Instituto Politécnico Nacional
Escuela Superior de Cómputo

Compiladores

Profr. Edgardo Adrián Franco Martínez

14 Lenguajes y gramáticas II
Jerarquía de las gramáticas

efranco.docencia@gmail.com
http://computacion.cs.cinvestav.mx/~efranco

Contenido

- Lenguaje generado por una gramática (Derivaciones)
  - Ejemplo
- Jerarquía de Chomsky
  - Gramáticas tipo 3
  - Gramáticas tipo 2
  - Gramáticas tipo 1
  - Gramáticas tipo 0
- Descripción de las gramáticas
- Ejercicios 04

Lenguaje generado por una gramática

- **Definición:** Decimos que la cadena w_1 ***deriva en un paso*** a la cadena w_2 **($w_1 \Rightarrow_G w_2$)** si y solo si existen cadenas $x, y \in \mathbf{V}^*$ tales que $w_1 = x\, \boldsymbol{u}\, y$ y $w_2 = x\, \boldsymbol{v}\, y$ y además existe una regla $u \rightarrow v$ en **P**.
  - Se acostumbra omitir el subíndice que indica la gramática G.
- **Definición:** una cadena $w \in \mathbf{V}^*$ es ***derivable*** a partir de la gramática G si y solo si existe una secuencia de derivación iniciando en el símbolo inicial y terminando en la cadena *w*:
  - $S = w_1 \Rightarrow w_2 \Rightarrow w_3 \Rightarrow \ldots \Rightarrow w_n = w.$
  - Escribimos $\alpha \Rightarrow_G \beta$ si α deriva a β en 0 o más pasos.
- **Definición:** el ***lenguaje generado*** por una gramática *G*, *L*(*G*), es igual al conjunto de las palabras en **VT** derivables a partir de *G*.
  - Una gramática describe las reglas sintácticas del lenguaje. Si una palabra u oración no sigue las reglas, entonces no pertenecen al lenguaje generado por la gramática.

Ejemplo

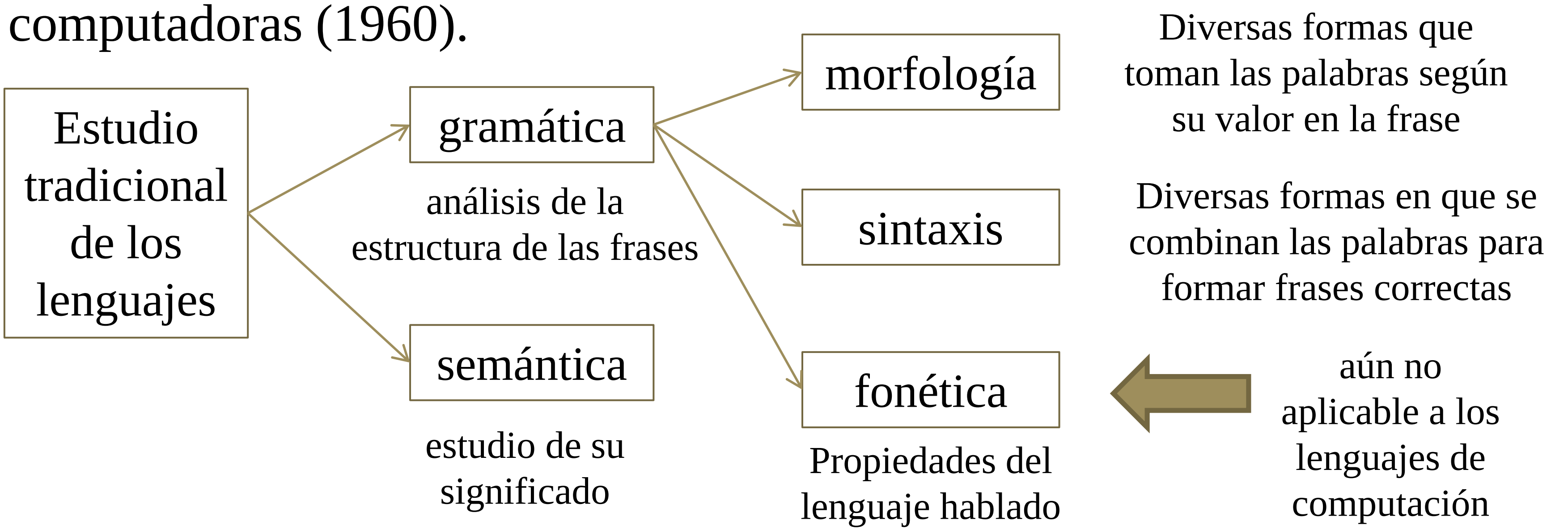
- **G = (VT , VN , S , P)**
  - **VN** = $\{S, A, B\}$
  - **VT** = $\{a, b, c\}$
  - **P**:
    - $S \rightarrow AccA$
    - $A \rightarrow BA \mid \lambda$
    - $B \rightarrow a \mid b \mid c$

