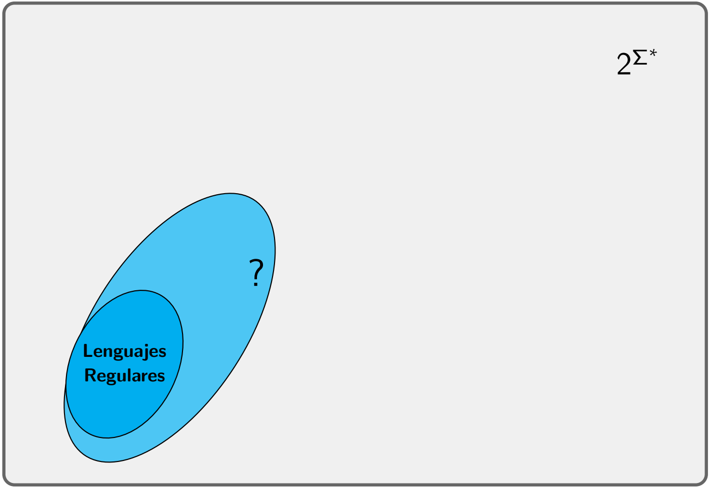
Gramáticas libres de contexto

Clase 15

IIC 2223

Prof. Cristian Riveros

¿dónde estamos?



¿qué le falta a los lenguajes regulares?

Outline

Definición de grámaticas

Árboles y derivaciones

Outline

Gramáticas libres de contexto

Definición

Una **gramática libre de contexto** (CFG) es una tupla:

$$\mathcal{G} = (V, \Sigma, P, S)$$

- V es un conjunto finito de **variables** o **no-terminales**.
- Σ es un alfabeto finito (o **terminales**) tal que $\Sigma \cap V = \varnothing$.
- $P \subseteq V \times (V \cup \Sigma)^*$ es un subconjunto finito de **reglas** o **producciones**.
- $S \in V$ es la **variable inicial**.

Gramáticas libres de contexto

Ejemplo

$$\mathcal{G}: \begin{array}{lcl} \mathsf{A} & \rightarrow & \mathsf{0\ A\ 1} \\ \mathsf{A} & \rightarrow & \mathsf{B} \\ \mathsf{B} & \rightarrow & \# \end{array}$$

La grámatica se define como $\mathcal{G} = (V, \Sigma, P, S)$ tal que:

- $V = \{ A, B \}$
- $\Sigma = \{ 0, 1, \# \}$
- $P = \{ (A, 0A1),\ (A, B),\ (B, \#) \}$
- $S = A$

Notación para gramáticas libres de contexto

Notación

- Para las **variables** en una gramática usaremos letras mayúsculas:

$$A, B, C, \ldots$$

- Para los **terminales** en una gramática usaremos letras minúsculas:

$$a, b, c, \ldots$$

- Para palabras en $(V \cup \Sigma)^*$ usaremos símbolos:

$$\alpha, \beta, \gamma, \ldots$$

- Para una producción $(A, \alpha) \in P$ la escribimos como:

$$A \to \alpha$$

Notación para gramáticas libres de contexto

Ejemplo anterior

$$\mathcal{G}: \begin{array}{ccc} \mathsf{A} & \to & \mathsf{0\ A\ 1} \\ \mathsf{A} & \to & \mathsf{B} \\ \mathsf{B} & \to & \# \end{array}$$

Esta grámatica se formaliza como $\mathcal{G} = (V, \Sigma, P, S)$ donde:

- $V = \{ A, B \}$ → variables en **mayus.**
- $\Sigma = \{ 0, 1, \# \}$ → letras en **minus.**
- $P = \{ A \to 0A1,\ A \to B,\ B \to \# \}$ → **producciones**
- $S = A$

