Árboles y variables inútiles

Clase 20

IIC2223 / IIC2224

Prof. Cristian Riveros

Outline

Lenguajes libres de contexto (clase ant.)

Árboles y derivaciones

Variables inútiles

Outline

Lenguajes libres de contexto (clase ant.)

Árboles y derivaciones

Variables inútiles

Recordatorio: 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**.

Recordatorio: Gramáticas libres de contexto

Ejemplo

Consideré la grámatica $\mathcal{G} = (V, \Sigma, P, S)$ tal que:

- $V = \{ X, Y \}$
- $\Sigma = \{ a, b \}$
- $P = \{ (X, aXb),\ (X, Y),\ (Y, \epsilon) \}$
- $S = X$

$$\begin{array}{rlcl} \mathcal{G}: & X & \to & aXb \\ & X & \to & Y \\ & Y & \to & \epsilon \end{array}$$

Recordatorio: 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:

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

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

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

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

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

Recordatorio: 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:

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

Recordatorio: 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 $(V \cup \Sigma)^*$

Recordatorio: 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.

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:

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

Outline

Lenguajes libres de contexto (clase ant.)

Árboles y derivaciones

Variables inútiles

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 $t := X(t_1, \ldots, t_k)$ es un árbol para todo $X \in V$.

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

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

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 $X \to X_1 \dots X_k \in P$ y
  $t_1, \dots, t_k$ son árboles de derivación con $\text{raiz}(t_i) = X_i$ para todo $i \leq k$
  entonces $t = X(t_1, \dots, t_k)$ 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. $\text{raiz}(t) = S$.

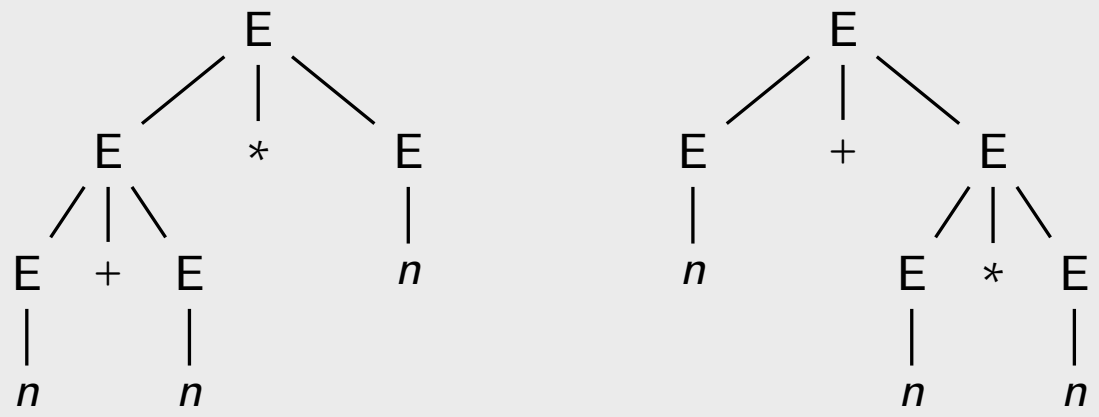
Los árboles de derivación son todos los árboles que parten desde S.

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 $\text{hijos}(t) = t_1 t_2 \dots t_k$, 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

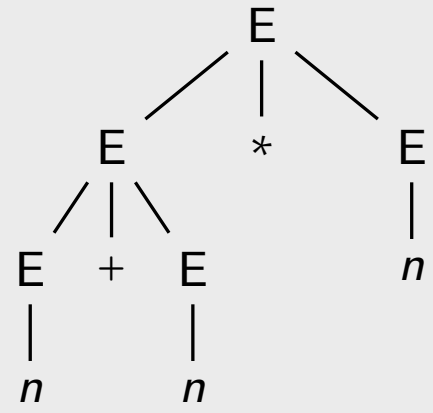
$w \in \mathcal{L}(\mathcal{G})$ si, y solo si, existe un árbol de derivación de $\mathcal{G}$ para w.