Cadena	**Regla**	**Derivación**
S	$S \rightarrow AccA$	$S \Rightarrow AccA$
$AccA$	$A \rightarrow BA$	$\Rightarrow BAccA$
$BAccA$	$B \rightarrow a$	$\Rightarrow aAccA$
$aAccA$	$A \rightarrow BA$	$\Rightarrow aBAccA$
$aBAccA$	$B \rightarrow b$	$\Rightarrow abAccA$
$abAccA$	$A \rightarrow \lambda$	$\Rightarrow abccA$
$abccA$	$A \rightarrow \lambda$	$\Rightarrow abcc$

$w_1 = abcc \in L(G)$ y $w_2 = acb \notin L(G)$

Jerarquía de Chomsky

➢En 1950 el lingüista norteamericano Avram Noam Chomsky introdujo la *teoría de las gramáticas transformacionales* o *teoría de lenguajes formales*, que convirtió la lingüística en una ciencia y proporcionó una herramienta que no sólo podía aplicarse a los lenguajes naturales, sino que facilitaba el estudio y la **formalización de los lenguajes para la programación** de computadoras (1960).

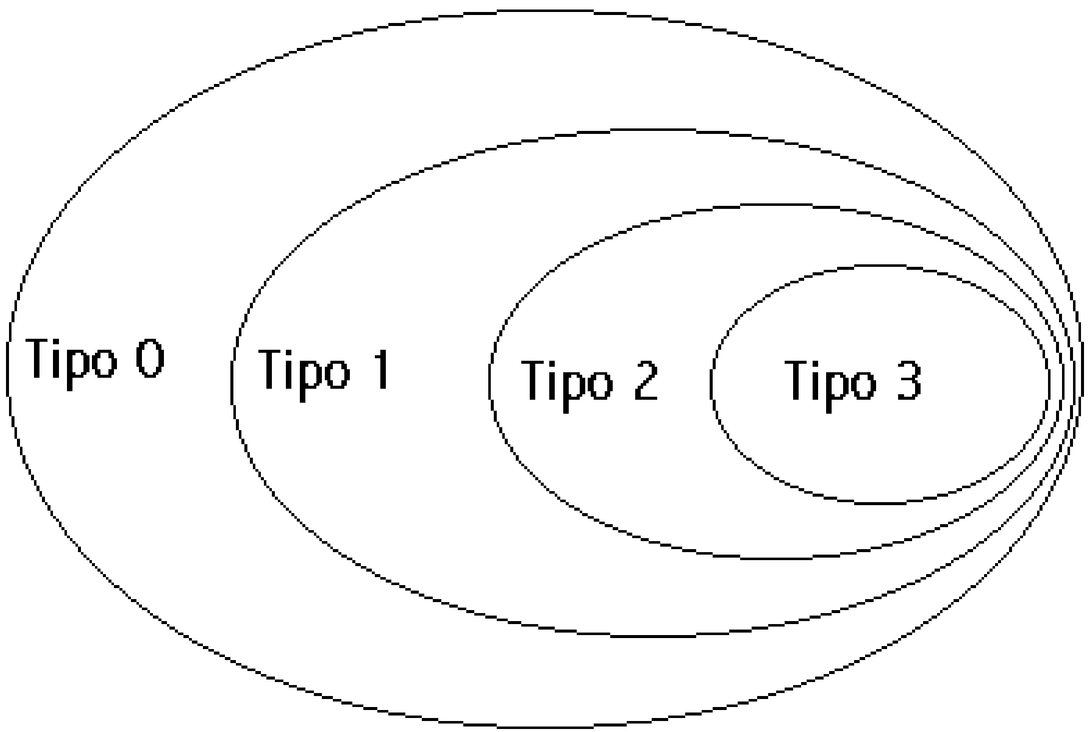


Jerarquía de Chomsky

- En función de la forma de sus producciones, se puede caracterizar qué tan compleja es una gramática formal.

