Práctica 6: Lenguajes formales y gramáticas

- Una gramática posee los siguientes símbolos:
  - inicial,
  - no terminales (son estructuras intermedias), y
  - terminales (son elementos del alfabeto).

Definición: Gramática es una tupla: (N, T, P, σ) donde:

- N es un conjunto finito de símbolos llamados **no terminantes.**
- T es un conjunto finito de símbolos, llamados **terminantes** o **alfabeto**, tal que $N \cap T = \emptyset$
- P es un conjunto finito de **reglas de producción**, donde

$$P \subseteq ((N \cup T)^* - T^*) \times (N \cup T)^*$$

Obs.:

- $(N \cup T)^* - T^*$ es el conjunto de cadenas de no terminales y terminales que contienen al menos un no terminal. Dado que $\lambda \notin (N \cup T)^* - T^*$ no puede haber reglas del tipo $\lambda \to \beta$.
- $(N \cup T)^*$ es el conjunto de cadenas sobre $N \cup T$.
- A una producción de la forma (α, β) la notaremos $\alpha \to \beta$.
- Una regla del tipo $\alpha \to \lambda$ recibe el nombre de **regla** λ.

- $\sigma \in N$ es el **símbolo inicial**

$$G_1 = (N, T, P, \sigma)$$

donde:

- $N = \{\sigma\}$
- $T = \{a, b\}$
- P está dado por las reglas de producción:

$$\sigma \to a\sigma bb$$
$$\sigma \to abb$$

- σ es el símbolo inicial

Definición 1. Una gramática se dice:

(a) *regular* si cada producción es de la forma: $A \to a$ o $A \to aB$ o $A \to \lambda$ donde $A, B \in N$ y $a \in T$,

Gramáticas regulares (tipo 3)

- Las gramáticas regulares o de tipo 3, también llamadas lineales, pueden ser clasificadas como derechas o izquierdas.
- Las reglas de producción de una gramática regular derecha se adhieren a las siguientes restricciones:
  - El lado izquierdo debe consistir en un solo no terminal.
  - El lado derecho está formado por un símbolo terminal, que puede estar seguido (o no) por un símbolo no terminal, o la cadena vacía.
- Es decir, las producciones de una gramática regular derecha pueden tener la forma:

$$A \to a$$
$$A \to aB$$
$$A \to \lambda$$

donde $A, B \in N$ y $a \in T$.

Gramáticas regulares (tipo 3)

- Alternativamente, en una gramática regular izquierda las reglas de producción son de la forma:

$$A \to a$$
$$A \to Ba$$
$$A \to \lambda$$

- Por ejemplo, reglas de la forma:

$$yW \to x$$
$$X \to xZy$$
$$YX \to WvZ$$

no están permitidas en una gramática regular. (¿Por qué?)

- Se puede convertir toda gramática regular derecha en izquierda y recíprocamente.

(b) *libre* (o independiente) de contexto si cada producción es de la forma $A \to \delta$ donde $A \in N$ y $\delta \in (N \cup T)^*$.
Otra definición:

- Como en las regulares, el lado izquierdo debe consistir en un solo no terminal.
- A diferencia de las regulares, no hay restricciones sobre la forma del lado derecho.

- Es decir, las reglas de producción tienen la forma:

$$A \to \delta$$

donde $A \in N$ y $\delta \in (N \cup T)^*$.

- El término *independiente del contexto* hace referencia a que, debido a que en el lado izquierdo el no terminal aparece solo, la regla se puede aplicar sin importar el contexto en que aparezca dicho no terminal.

- En contraste, la regla $xNy \to xzy$ permite reemplazar el no terminal N por z sólo cuando se encuentre en el contexto de x e y.

Gramáticas independientes del contexto (tipo 2)

- Dado que no hay restricciones sobre la forma de δ en $A \to \delta$, podría ocurrir que δ contenga más de un no terminal, como en $A \to zXYz$.
- Pero ¿cuál de los no terminales reemplazamos en el paso siguiente?
- El enfoque más común es el de la **derivación por la izquierda**, que consiste en reemplazar el no terminal situado más a la izquierda.
- Análogamente, se podría aplicar derivación por la derecha, o seguir algún otro patrón.
- Resulta ser que el orden en que se apliquen las reglas no afecta la determinación de si una cadena puede ser generada por la gramática o no.
- Esto es consecuencia de que *si existe una derivación que genera una cadena, entonces también existe una derivación por izquierda que la genera.*

(c) *sensible al contexto* si cada producción es de la forma $\alpha A\beta \to \alpha\delta\beta$ donde $A \in N, \alpha, \beta \in (N \cup T)^*$ y $\delta \in (N \cup T)^+$.

