



## Gramáticas Libres de Contexto

Este tipo de gramática se usa mucho en **procesamiento de lenguaje natural (PLN)** y **compiladores** para definir **estructuras válidas de una lengua**, ya sea un lenguaje natural (como el español o inglés) o un lenguaje formal (como el código fuente de un programa). Las gramáticas Libre de Contexto se utilizan principalmente para analizar la lingüística para poder entender la estructura del lenguaje natural hablado o escrito. Se estudia la forma de generar o aceptar cadenas o strings de una forma declarativa.

Una gramática libre o independiente del contexto es una notación formal que sirve para expresar las definiciones recursivas de los lenguajes. La gramática:

$$R \rightarrow X R X \mid S$$

$$S \rightarrow a T b \mid b T a$$

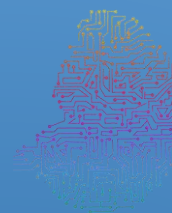
$$T \rightarrow X T X \mid X \mid \varepsilon$$

$$X \rightarrow a \mid b$$

### EXPLICACION DE LAS GRAMÁTICA FORMALES:

Una gramática formal es un conjunto de reglas que nos permite generar cadenas en un lenguaje. Estas reglas nos dicen cómo construir frases o palabras válidas usando símbolos.

Una gramática tiene cuatro componentes principales:





$$G = (V, T, P, S)$$

V: Conjunto de variables (también llamadas no terminales).

T: Conjunto de terminales.

P: Reglas de producción.

S: Símbolo inicial.



### ¿QUÉ ES UN TERMINAL?

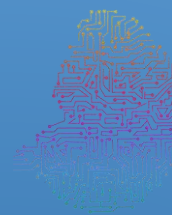
Los terminales son los símbolos finales del lenguaje, es decir, los que aparecen en la cadena generada. No se pueden reemplazar por nada más. Son como las letras o palabras reales de un idioma.



$$X \rightarrow a \mid b$$

Los terminales son 2: a y b.

Una vez que aparece una a o una b, ya no se puede cambiar más.



### ¿QUÉ ES UN NO TERMINAL O VARIABLE?

Son símbolos auxiliares que se usan para generar otras cosas. Aparecen durante la derivación, pero deben desaparecer al final. Se reemplazan usando las reglas de producción hasta llegar a una cadena compuesta solo por terminales.





R, S, T, X

Son 4 variables se usan para derivar, pero no deberían aparecer en la cadena final.

Supongamos que usamos esta parte:

$S \rightarrow a T b$

$T \rightarrow \varepsilon$

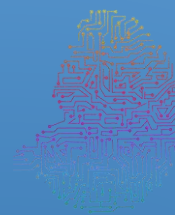
Derivación:

- Comenzamos con S
- Usamos la regla  $S \rightarrow a T b$
- Ahora tenemos a T b
- Sustituimos  $T \rightarrow \varepsilon$  (vacío)
- Resultado: ab

¿QUÉ GENERA ESTA GRAMÁTICA?

Esta gramática genera cadenas formadas solo por las letras a y b, combinadas de formas específicas:

- El símbolo inicial R permite construir cadenas simétricas (espejadas) usando  $X R X$ , o bien producir algo a través de S.
- S genera estructuras como a T b o b T a, lo que aporta un núcleo con un contenido variable entre a y b.





- T produce cadenas intermedias formadas por combinaciones de a y b, incluso puede ser vacío ( $\epsilon$ ).
- X es simplemente un a o b.

Entonces, el lenguaje  $L(G)$  contiene cadenas con una simetría parcial o total, con combinaciones equilibradas entre a y b.

### 3 CADENAS EN $L(G)$

R  $\rightarrow$  S

$\rightarrow$  aT b

$\rightarrow$  aX b

$\rightarrow$  aab

R  $\rightarrow$  XRX

$\rightarrow$  aRb

$\rightarrow$  aSb

$\rightarrow$  aaT bb

$\rightarrow$  aaX bb

$\rightarrow$  aaa bb

R  $\rightarrow$  S

$\rightarrow$  bT a

$\rightarrow$  bXT X a

$\rightarrow$  baX b a

$\rightarrow$  bab b a

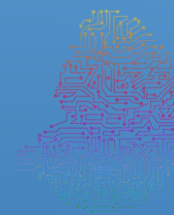
### LA CADENA MÍNIMA POSIBLE

R  $\rightarrow$  S

$\rightarrow$  aT b

$\rightarrow$  a $\epsilon$  b

$\rightarrow$  a b





Ahora podemos responder las preguntas una por una si son verdaderas o falsas:

V o F:  $T \Rightarrow aba$

**Falso** (eso es una derivación completa, no una directa)

V o F:  $T \Rightarrow^* aba$

**Verdadero** (por  $T \rightarrow X \ T \ X \rightarrow a \ \varepsilon \ b \rightarrow aba$ )

V o F:  $T \Rightarrow T$

**Falso** (no hay una regla  $T \rightarrow T$ )

V o F:  $T \Rightarrow^* T$

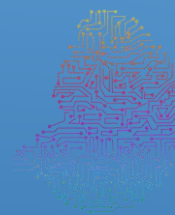
**Verdadero** (por 0 pasos:  $T \Rightarrow^* T$ )

V o F:  $XXX \Rightarrow^* aba$

**Verdadero** ( $X \rightarrow a$  o  $b$ , entonces es posible generar  $aba$ )

V o F:  $X \Rightarrow^* aba$

**Falso** ( $X$  solo genera  $a$  o  $b$ , no cadenas largas)





V o F:  $T \Rightarrow^* XX$

**Verdadero** (T puede derivar (en cero o más pasos) a dos X seguidas)



V o F:  $T \Rightarrow^* XXX$

**Verdadero** (puede generar 3 letras, como aba)

V o F:  $S \Rightarrow^* \varepsilon$

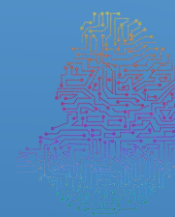
**Falso** (S siempre produce al menos dos letras: a....b o b....a)



## ¿QUÉ ES UN ÁRBOL DE DERIVACIÓN?

Un árbol de derivación (también llamado árbol sintáctico) es una representación gráfica que muestra cómo una cadena se genera a partir de una gramática.