- Noam Chomsky mostró que esta caracterización clasifica jerárquicamente a las gramáticas formales: Gramáticas en un nivel están incluidas en los siguientes niveles y la inclusión entre niveles es propia.

Jerarquía de Chomsky



Jerarquía de Chomsky

Gramáticas Tipo 3 *(gramáticas regulares)*

- Generan los **lenguajes regulares**. Las reglas (producciones) se restringen a un único no terminal en la parte izquierda y una parte derecha compuesta por un único terminal que puede estar seguido o no de un único no terminal. Es decir, normas del tipo:

$$A \rightarrow aB$$

$$A \rightarrow a$$

  - Estos lenguajes son los que pueden ser decididos por un **autómata finito *(regular)*.** Los lenguajes regulares se utilizan para definir **estructura léxica** de los lenguajes de programación. **Definen la sintaxis** de los **identificadores**, **números**, **cadenas** y otros **elemntos básicos** del lenguaje.

Jerarquía de Chomsky

Gramáticas Tipo 2*(independientes o libres de contexto)*

- Generan los lenguajes libres de contexto. Están definidas por reglas de la forma:
  - $A \rightarrow \gamma$
  - **A** es un no terminal
  - **γ** es una cadena de terminales y no terminales.
- Se denominan independientes de contexto porque A puede sustituirse por γ independientemente de las cadenas por las que esté acompañada.
- Estos lenguajes son todos los lenguajes que pueden ser reconocidos por los autómatas de pila.
  - Los lenguajes independientes de contexto **constituyen la base teórica para la sintaxis de la mayoría de los lenguajes de programación**. Definen la **sintaxis de las declaraciones**, las **proposiciones**, **las expresiones**, etc.*(i.e. la estructura de un programa).*

Jerarquía de Chomsky

Gramáticas Tipo 1 *(dependientes de contexto)*

- Generan los lenguajes dependientes de contexto. Contienen reglas de producción de la forma:

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

  A es un no terminal

  α, **β** y **γ** **son** cadenas de terminales y no terminales.

  α y **β** pueden ser vacíos, pero **γ** ha de ser distinto del vacío.

Jerarquía de Chomsky

Gramáticas Tipo 1 (Continuación)

- Se denominan gramáticas dependientes del contexto, porque, como se observa, A puede ser sustituido por γ si está acompañada de α por la izquierda y de β por la derecha.

- Estos lenguajes son todos los lenguajes que pueden ser reconocidos por autómatas lineales acotados.

Jerarquía de Chomsky

- **Gramáticas Tipo 0** *(sin restricciones, recursivas)*
- Incluyen todas las gramáticas formales.
- El más general, al que pertenece la semántica de los lenguajes naturales y artificiales.
- A estos lenguajes no se les impone restricción alguna.
- Estos lenguajes son todos los lenguajes que pueden ser reconocidos por una *máquina de Turing*.

Jerarquía de Chomsky
(Observaciones)

- Se dice que un lenguaje es de tipo **k** $[k = 0, k = 1, k = 2, k = 3]$ cuando existe una gramática de tipo k que genera ese lenguaje.

- La clasificación de la gramática será la correspondiente al tipo de la producción de menor clasificación.

Jerarquía de Chomsky
(Observaciones)

Gramática	Lenguaje	Reglas de Producción	Si $\mu \rightarrow \varphi$, relación entre $\|\mu\|$ y $\|\varphi\|$	Solución
Tipo-0	Recursivas	Sin restricciones		Máquinas de Turing
Tipo-1	Dependiente de contexto	$\alpha A\beta \rightarrow \alpha\gamma\beta$	$\|\mu\| \leq \|\varphi\|$	Autómatas lineales acotados
Tipo-2	Independiente de contexto	$A \rightarrow \gamma$	$\|\mu\| = 1$	Autómatas de pila
Tipo-3	Regular	$A \rightarrow aB$ $A \rightarrow a$	$\|\mu\| = 1$	Autómatas finitos, regulares

$$G_3 \subset G_2 \subset G_1 \subset G_0$$

Descripción de las gramáticas

Gramáticas Regulares (tipo 3 o G_3)

- **Gramáticas Regulares (tipo 3 o G_3)**
  - El lado **izquierdo consiste sólo de una variable**.
  - El lado derecho consiste de
    - **Un símbolo terminal seguido de una variable** ó
    - **Sólo un símbolo** terminal ó
    - La cadena vacía.

