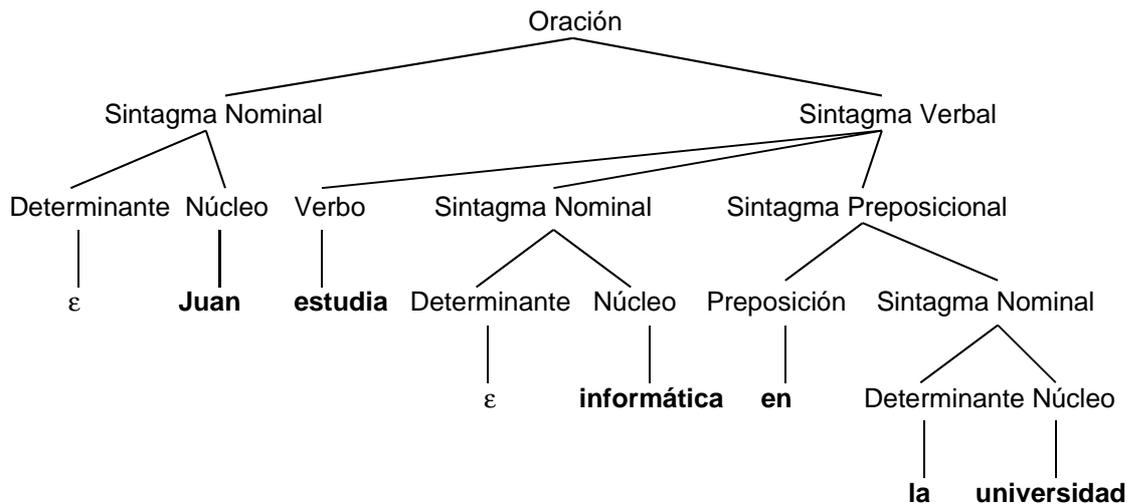


## Gramáticas

Análisis sintáctico de la frase "Juan estudia informática en la universidad"



<oración> → <sintagma nominal> <sintagma verbal>

<sintagma nominal> → <determinante> <núcleo>

<sintagma verbal> → <verbo> <sintagma nominal> <sintagma preposicional>

<sintagma preposicional> → <preposición> <sintagma nominal>

<determinante> → ε | la

<núcleo> → Juan | informática | universidad

<verbo> → estudia

<preposición> → en

**Definición:** Una Gramática es una cuádrupla  $G \equiv (V, \Sigma, S, P)$  donde

**V:** Es el conjunto de variables gramaticales o símbolos no terminales.  
V es finito,  $V \neq \emptyset$

**$\Sigma$ :** Es el conjunto de símbolos terminales o alfabeto de la gramática.  $\Sigma$  es finito,  $\Sigma \neq \emptyset$ ,  $\Sigma \cap V = \emptyset$ .

**S:**  $S \in V$ , es el símbolo de arranque o símbolo inicial de la gramática

**P:** Es el conjunto de reglas de producción.  
 $P \subseteq (V \cup \Sigma)^* \times (V \cup \Sigma)^*$        $P \neq \emptyset$

Los elementos de P son pares donde la primera componente es una variable sintáctica y la segunda una cadena de símbolos terminales (de  $\Sigma$ ) y no terminales (de V). En el primer elemento del par ha de haber al menos un símbolo no terminal (de V)

Estas componentes se suelen representar de la forma:

$\alpha \rightarrow \beta$  o bien  $\alpha ::= \beta$      $\alpha, \beta \in (V \cup \Sigma)^*$

Se suele especificar P listando todas sus producciones (reglas) colocando primero las del símbolo de arranque, S

## Convenios a la hora de escribir gramáticas

- Los símbolos **terminales** se representan mediante letras minúsculas del principio del alfabeto latino: **a, b, c...**
- Los símbolos **no terminales** se representan usando las primeras letras del alfabeto latino en mayúsculas: **A, B, C, D,...**
- Los **símbolos gramaticales** (terminales y no terminal) se representarán utilizando en mayúsculas las últimas letras del alfabeto latino: **X, Y, Z, W,...**
- Las **cadenas de símbolos terminales** se representarán usando en minúsculas las últimas letras del alfabeto latino: **w, x, y, z,...**
- Las **cadenas de símbolos gramaticales** (terminales y no terminales) se representarán usando letras del alfabeto griego:  **$\alpha, \beta, \gamma, \delta, \dots$**
- Si existen **diferentes alternativas** para un mismo símbolo no terminal, las diferentes alternativas se listan separadas por barras verticales:  $A \rightarrow \alpha|\beta|\dots|\gamma|\zeta$

## Clasificación de Chomsky

- **Gramáticas estructuradas por frases, semi-Thue o de tipo 0:**

Son las más generales. Sus reglas no tienen ninguna restricción y son de la forma:

$$\alpha \rightarrow \beta$$

Donde  $\alpha$  y  $\beta$  son cadenas arbitrarias de símbolos gramaticales con  $\alpha \neq \epsilon$ .

- **Gramáticas sensibles al contexto o de tipo 1:**

Sus reglas son de la forma:

$$\alpha \rightarrow \beta$$

con la restricción  $|\beta| \geq |\alpha|$ . Se puede demostrar que este tipo de reglas son equivalentes a las de la forma:

$$\alpha A \beta \rightarrow \alpha \gamma \beta$$

- **Gramáticas independientes del contexto (CFG) o de tipo 2:**

Sus reglas tienen la forma

$$A \rightarrow \alpha$$

La mayor parte de los lenguajes de programación están generados por una CFG

- **Gramáticas regulares o de tipo 3:**

Si todas las producciones de una gramática son de la forma:

$$\begin{aligned} A &\rightarrow wB \mid w && \text{o bien} \\ A &\rightarrow Bw \mid w \end{aligned}$$

Donde  $A$  y  $B$  son variables sintácticas y  $w$  una cadena arbitraria de terminales (podría ser vacía), se dice que la gramática es lineal por la derecha (izquierda). Una gramática lineal por la izquierda o por la derecha se denomina una gramática regular.