Muestra, paso a paso, cómo se usan las reglas de producción desde el símbolo inicial hasta obtener una cadena de solo terminales.

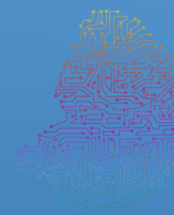
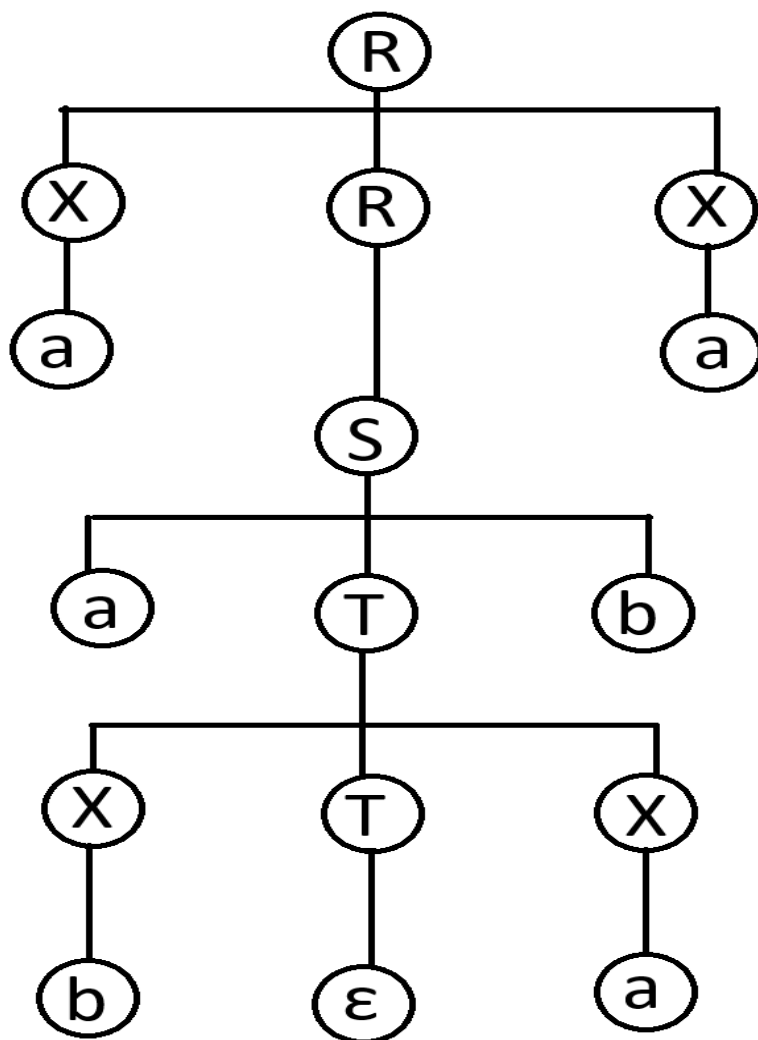


Estructura:





- El nodo raíz es el símbolo inicial de la gramática (por ejemplo, R).
- Cada rama representa una regla de producción que se aplica.
- Las hojas (parte inferior del árbol) son los símbolos terminales que forman la cadena final.
- Las ramas intermedias son variables (no terminales).





Un árbol de derivación te muestra de forma visual cómo una cadena se genera paso a paso usando las reglas de la gramática. Es clave para entender el proceso de derivación y para analizar si una cadena pertenece al lenguaje.



## CONCLUSION:

En este trabajo realizado se responden todas las preguntas.

## Trabajo de Procesamiento de Lenguaje Natural (NLP)



**Tema:**

**Gramática Libre de Contexto**

---

**Alumno:**

**Monzón Cristian**

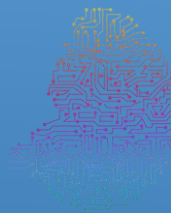
**Profesora:**

**Scudero Yanina Ximena**

**Materia:**

**TECNICAS PROCESAMIENTO HABLA**

**Carrera:**







## Ciencia de Datos e Inteligencia Artificial

---

### Objetivo del trabajo:

1. Responder los siguientes puntos respecto a la gramática libre de contexto G que se muestra a continuación:

- a. ¿Cuántas variables tiene G?
- b. ¿Cuántos terminales tiene G?
- c. ¿Cuál es el símbolo inicial de G?
- d. Dar tres cadenas en  $L(G)$ .
- e. Dar la cadena mínima posible.
- f. V o F:  $T \Rightarrow aba$ .
- g. V o F:  $T \Rightarrow *aba$ .
- h. V o F:  $T \Rightarrow T$ .
- i. V o F:  $T \Rightarrow *T$ .
- j. V o F:  $XXX \Rightarrow *aba$ .
- k. V o F:  $X \Rightarrow *aba$ .
- l. V o F:  $T \Rightarrow *XX$ .
- m. V o F:  $T \Rightarrow *XXX$ .
- n. V o F:  $S \Rightarrow *\epsilon$ .
- ñ. Describa en español el lenguaje  $L(G)$ .
- o. Árbol de derivación: cadena aababa