P.g.: $A \rightarrow aB \mid a \mid \lambda$

Descripción de las gramáticas

Gramáticas Regulares (tipo 3 o G_3)

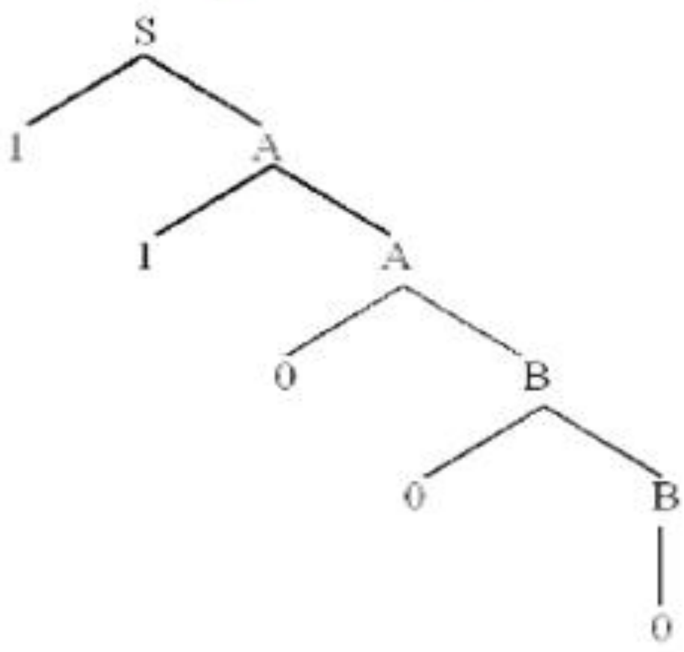
Ejemplo: Las dos gramáticas G_r y G_l generan el lenguaje regular 11^*00^*

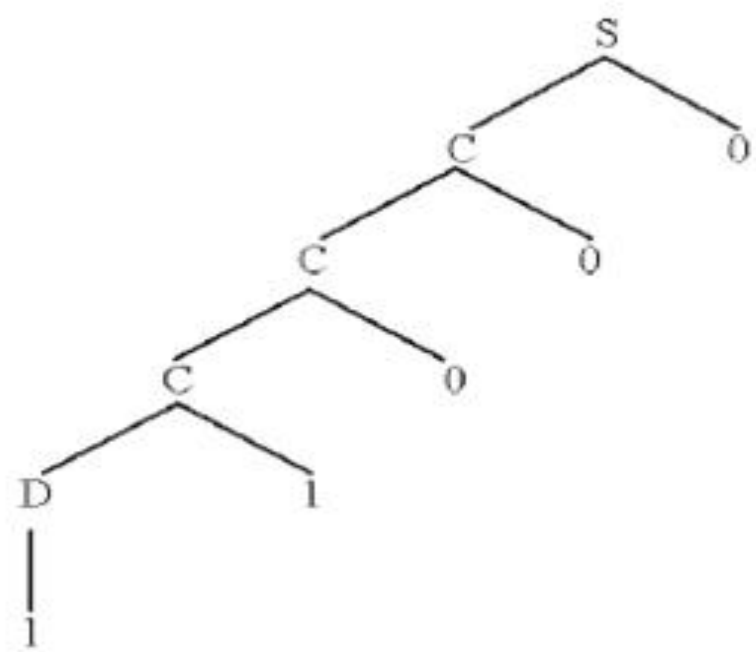
Regular por la derecha

$$G_r = \begin{cases} S \rightarrow 1A \\ A \rightarrow 1A \mid 0B \mid 0 \\ B \rightarrow 0B \mid 0 \end{cases}$$

Regular por la izquierda

$$G_l = \begin{cases} S \rightarrow C0 \\ C \rightarrow C0 \mid D1 \mid 1 \\ D \rightarrow D1 \mid 1 \end{cases}$$





Descripción de las gramáticas

Gramáticas Libres de Contexto, *GLC*, (tipo 2 o G_2)

- **Gramáticas Libres de Contexto, *GLC*, (tipo 2 o G_2)**
  - El lado **izquierdo** consiste **sólo de una variable.**
  - No hay restricciones para el lado derecho.

P.g.: $S \rightarrow aSb \mid ab \mid \lambda$

Descripción de las gramáticas

Gramáticas Libres de Contexto, GLC, (tipo 2 o G2)

Ejemplo: Las dos gramáticas G_1 y G_2 generan el lenguaje independiente del contexto $0^n 1^n 2^m$ con $n, m \geq 0$.