Simplificación para gramáticas libres de contexto

Simplificación

Si tenemos un conjunto de reglas de la forma:

$$\begin{array}{ccc} A & \rightarrow & \alpha_1 \\ A & \rightarrow & \alpha_2 \\ & \cdots & \\ A & \rightarrow & \alpha_n \end{array}$$

entonces escribimos estas reglas **sucintamente** como:

$$A \quad \rightarrow \quad \alpha_1 \mid \alpha_2 \mid \cdots \mid \alpha_n$$

(recordar que: $\alpha_1, \alpha_2, \ldots, \alpha_n \in (\Sigma \cup V)^*$)

Simplificación para gramáticas libres de contexto

Ejemplo anterior

$$
\begin{array}{llll}
\mathcal{G}: & \mathsf{A} & \rightarrow & \mathsf{0\ A\ 1} \\
 & \mathsf{A} & \rightarrow & \mathsf{B} \\
 & \mathsf{B} & \rightarrow & \#
\end{array}
$$

Esta grámatica la escribiremos en notación **sucinta** como:

$$
\begin{array}{llll}
\mathcal{G}: & \mathsf{A} & \rightarrow & \mathsf{0\ A\ 1} \mid \mathsf{B} \\
 & \mathsf{B} & \rightarrow & \#
\end{array}
$$

Producciones

Sea $\mathcal{G} = (V, \Sigma, P, S)$ una CFG.

Definición

Definimos la relación $\Rightarrow \subseteq (V \cup \Sigma)^* \times (V \cup \Sigma)^*$ de **producción** tal que:

$$\alpha \cdot A \cdot \beta \Rightarrow \alpha \cdot \gamma \cdot \beta \qquad \text{si, y solo si,} \qquad (A \to \gamma) \in P$$

para todo $A \in V$ y $\alpha, \beta, \gamma \in (V \cup \Sigma)^*$.

Si $\alpha A \beta \Rightarrow \alpha \gamma \beta$ entonces decimos que

- $\alpha A \beta$ **produce** $\alpha \gamma \beta$ o
- $\alpha \gamma \beta$ **es producible** desde $\alpha A \beta$.

$\alpha A \beta \Rightarrow \alpha \gamma \beta$ es **reemplazar** γ en A en la palabra $\alpha A \beta$.

Producciones

¿cuál de las siguientes producciones son correctas?

$$\mathcal{G}: \begin{array}{lcl} \mathsf{A} & \rightarrow & \mathsf{0\ A\ 1} \mid \mathsf{B} \\ \mathsf{B} & \rightarrow & \# \end{array}$$

- $A \Rightarrow B$?
- $00A11 \Rightarrow 000A111$?
- $000B111 \Rightarrow 000A111$?
- $0A0A1BA \Rightarrow 0A0A1B0A1$?

Derivaciones

Sea $\mathcal{G} = (V, \Sigma, P, S)$ una CFG.

Definición

Dada dos palabras $\alpha, \beta \in (V \cup \Sigma)^*$ decimos que α **deriva** β:

$$\alpha \overset{\star}{\Rightarrow} \beta$$

Si existe $\alpha_1, \alpha_2, \ldots, \alpha_n \in (V \cup \Sigma)^*$ tal que:

$$\alpha \Rightarrow \alpha_1 \Rightarrow \alpha_2 \Rightarrow \ldots \Rightarrow \beta$$

Derivaciones

Sea $\mathcal{G} = (V, \Sigma, P, S)$ una CFG.

Definición

Dada dos palabras $\alpha, \beta \in (V \cup \Sigma)^*$ decimos que α **deriva** β:

$$\alpha \overset{\star}{\Rightarrow} \beta$$

con $\overset{\star}{\Rightarrow}$ es la **clausura refleja y transitiva** de $\Rightarrow$, esto es:

1. $\alpha \overset{\star}{\Rightarrow} \alpha$
2. $\alpha \overset{\star}{\Rightarrow} \beta$ si, y solo si, existe γ tal que $\alpha \overset{\star}{\Rightarrow} \gamma$ y $\gamma \Rightarrow \beta$

para todo $\alpha, \beta \in (V \cup \Sigma)^*$.