- Obs.:
  - La producción $abA \to baab$ no es sensible al contexto.
    Lo sería si fuese $abA \to abab$.
  - La producción $aSb \to ab$ no es de tipo 1, pues $\delta = \lambda \notin (N \cup T)^+$.

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

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

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

   $$\sigma \to b\sigma, \sigma \to aA, A \to a\sigma,$$

   $$A \to bA, A \to a, \sigma \to b$$

   Regular.

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

$$\sigma \to AB, AB \to BA, A \to aA,$$

$$B \to Bb, A \to a, B \to b$$

Sensible al contexto.

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

$$\sigma \to A, \quad \sigma \to AAB, \quad Aa \to ABa, \quad A \to aa,$$

$$Bb \to ABb, \quad AB \to ABB, \quad B \to b.$$

Sensible al contexto.

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

$$\sigma \to BAB, \quad \sigma \to ABA, \quad A \to AB, \quad B \to BA,$$

$$A \to aA, \quad Aab, \quad B \to b.$$

Independiente de contexto (libre).

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

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

$$\begin{gathered} \sigma \\ =< \sigma \to b\sigma > \\ b\sigma \\ =< \sigma \to b\sigma > \\ bb\sigma \\ =< \sigma \to aA > \\ bbaA \\ =< A \to bA > \\ bbabA \\ =< A \to bA > \\ bbabbA \\ =< A \to a\sigma > \\ bbabba\sigma \\ =< \sigma \to b > \\ bbabbab. \end{gathered}$$

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

$$\begin{gathered} \sigma \\ =< \sigma \to AB > \\ AB \\ =< A \to aA, B \to Bb > \\ aABb \\ =< AB \to BA > \\ aBAb \\ =< A \to a, B \to b > \\ abab. \end{gathered}$$

c)

$$\sigma$$
$$=< \sigma \to AAB >$$
$$AAB$$
$$=< B \to b >$$
$$AAb$$
$$=< A \to aA >$$
$$Aaab$$
$$=< Aa \to ABa >$$
$$ABaab$$
$$=< B \to b >$$
$$Abaab$$
$$=< A \to aa >$$
$$aabaab$$

d)

$$\sigma$$
$$=< \sigma \to ABA >$$
$$ABA$$
$$=< A \to AB$$
$$ABAB$$
$$=< B \to b >$$
$$ABAb$$
$$=< A \to ab >$$
$$ABabb$$
$$=< B \to b >$$
$$Ababb$$
$$=< A \to ab >$$
$$abbabb$$

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:

$$\sigma \to aaBA, \quad \sigma \to ABB, \quad A \to aaB, \quad A \to \lambda$$

$$aBa \to A, \quad Aaa \to B, \quad B \to AabaB, \quad B \to bbb$$

   Ya sea, comenzando con $\sigma \to aaBA$, o $\sigma \to ABB$, al aplicar cualquiera de las reglas de producción posibles, siempre se llega a algo con final en B o bbb. Este final es imposible de cambiar por a. Por lo tanto $abbbabaaba$ no está en dicho lenguaje.

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

   - Gramática regular:
     I. Cadenas sobre el alfabeto $\{a, b\}$ que comiencen con a.
     II. Cadenas sobre el alfabeto $\{a, b\}$ que contengan exactamente una a y terminen con al menos una b.
     III. Cadenas sobre el alfabeto $\{a, b\}$ que terminen con ba.
     IV. Cadenas sobre el alfabeto $\{a, b\}$ que contengan ba.
     V. Cadenas sobre el alfabeto $\{a, b\}$ que no terminen con ab.
   - a) Cadenas sobre el alfabeto $\{a, b\}$ de la forma $a^n b^n$ para $n \geq 0$.

b) Cadenas s

5. Sea $\mathcal{L}$ el ocnjunto 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 generada por la gramática pero que no está en $\mathcal{L}$, o una cadena que está en $\mathcal{L}$ pero no es generada por la gramática). En todas las gramáticas el símbolos inicial es S.

   a) $S \to aSb|bSa|\lambda$

6. I. Muestre que si cada producción de una gramática G es de la forma.

   $$A \to \quad o \quad A \to \alpha B \quad o \quad A \to \lambda \quad \text{donde} \quad A, B \in N \quad y \quad \alpha \in T^+$$

   entonces existe una gramática regular G' equivalente (es decir, tal que $L(G') = L(G)$).

   II. Aplique el apartado anterior para modificar la gramática siguiente (símbolo inicial: S) de manera de formar una gramática regular equivalente.

   $$S \to yX$$
   $$X \to xxX$$
   $$X \to yY$$
   $$Y \to \lambda$$
   $$A \to xX$$

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

   $$A \to BC \quad o \quad A \to a \quad o \quad S \to \lambda$$

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

8. Convertir a la forma normal de Chomsky:

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