3 Se dice que un árbol es un *árbol de parseo* de una gramática libre de contexto $G = (N, T, P, S)$ si:

- cada nodo tiene una etiqueta en $N \cup T \cup \{\lambda\}$
- la etiqueta de la raíz es S
- las etiquetas de los nodos interiores están en N
- si el nodo n tiene etiqueta A e hijos $n_1, \ldots, n_k$ (de izquierda a derecha) con etiquetas $X_1, \ldots, X_k$, entonces $A \to X_1 \ldots X_k$ es una producción en P

- si un nodo n tiene etiqueta λ, entonces n es una hoja y es hijo único

Una gramática libre de contexto G es *ambigua* si alguna palabra en $L(G)$ tiene más de un árbol de parseo.

Un lenguaje libre de contexto para el cual toda gramática es ambigua se dice *inherentemente ambiguo*.

11. Considere la gramática libre de contexto $G = (\{E, V\}, \{+, *, (,), x, y, z\}, P, E)$ donde P consiste de las producciones:

$$E \to E + E \mid E * E \mid V \mid (E)$$

$$V \to x \mid y \mid z$$

   (a) Pruebe que la gramática es ambigua.

   (b) Dé una gramática libre de contexto equivalente que resuelva la ambigüedad mediante la precedencia usual entre la multiplicación y la suma.