Notar que $\Rightarrow$ y $\overset{\star}{\Rightarrow}$ son relaciones entre palabras en $(V \cup \Sigma)^*$

Derivaciones

¿cuál de las siguientes derivaciones son correctas?

$$\mathcal{G}: \quad \begin{array}{lcl} \mathsf{A} & \rightarrow & \mathsf{0\ A\ 1} \mid \mathsf{B} \\ \mathsf{B} & \rightarrow & \# \end{array}$$

- $A \stackrel{\star}{\Rightarrow} 000A111$?
- $00A11 \stackrel{\star}{\Rightarrow} 000B11$?
- $00A11 \stackrel{\star}{\Rightarrow} 000\#111$?

Lenguaje definido por una gramática

Sea $\mathcal{G} = (V, \Sigma, P, S)$ una CFG.

Definición

El **lenguaje** de una grámatica $\mathcal{G}$ se define como:

$$\mathcal{L}(\mathcal{G}) = \left\{ w \in \Sigma^* \mid S \overset{\star}{\Rightarrow} w \right\}$$

$\mathcal{L}(\mathcal{G})$ son todas las palabras en Σ^* que se pueden derivar desde S.

Lenguaje definido por una gramática

¿qué palabras están en $\mathcal{L}(\mathcal{G})$?

$$\mathcal{G}: \quad \begin{array}{lcl} \mathrm{A} & \to & 0\ \mathrm{A}\ 1 \mid \mathrm{B} \\ \mathrm{B} & \to & \# \end{array}$$

- Como $A \overset{\star}{\Rightarrow} 000\#111$, entonces $000\#111 \in \mathcal{L}(\mathcal{G})$.
- En general, uno puede demostrar **por inducción** que:

$$\mathcal{L}(\mathcal{G}) = \left\{0^n\#1^n \mid n \geq 0\right\}$$

Lenguaje definido por una gramática

¿qué lenguaje define cada grámatica libre de contexto?

1. $$\mathcal{G}: \quad \begin{array}{rcl} \mathsf{S} & \rightarrow & \mathsf{A\ S} \mid \epsilon \\ \mathsf{A} & \rightarrow & \mathsf{aa} \mid \mathsf{ab} \mid \mathsf{ba} \mid \mathsf{bb} \end{array}$$

2. $$\mathcal{G}: \quad \begin{array}{rcl} \mathsf{S} & \rightarrow & \mathsf{S + S} \mid \mathsf{S \times S} \mid \mathsf{(S)} \mid \mathsf{A} \\ \mathsf{A} & \rightarrow & \mathsf{0} \mid \mathsf{1} \mid \ldots \mid \mathsf{9} \end{array}$$

3. $$\mathcal{G}: \quad \mathsf{S} \rightarrow \mathsf{a\ S\ b} \mid \mathsf{S\ S} \mid \epsilon$$

Lenguaje definido por una gramática

¿cuál es una gramática para cada lenguaje?

1. $L_1 = \left\{ 0^n 1^n \mid n \geq 0 \right\} \cup \left\{ 1^n 0^n \mid n \geq 0 \right\}$

2. $L_2 = \left\{ w \in \{a, b\}^* \mid w = w^{\text{rev}} \right\}$

Lenguajes libres de contexto

Definición

Diremos que $L \subseteq \Sigma^*$ es un **lenguaje libre de contexto** ssi existe una gramática libre de contexto $\mathcal{G}$ tal que:

$$L = \mathcal{L}(\mathcal{G})$$

Ejemplos

Los siguientes son lenguajes libres de contexto:

- $L = \{0^n \# 1^n \mid n \geq 0\}$
- $\text{Par} = \{\, w \in \{a, b\}^* \mid w \text{ tiene largo par} \,\}$
- $\text{Pal} = \{\, w \in \{a, b\}^* \mid w = w^{\text{rev}} \,\}$

Outline

Árboles ordenados y etiquetados

Definiciones

El conjunto de **árboles ordenados y etiquetados** (o solo **árboles**) sobre etiquetas Σ y V, se define recursivamente como:

- $t := a$ es un árbol para todo $a \in \Sigma$.
- si $t_1, \ldots, t_k$ son árboles,
  entonces $t := A(t_1, \ldots, t_k)$ es un árbol para todo $A \in V$.

