



INSTITUTO POLITÉCNICO NACIONAL  
ESCUELA SUPERIOR DE CÓMPUTO

COMPILADORES

La jerarquía de Chomsky

Grupo: 3CM2

*Integrantes:*

López Manríquez Ángel

*Profesor:*

Tecla Parra Roberto

Fecha de realización: 23 de junio de 2019

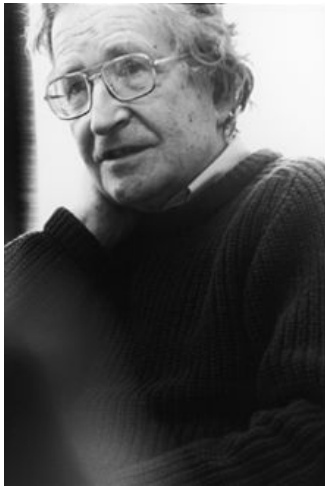
# Índice

1. La jerarquía	2
2. Lenguajes Recursivamente Enumerables (de tipo 0)	3
3. Lenguajes Dependientes del Contexto (sensibles al contexto, de tipo 1)	3
4. Lenguajes Independientes del Contexto (Libres de contexto, de tipo 2)	4
5. Lenguajes Regulares (de tipo 3)	4

# Jerarquia de Chomsky

López Manríquez Ángel  
3CM2

23 de junio de 2019



En lingüística la jerarquía de Chomsky (ocasionalmente también llamada la jerarquía de Chomsky–Schützenberger) es una clasificación jerárquica de distintos tipos de gramáticas formales que generan lenguajes formales. Esta jerarquía fue descrita por Noam Chomsky en 1956.

## 1. La jerarquia

La Jerarquía de Chomsky consta de cuatro niveles:

- Gramáticas de tipo 0 (sin restricciones), que incluye a todas las gramáticas formales. Estas gramáticas generan todos los lenguajes capaces de ser reconocidos por una máquina de Turing. Los lenguajes son conocidos como lenguajes recursivamente enumerables. Nótese que esta categoría es diferente de la de los lenguajes recursivos, cuya decisión puede ser realizada por una máquina de Turing que se detenga.
- Gramáticas de tipo 1 (gramáticas sensibles al contexto) generan los lenguajes sensibles al contexto. Estas gramáticas tienen reglas de la forma  $\alpha A \beta \rightarrow \alpha \gamma \beta$  con  $A$  un no terminal y  $\alpha$ ,  $\beta$  y  $\gamma$  cadenas de terminales y no terminales. Las cadenas  $\alpha$  y  $\beta$  pueden ser vacías, pero  $\gamma$  no puede serlo. La regla  $S \rightarrow \epsilon$  está permitida si  $S$  no aparece en la parte derecha de ninguna regla. Los lenguajes descritos por estas gramáticas son exactamente todos aquellos lenguajes reconocidos por una máquina de Turing determinista cuya cinta de memoria está acotada por un cierto número entero de veces sobre la longitud de entrada, también conocidas como autómatas linealmente acotados.
- Gramáticas de tipo 2 (gramáticas libres del contexto) generan los lenguajes independientes del contexto. Las reglas son de la forma  $A \rightarrow \gamma$  con  $A$  un no terminal y  $\gamma$  una cadena de terminales y no terminales. Estos lenguajes son aquellos que pueden ser reconocidos por un autómata con pila.

- Gramáticas de tipo 3 (gramáticas regulares) generan los lenguajes regulares. Estas gramáticas se restringen a aquellas reglas que tienen en la parte izquierda un no terminal, y en la parte derecha un solo terminal, posiblemente seguido de un no terminal. La regla  $S \rightarrow \epsilon$  también está permitida si  $S$  no aparece en la parte derecha de ninguna regla. Estos lenguajes son aquellos que pueden ser aceptados por un autómata finito. También esta familia de lenguajes pueden ser obtenidas por medio de expresiones regulares.

Nótese que el conjunto de gramáticas correspondiente a los lenguajes recursivos no es un miembro de la jerarquía.

Cada lenguaje regular es a su vez libre del contexto, asimismo un lenguaje libre del contexto es también dependiente del contexto, éste es recursivo y a su vez, recursivamente enumerable. Las inclusiones son, sin embargo, propias, es decir, existen en cada nivel lenguajes que no están en niveles anteriores.

Tipo	Lenguaje	Autómata	Normas de producción de gramáticas
0	recursivamente enumerable (LRE)	Máquina de Turing (MT)	Sin restricciones
1	dependiente del contexto (LSC)	Autómata linealmente acotado	$\alpha A \beta \rightarrow \alpha \gamma \beta$
2	independiente del contexto (LLC)	Autómata con pila	$A \rightarrow \gamma$
3	regular (LR)	Autómata finito	$A \rightarrow aB$ $A \rightarrow a$

## 2. Lenguajes Recursivamente Enumerables (de tipo 0)

Las gramáticas que generan estos lenguajes pueden tener reglas compresoras. Las reglas de producción son de la siguiente forma:

$$P = \{(u \rightarrow v) | u = xAy; u \in \Sigma^+; v, x, y \in \Sigma^*; A \in N\}$$

## 3. Lenguajes Dependientes del Contexto (sensibles al contexto, de tipo 1)

No existen reglas compresoras en toda la teoría, salvo, opcionalmente, la que deriva el axioma a la palabra vacía.

Existen reglas en las que un símbolo no terminal puede derivar a formas sentenciales distintas, según los símbolos que aparezcan a su alrededor.

Las reglas de producción son de la siguiente forma:

$$P = \{(S \rightarrow \lambda) \vee (xAy \rightarrow xvy) \mid v \in \Sigma^+; x, y \in \Sigma^*; A \in N\}$$

## 4. Lenguajes Independientes del Contexto (Libres de contexto, de tipo 2)

La mayoría de los lenguajes de programación entran en ésta categoría en cuanto su forma sintáctica, aunque en realidad los lenguajes de programación son dependientes del contexto, se reconocen a través de lenguajes de tipo 2 porque su reconocimiento es de  $O(n)$  mientras que los de tipo 1 tienen un orden de reconocimiento  $O(n^3)$  en el peor caso. Por este motivo se ejecuta un análisis semántico para reconocer si el programa es correcto.

Las reglas de producción son de la siguiente manera:

$$P = \{(S \rightarrow \lambda) \vee (A \rightarrow v) \mid v \in \Sigma^+; A \in N\}$$

## 5. Lenguajes Regulares (de tipo 3)

Son los lenguajes más simples dentro la Jerarquía de Chomsky. Se suelen expresar mediante expresiones regulares.

Existen 2 tipos: lineales por la derecha y lineales por la izquierda. Las reglas de producción son de la siguiente forma:

Lineales por la derecha:

$$P = \{(S \rightarrow \lambda) \vee (A \rightarrow aB) \vee (A \rightarrow a) \mid a \in T; A, B \in N\}$$

Lineales por la izquierda:

$$P = \{(S \rightarrow \lambda) \vee (A \rightarrow Ba) \vee (A \rightarrow a) \mid a \in T; A, B \in N\}$$