PRÁCTICA 6: *Lenguajes Formales y Gramáticas*

Pablo Verdes Dante Zanarini Pamela Viale Alejandro Hernandez Mauro Lucci

Definición 1 Una gramática se dice:
regular si cada producción es de la forma[1]:

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

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:

$$\alpha A \beta \to \alpha \delta \beta \text{ donde } A \in N,\ \alpha, \beta \in (N \cup T)^* \text{ y } \delta \in (N \cup T)^+$$

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:

   1) $\sigma \to b\sigma$.
   2) $\sigma \to aA$.
   3) $\sigma \to b$.
   4) $A \to a\sigma$.
   5) $A \to bA$.
   6) $A \to a$.

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

   1) $\sigma \to AB$.
   2) $AB \to BA$.
   3) $A \to aA$.
   4) $B \to Bb$.
   5) $A \to a$.
   6) $B \to b$.

[1] Algunas definiciones permiten reemplazar a por una cadena de uno o más terminales; se puede mostrar (ver ejercicio 6) que las dos definiciones son equivalentes.

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

1) $\sigma \to A$.
2) $\sigma \to AAB$.
3) $Aa \to ABa$.
4) $A \to aa$.
5) $Bb \to ABb$.
6) $AB \to ABB$.
7) $B \to b$.

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

1) $\sigma \to BAB$.
2) $\sigma \to ABA$.
3) $A \to AB$.
4) $B \to BA$.
5) $A \to aA$.
6) $A \to ab$.
7) $B \to b$.

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

   a) Cadena $bbabbab$ en la gramática $1a$.

   b) Cadena $abab$ en la gramática $1b$.

   c) Cadena $aabaab$ en la gramática $1c$.

   d) Cadena $abbabb$ en la gramática $1d$.

3. Muestre que la cadena $abbbabaaba$ no esta 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 $T \cup N$ 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. De 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:
    1) $true \in \mathcal{L}$.
    2) $p \in \mathcal{L}$.
    3) $q \in \mathcal{L}$.
    4) Si $A \in \mathcal{L}$, entonces $(\neg A) \in \mathcal{L}$.
    5) Si $A, B \in \mathcal{L}$, entonces $(A \wedge B) \in \mathcal{L}$.
    6) Si $A, B \in \mathcal{L}$, entonces $(A \vee B) \in \mathcal{L}$.
  - *h*) Cadenas palindromos[2] sobre el alfabeto $\{a, b\}$.
  - *i*) $\left\{a^n b^n c^k / n, k \in \mathbb{N}\right\}$.
  - *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 una c.
  - *m*) $\{a^n b^{2n} c^m / n, m \in \mathbb{N}_0\}$.
  - *n*) $\{a^n b^m c^m / n, m \in \mathbb{N}_0\}$.

[2]Una cadena $x_1 x_2 \ldots x_n$ es palindromo sii $x_1 x_2 \ldots x_n = x_n \ldots x_2 x_1$

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, de un contraejemplo (es decir, una cadena generada por la gramática pero que no esta en $\mathcal{L}$, o una cadena que esta en $\mathcal{L}$ pero no es generada por la gramática). En todas las gramáticas el símbolo inicial es S.

   a) $S \to aSb|bSa|\lambda$.
   b) $S \to aSb|bSa|\lambda|SS$.
   c) $S \to aB|bA|\lambda$, $B \to b|bA$, $A \to a|aB$.
   d) $S \to aSb|baS|abS|bSa|\lambda$.
   e) $S \to aB|bA$, $A \to a|Sa$, $B \to b|SB$.

6.

   a) Muestre que si cada producción de una gramática G es de la forma $A \to \alpha$, $A \to \alpha B$ o $A \to \lambda$ donde $A, B \in N$ y $\alpha \in T^+$ entonces existe una gramática regular G' equivalente (es decir, tal que $L(G') = L(G)$).
   b) Aplique el apartado anterior para modificar la gramática siguiente (símbolo inicial: S) de manera de formar una gramática regular equivalente:

      - $S \to yX$.
      - $X \to xxX$.
      - $X \to yY$.
      - $Y \to \lambda$.

7. Muestre que un lenguaje regular que no contiene a λ puede ser generado por una gramática que no contiene reglas de la forma $A \to \lambda$.

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

$$A \to BC \text{ o } A \to a \text{ o } S \to \lambda$$

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

8. Convierta a una gramática equivalente que este en forma normal de Chomsky:

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

9.

   a) Se puede probar que el lenguaje $\mathcal{L} = \{a^n b^n c^n | n \in \mathbb{N}\}$ 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.

10. 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 9a para probar que el lenguaje $\{w \in \{a,b,c\}^* / N_a(w) = N_b(w) = N_c(w)\}$ no es independiente del contexto.

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

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

12. Sea la gramática con producciones:

$$S \rightarrow aS|Sb|a|b$$

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

13. Sea la gramática con producciones

$$S \rightarrow aSbS|bSaS|\lambda$$

    y símbolo inicial S. Pruebe por inducción sobre el numero de pasos en la derivación, que toda cadena generada por la gramática tiene la misma cantidad de a's y b's.

14. Probar que la gramática

$$S \rightarrow \lambda|0|1|0S0|1S1$$

    con símbolo inicial S genera el lenguaje formado por las cadenas palindromos sobre el alfabeto $\{0,1\}$.

Definición 3 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 mas de un árbol de parseo.

Un lenguaje libre de contexto para el que toda gramática es ambigua, es inherentemente ambiguo.

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

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

$$V \to x | y | z$$

   a) Probar que la gramática es ambigua.
   b) Dar una gramática libre de contexto equivalente que resuelva la ambigüedad mediante la precedencia usual entre la multiplicación y la suma.