Para un árbol $t = A(t_1, \ldots, t_k)$ cualquiera se define:

- $\text{raiz}(t) = A$
- $\text{hijos}(t) = t_1, \ldots, t_k$

Si $t = a$, entonces decimos que t es una **hoja**, $\text{raiz}(t) = a$ y $\text{hijos}(t) = \epsilon$.

Árboles de derivación de una gramática

Fije una gramática libre de contexto $\mathcal{G} = (V, \Sigma, P, S)$.

Definiciones

Se define el conjunto de **árboles de derivación** recursivamente como:

- Si $a \in \Sigma$, entonces $t = a$ es un árbol de derivación.
- Si $A \to A_1 \dots A_k \in P$ y
  $t_1, \dots, t_k$ son árboles de derivación con $\text{raiz}(t_i) = A_i$ para todo $i \leq k$
  entonces $t = A(t_1, \dots, t_k)$ es un árbol de derivación.

Decimos que t es un **árbol de derivación de $\mathcal{G}$** si:

1. t es un árbol de derivación y
2. $\text{raiz}(t) = S$.

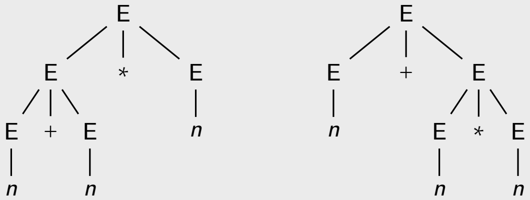
Los árboles de derivación son todos los árboles que parten desde S.

Árboles de derivación de una gramática

Ejemplo de árbol de derivación

$$\mathcal{G}: \quad E \;\rightarrow\; E + E \mid E * E \mid n$$

Algunos **árboles de derivación** para $\mathcal{G}$:



Árbol de derivación para una palabra

Sea $\mathcal{G} = (V, \Sigma, P, S)$ una CFG y $w \in \Sigma^*$.

Definiciones

Se define la función **yield** sobre árboles, recursivamente como:

- Si $t = a \in \Sigma$, entonces $\text{yield}(t) = a$.
- Si t no es una hoja y $\text{hijos}(t) = t_1 t_2 \ldots t_k$, entonces:

$$\text{yield}(t) \ = \ \text{yield}(t_1) \cdot \text{yield}(t_2) \cdot \ldots \cdot \text{yield}(t_k)$$

Decimos que t es un **árbol de derivación de $\mathcal{G}$ para w** si:

1. t es un árbol de derivación de $\mathcal{G}$ y
2. $\text{yield}(t) = w$.

Las hojas de t forman la palabra w.

Equivalencia entre árboles de derivación y derivaciones

Sea $\mathcal{G} = (V, \Sigma, P, S)$ una CFG y $w \in \Sigma^*$.

Proposición

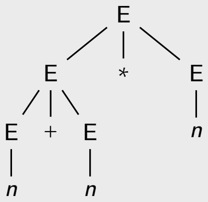
$w \in \mathcal{L}(\mathcal{G})$ si, y solo si, existe un árbol de derivación de $\mathcal{G}$ para w.

Un arbol de derivación es la **representación gráfica** de una derivación.

