

Gramáticas

Dado un conjunto de símbolos o alfabeto V , el conjunto de todas cadenas posibles sobre ese alfabeto es V^* (el monoide libre sobre V mediante el operador concatenación) y V^+ será el mismo conjunto sin la cadena vacía.

Un lenguaje sobre el alfabeto V es un subconjunto de V^* . Existen un número no enumerable de lenguajes, todos ellos representables mediante gramáticas.

Una gramática se define como la cuádrupla $G=(N,V,P,S)$, donde V es el alfabeto de G y N es un conjunto finito de símbolos **no** terminales, $V \cap N = \emptyset$. Se conoce a $W=V \cup N$ como el vocabulario de G , una frase $[zeta] \in W^*$ es una cadena de símbolos de W . $P \subseteq W^* \times W^*$ es un conjunto de reglas de producción o reescritura, que transforman una frase (en la que por lo menos hay un no terminal) en otra frase del mismo vocabulario. Estas reglas se denotan como $[zeta]_1 A [zeta]_2 \rightarrow [zeta]_3$, donde $[zeta]_1, [zeta]_2, [zeta]_3 \in W^*$ y $A \in N$. Finalmente, $S \in N$ es el símbolo no terminal inicial o axioma de G .

Una gramática es una **estructura algebraica** formada por cuatro elementos fundamentales:

$$G = \{ NT, T, S, P \}$$

Donde:

- NT es el conjunto de elementos **No Terminales**
- T es el conjunto de elementos **Terminales**
- S es el **Símbolo inicial** de la gramática
- P es el conjunto de **Reglas de Producción**

Al proceso de obtener una derivación de una cadena a partir del axioma de una gramática y aplicando reglas de ésta, se le denomina *análisis sintáctico*. Se dice que una gramática es *ambigua* si para alguna cadena del lenguaje existe más de una posible derivación para generarla.

Ejemplo:

Derivación en G : $S \Rightarrow McF \Rightarrow aMcAF \Rightarrow aMcaF \Rightarrow aaMcAaF \Rightarrow aaMcaAF \Rightarrow aaMcaaF \Rightarrow aabMcBaaF \Rightarrow aabMcaBaF \Rightarrow aabMcaaBF \Rightarrow aabMcaabF \Rightarrow aabcXaabF \Rightarrow aabcaXabF \Rightarrow aabcaaXbF \Rightarrow aabcaabXF \Rightarrow aabcaab$

Chomsky dividió las gramáticas (y por lo tanto los lenguajes) en una jerarquía que va del tipo 0 a 3, en orden decreciente de complejidad, y en base a la forma de sus reglas:

0 o "No restringida o recursivamente e numerables"

$$\begin{aligned} & x \rightarrow y \\ & x \in (NT/T)^+ \\ & y \in (NT/T)^* \end{aligned}$$

- "x puede ser sustituido por y si x está, ya sea, en los símbolos **No Terminales** o los símbolos **Terminales**, sin incluir la cadena vacía e y está en los símbolos **No Terminales** o **Terminales**, incluyendo la cadena vacía."
 - Los lenguajes generados por este tipo de gramáticas se llaman "lenguajes sin restricciones"
 - Nota: "+" significa "sin incluir la cadena vacía" y "*" significa "incluyendo la cadena vacía". "/" significa "o"
 - Estos lenguajes también son denominados "recursivamente enumerables"
- Las máquinas que los aceptan son las máquinas de Turing (y equivalentes no deterministas).

1 o "Sensible al contexto"

$$\begin{aligned} & \alpha \rightarrow \beta; |\alpha| \leq |\beta| \\ & \alpha = z_1 x z_2 \\ & \beta = z_1 y z_2 \\ & z_1, z_2 \in T^* \\ & x \in NT \\ & y \in (NT/T)^+ \end{aligned}$$

- "α puede ser reemplazado por β si la longitud de α es menor o igual a la longitud de β, siendo α un símbolo **Terminal** o una cadena vacía z₁, seguido de un símbolo **No Terminal** X, seguido de otro símbolo **Terminal** o una cadena vacía z₂. En el caso de β, z₁ debe ser el mismo símbolo z₁ de α seguido de un símbolo **No Terminal** o **Terminal** sin ser la cadena vacía, seguido del símbolo z₂."
- Las máquinas que los aceptan son autómatas linealmente acotados(linear-bounded).

2 o "Libre de contexto"

$$\begin{aligned}x &\rightarrow y \\ x &\in NT \\ y &\in (NT/T)^*\end{aligned}$$

“x puede ser reemplazado por y si x pertenece a los símbolos **No Terminales** e y es un **Terminal** o **No Terminal**, incluyendo la cadena vacía.”

Máquinas que los pueden leer:

Máquinas que los aceptan: Autómata a Pila (Pushdown Automaton)

3 o "Regular"

$$\begin{aligned}\alpha &\rightarrow \beta \\ \alpha &\in NT \\ \beta &\in \begin{cases} aB \\ Ba \\ b \end{cases} \\ B &\in NT \\ a &\in T^+ \\ b &\in T^*\end{aligned}$$

También llamada "De contexto regular"

“α puede ser reemplazado por β si α pertenece a los símbolos **No Terminales** y β es uno de estos 3:

- Un símbolo **Terminal** no nulo seguido de un **No Terminal**.
- Un símbolo **No Terminal** seguido de un símbolo **Terminal** no nulo.
- Un símbolo **Terminal** pudiendo ser la cadena vacía.”

Máquinas que los aceptan: autómata finito, determinista o no determinista.