

INF155: Informática Teórica

Clase 12: Gramáticas de Contexto Libre

Aldo Berrios Valenzuela

Jueves 14 de Abril de 2016

1. Gramáticas

Definición 1.1. Una *gramática* $G = (N, \Sigma, P, S)$ consta de:

- N : El alfabeto de *no-terminales*. (es una v (nu) mayúscula, no una n)
- Σ : El alfabeto de *terminales*.
- P : conjunto de *producciones* de la forma $\alpha \rightarrow \beta$ con $\alpha \in (N \cup \Sigma)^*$, conteniendo al menos un símbolo de N , $\beta \in (N \cup \Sigma)^*$
- S : *símbolo de partida*, $S \in N$.

Notar que $N \cap \Sigma = \emptyset$ y $N \cup \Sigma = V$ (*vocabulario*).

Definición 1.2. Sea $G = (N, \Sigma, P, S)$ una gramática. Definimos la *relación de derivación* de G sobre $(N \cup \Sigma)^*$ a través de:

$$u\alpha v \Rightarrow_G u\beta v$$

para $u, v \in (\Sigma \cup N)^*$ si $\alpha \rightarrow \beta \in P$

Definición 1.3. Sea $G = (N, \Sigma, P, S)$ una gramática. Se define el *lenguaje generado* por G como:

$$\mathcal{L}(G) = \{\sigma : S \Rightarrow_G^* \sigma \wedge \sigma \in \Sigma^*\} \quad (1.1)$$

1.1. Jerarquía de Chomsky

Sea $G = (N, \Sigma, P, S)$ una gramática. Se dice que es:

- **Tipo 0 (irrestringida)** – Como definido.
- **Tipo 1 (sensibles al contexto)** – Toda las producciones $\alpha \rightarrow \beta$ cumplen:

$$|\alpha| \leq |\beta| \quad ; \text{ las palabras nunca se acortan} \quad (1.2)$$

Nótese que $\alpha \neq \epsilon$ por definición del conjunto de producciones. (Una manera alternativa de describirlas es decir que las producciones tienen la forma $\alpha A \beta \rightarrow \alpha \gamma \beta$, con $\alpha, \beta \in (N \cup \Sigma)^*$, $A \in N$, $\gamma \in (N \cup \Sigma)^+$)

Acá sólo podemos reemplazar la producción A por γ sólo si lo rodean α y β ... por eso es sensible al contexto...

- **Tipo 2 (contexto libre)** – Todas las producciones tienen la forma $A \rightarrow \alpha$, con $A \in N$, $\alpha \in (N \cup \Sigma)^+$
- **Tipo 3 (regular)** – Todas las producciones sólo pueden ser de la forma:

$$A \rightarrow \alpha B, \quad \text{con } A, B \in N, \alpha \in \Sigma^*$$

$$A \rightarrow \alpha, \quad \text{con } A \in N, \alpha \in \Sigma^+$$

Observación: Las gramáticas de tipo 1 son un caso particular de tipo 0, las de tipo 2 particulares de tipo 1 y así sucesivamente.

Teorema 1.1. Las gramáticas de tipo 3 generan los lenguajes regulares (gramática tipo 3 \Leftrightarrow lenguaje regular).

Demostración. Comenzamos demostrando

$$\text{lenguaje regular} \Rightarrow \text{gramática tipo 3 (sin } \epsilon) \quad (1.3)$$

L regular $\Rightarrow L = \mathcal{L}(M)$, con M un DFA tal que el estado inicial no es final. Sea $M = (Q, \Sigma, \delta, q_0, F)$ el DFA. Sin pérdida de generalidad $Q \cap \Sigma = \emptyset$ (la idea es usar los estados como no terminales).

Definimos la gramática $G = (Q, \Sigma, P, q_0)$. Las producciones son:

- Si $\delta(q, a) = p$, entonces $q \rightarrow ap$
- Si $\delta(q, a) = p$, y p es final, $q \rightarrow a$.

$$\therefore \mathcal{L}(G) = \mathcal{L}(M).$$

Ya demostrado (1.3), nos falta demostrar que

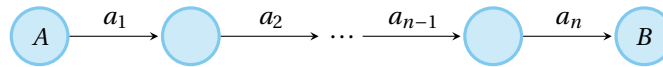
$$\text{gramática tipo 3} \Rightarrow \text{lenguaje regular} \quad (1.4)$$

Sea $G = (N, \Sigma, P, S)$ una gramática de tipo 3. Construimos un NFA que acepte $\mathcal{L}(G)$ vía:

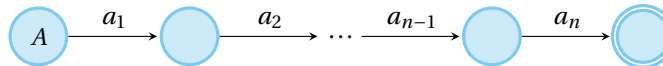
$$M = (N \cup \{q_1, \dots\}, \Sigma, S, F)$$

con:

- Si $A \rightarrow a_1 a_2 \dots a_n B$:



- Si $A \rightarrow a_1 a_2 \dots a_n$:



Con esto, queda demostrado (1.4).

Finalmente, se tiene que las gramáticas de tipo 3 generan lenguajes regulares y viceversa. \square

Dado que las gramáticas de tipo 3 sólo generan lenguajes regulares, podemos modificar nuestro triángulo NFA - DFA - RE agregándole un “cototo”:

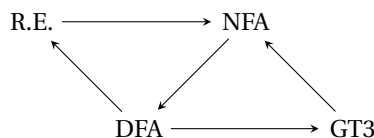


Figura 1: GT3 son las gramáticas de tipo 3 son aquellas que generan R.E.

1.2. Gramáticas de Contexto Libre

Sea $G = (N, \Sigma, P, S)$ una gramática de contexto libre (CFG). Algunas convenciones:

- Mayúsculas son no-terminales (A, B, \dots, S, \dots)
- Otros símbolos (minúsculas, a, b, \dots) son terminales.
- Se anotan sólo las producciones, el símbolo no terminal de la primera es el símbolo de partida.

Ejemplo (ambas producciones forman parte de la misma gramática):

$$S \rightarrow ab$$

$$S \rightarrow aSb$$

Vemos que:

$$S \Rightarrow ab$$

$$S \Rightarrow aSb$$

$$\Rightarrow a^2 S b^2$$

$$\vdots$$

$$\Rightarrow a^{n-1} S b^{n-1}, \quad n \in \mathbb{N}$$

$$\Rightarrow a^n b^n, \quad n \in \mathbb{N}$$

\therefore Hay gramáticas de contexto libre que generan lenguajes que NO son regulares.

Por lo general, cuando hablemos de gramáticas, sólo nombraremos el conjunto de producciones que esta trae para simplificar.

1.2.1. Ejemplos de gramáticas de contexto libre

Un ejemplo de gramática puede ser:

$$E \rightarrow E + T$$

$$E \rightarrow T$$

$$T \rightarrow T * F$$

$$T \rightarrow F$$

$$F \rightarrow (E)$$

$$F \rightarrow a$$

Como puede ver, el CFG anterior representa la sintaxis que se debe seguir al resolver una expresión algebraica. Es muy similar al BNF.

Nótese que existen otras CFG equivalentes a las que se mencionan en el párrafo anterior, por ejemplo:

$$E \rightarrow E + E$$

$$E \rightarrow E * E$$

$$E \rightarrow (E)$$

$$E \rightarrow a$$

También puede ser:

$$E \rightarrow E + T$$

$$E \rightarrow E * T$$

$$E \rightarrow T$$

$$T \rightarrow (E)$$

$$T \rightarrow a$$