Equivalencia entre árboles de derivación y derivaciones

Ejemplo

$$E \;\rightarrow\; E + E \mid E * E \mid n$$



1) $E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow n + E * E \Rightarrow n + n * E \Rightarrow n + n * n$
2) $E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow E + n * E \Rightarrow E + n * n \Rightarrow n + n * n$
3) $E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow E + E * n \Rightarrow E + n * n \Rightarrow n + n * n$
4) $E \Rightarrow E * E \Rightarrow E * n \Rightarrow E + E * n \Rightarrow n + E * n \Rightarrow n + n * n$
5) $E \Rightarrow E * E \Rightarrow E * n \Rightarrow E + E * n \Rightarrow E + n * n \Rightarrow n + n * n$
6) ...

Dado un árbol de derivación, ¿con cuál derivación nos quedamos?

Derivaciones por la izquierda y por la derecha

Sea $\mathcal{G} = (V, \Sigma, P, S)$ una CFG.

Definición

- Definimos la **derivación por la izquierda** $\underset{\text{lm}}{\Rightarrow} \subseteq (V \cup \Sigma)^* \times (V \cup \Sigma)^*$:

$$\mathrm{w} \cdot A \cdot \beta \underset{\text{lm}}{\Rightarrow} \mathrm{w} \cdot \gamma \cdot \beta \qquad \text{si, y solo si,} \qquad A \to \gamma \in P$$

  para todo $A \in V$, $\mathrm{w} \in \Sigma^*$ y $\beta, \gamma \in (V \cup \Sigma)^*$.

- Definimos la **derivación por la derecha** $\underset{\text{rm}}{\Rightarrow} \subseteq (V \cup \Sigma)^* \times (V \cup \Sigma)^*$:

$$\alpha \cdot A \cdot \mathrm{w} \underset{\text{rm}}{\Rightarrow} \alpha \cdot \gamma \cdot \mathrm{w} \qquad \text{si, y solo si,} \qquad A \to \gamma \in P$$

  para todo $A \in V$, $\mathrm{w} \in \Sigma^*$ y $\alpha, \gamma \in (V \cup \Sigma)^*$.

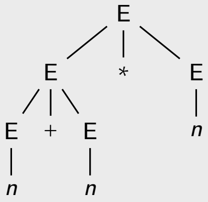
Se define $\overset{\star}{\underset{\text{lm}}{\Rightarrow}}$ y $\overset{\star}{\underset{\text{rm}}{\Rightarrow}}$ como la **clausura refleja y transitiva** de $\underset{\text{lm}}{\Rightarrow}$ y $\underset{\text{rm}}{\Rightarrow}$, resp.

$\underset{\text{lm}}{\Rightarrow}$ y $\underset{\text{rm}}{\Rightarrow}$ solo reemplaza **a la izquierda** (leftmost) y **derecha** (rightmost).

Derivaciones por la izquierda y por la derecha

Ejemplo anterior

$$E \;\rightarrow\; E+E \mid E*E \mid n$$



Derivación por la izquierda (lm)

$$E \underset{\text{lm}}{\Rightarrow} E*E \underset{\text{lm}}{\Rightarrow} E+E*E \underset{\text{lm}}{\Rightarrow} n+E*E \underset{\text{lm}}{\Rightarrow} n+n*E \underset{\text{lm}}{\Rightarrow} n+n*n$$

Derivación por la derecha (rm)

$$E \underset{\text{rm}}{\Rightarrow} E*E \underset{\text{rm}}{\Rightarrow} E*n \underset{\text{rm}}{\Rightarrow} E+E*n \underset{\text{rm}}{\Rightarrow} E+n*n \underset{\text{rm}}{\Rightarrow} n+n*n$$

¿cuál es la relación entre el **tipo de derivación** y el **recorrido del árbol**?

Derivaciones por la izquierda y por la derecha

Sabemos que ...

- Por cada derivación, existe un único árbol de derivación.
- Por cada árbol de derivación existen **múltiples** posibles derivaciones.

Proposición

Por cada árbol de derivación, existe
una **único** derivación por la izquierda y una **única** derivación por la derecha.

Por lo tanto, desde ahora podemos hablar de
árbol de derivación y derivación (izquierda o derecha) indistintamente.