Lenguaje

GIC en formato no estricto	GIC en formato estricto
$G_1 = \begin{cases} S \to AB \\ A \to 0A1 \mid \lambda \\ B \to 2B \mid \lambda \end{cases}$	$G_2 = \begin{cases} S \to AB \mid A \mid B \mid \lambda \\ A \to 0A1 \mid 01 \\ B \to 2B \mid 2 \end{cases}$

Descripción de las gramáticas

Gramáticas Sensitivas al Contexto (tipo 1 o G_1)

- **Gramáticas Sensitivas al Contexto (tipo 1 o G_1)** A es un símbolo no terminal. Además**, las reglas son no-contractivas**, i.e. la longitud del lado izquierdo es menor o igual a la longitud del lado derecho. Esta propiedad de no-contracción garantiza que un lenguaje sensitivo al contexto no contiene λ.

P.g.:

$$S \rightarrow abc \mid aAbc \qquad Ab \rightarrow bA \qquad Ac \rightarrow Bbcc$$

$$bB \rightarrow Bb \qquad aB \rightarrow aa \mid aaA$$

Descripción de las gramáticas

Gramáticas Sensitivas al Contexto (tipo 1 o G_1)

Ejemplo: Sea la GDC $G = (\{a, b, c\}, \{S, M\}, S, P)$ donde

$$P = \begin{cases} S \to aMc \mid aSMc \\ aM \to ab \\ bM \to bb \\ cM \to Mc \end{cases}$$

La gramática G genera el lenguaje dependiente del contexto $a^n b^n c^n$ con $n > 0$. Un ejemplo de derivación sería:

$$S \Rightarrow aSMc \Rightarrow aaMcMc \Rightarrow aabcMc \Rightarrow aabMcc \Rightarrow aabbcc$$

Descripción de las gramáticas

Gramáticas sin restricción (tipo 0 o G_0)

- **Gramáticas sin restricción (tipo 0 o G_0)**, no hay restricciones para las reglas, excepto que el lado izquierdo no es λ.

P.g.:

$$
\begin{array}{lll}
S \rightarrow aSBC \mid aBC & CB \rightarrow BC & aB \rightarrow ab \\
bB \rightarrow bb & bC \rightarrow bc & cC \rightarrow cc \\
A \rightarrow bc & &
\end{array}
$$

Descripción de las gramáticas

Gramáticas sin restricción (tipo 0 o G_0)

Las gramáticas GEFs se caracterizan frente al resto (dejando aparte el caso $S \rightarrow \lambda$) en que admite reglas **compresoras**. Una regla compresora es aquella regla que cumple que el tamaño de su lado derecho es menor que el tamaño de su lado izquierdo.

Ejemplo: Sea la GEF $G = (\{a, b, c\}, \{S, M\}, S, P)$ donde

$$P = \begin{cases} S \rightarrow abMSc \mid \lambda \\ bMa \rightarrow abM \\ bMc \rightarrow bc \\ bMb \rightarrow bbM \end{cases}$$

La gramática G genera el lenguaje $a^n b^n c^n$ con $n \geq 0$. Un ejemplo de derivación sería:

$$S \Rightarrow abMSc \Rightarrow abMabMScc \Rightarrow abMabMcc \Rightarrow abMabcc \Rightarrow aabMbcc \Rightarrow aabbMcc \Rightarrow aabbcc$$

Ejercicios 04

- Clasificar las siguientes gramáticas dadas sus reglas de producción.

Ga	Gb	Gc	Gd	Ge
$Z \rightarrow yX$ $X \rightarrow y$ $X \rightarrow \lambda$ $yX \rightarrow x$	$yW \rightarrow x$ $X \rightarrow xZy$ $YX \rightarrow WvZ$	$E \rightarrow E+T$ $E \rightarrow E-T$ $E \rightarrow T$ $T \rightarrow T*F$ $T \rightarrow T/F$ $T \rightarrow F$ $F \rightarrow (E)$ $F \rightarrow id$	$S \rightarrow aAbc$ $Ab \rightarrow bA$ $Ac \rightarrow Bbcc$ $bB \rightarrow Bba$ $B \rightarrow aa$ $B \rightarrow aaA$	$S \rightarrow aS$ $S \rightarrow aN$ $N \rightarrow bN$ $N \rightarrow bM$ $N \rightarrow b$ $M \rightarrow c$

Ejercicios 04

- **Fecha de entrega**

Entregar en formato digital vía Web, con el titulo *"Ejercicios 04 Clasificación de gramáticas"* a más tardar el día lunes 04 de abril de 2011.