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** $\underset{\text{lm}}{\Rightarrow} \subseteq (V \cup \Sigma)^* \times (V \cup \Sigma)^*$:

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

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

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

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

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

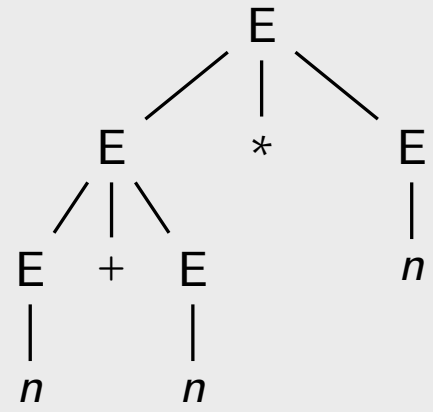
Se define $\overset{\star}{\underset{\text{lm}}{\Rightarrow}}$ y $\overset{\star}{\underset{\text{rm}}{\Rightarrow}}$ como la **clausura refleja y transitiva** de $\underset{\text{lm}}{\Rightarrow}$ y $\underset{\text{rm}}{\Rightarrow}$, resp.

$\underset{\text{lm}}{\Rightarrow}$ 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 **única** 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.

Outline

Lenguajes libres de contexto (clase ant.)

Árboles y derivaciones

Variables inútiles

¿cómo podemos simplificar está gramática?

$$
\begin{array}{lcl}
S & \rightarrow & aAa \mid aBD \\
A & \rightarrow & B \mid D \\
B & \rightarrow & aBa \mid b \\
C & \rightarrow & aCC \mid bC \\
D & \rightarrow & aDCa \mid CFa \\
F & \rightarrow & aFDa \mid aab \\
H & \rightarrow & ab
\end{array}
$$

¿cuáles son variables **inútiles**?

Variables inútiles

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

Definición

Diremos que una variable $X \in V$ es **útil** si existe una derivación:

$$S \overset{\star}{\Rightarrow} \alpha X \beta \overset{\star}{\Rightarrow} w$$

Al contrario, diremos que una variable X es **inútil** si NO es útil.

¿qué variables son inútiles?

$$\begin{aligned} S &\to aAa \mid aBC \\ A &\to aS \mid bD \\ B &\to aBa \mid b \\ C &\to abb \mid DD \\ D &\to aDa \end{aligned}$$

Variables inútiles

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

Definición

Diremos que una variable $X \in V$ es **útil** si existe una derivación:

$$S \overset{\star}{\Rightarrow} \alpha X \beta \overset{\star}{\Rightarrow} w$$

Al contrario, diremos que una variable X es **inútil** si NO es útil.

¿cómo podemos determinar si un símbolo es **inútil**?

Desde el lado positivo: símbolos útiles

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

Definición

Para una variable $X \in V$:

1. Decimos que X es **alcanzable** si existe una derivación:

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

2. Decimos que X es **generadora** si existe una derivación:

$$X \overset{\star}{\Rightarrow} w$$

¿cómo determinamos si una variable es **alcanzable** o **generadora**?

¿cómo determinamos si una variable es **alcanzable**?

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

Propiedad

Para toda variable $X \in V - \{S\}$,

existe una producción $Y \to \alpha X \beta$ en P tal que $Y \in V$ es alcanzable

si, y solo si, X es **alcanzable**.

Demostración: ejercicio.

¿cómo determinamos si una variable es **alcanzable**?

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

output: Conjunto C de variables alcanzables

Function `alcanzables` $(\mathcal{G})$

- **let** $C_0 := \{S\}$
- **let** $C := \varnothing$
- **while** $C_0 \neq \varnothing$ **do**
  - **take** $Y \in C_0$
  - $C_0 := C_0 - \{Y\}$
  - $C := C \cup \{Y\}$
  - **foreach** $X \in V - C$ *tal que existe una regla* $\left(Y \to \alpha X \beta\right) \in P$ **do**
    - $C_0 := C_0 \cup \{X\}$
- **return** C

¿cómo determinamos si una variable es **alcanzable**?

¿cuáles son las variables alcanzables?

$$
\begin{array}{lcl}
S & \rightarrow & aAa \\
A & \rightarrow & aB \\
B & \rightarrow & aBa \mid b \\
C & \rightarrow & abb
\end{array}
$$

¿cómo determinamos si una variable es **generadora**?

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

Propiedad

Para toda variable $X \in V$:

existe una regla $X \to \alpha$ tal que
todas las variables en α son generadoras

si, y solo si, X es **generadora**.

Demostración: ejercicio.

¿cómo determinamos si una variable es **generadora**?

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

output: Conjunto G de variables generadoras

Function `Generadores` $(\mathcal{G})$

- **let** $G_0 := \{X \in V \mid (X \to w) \in P\}$
- **let** $G := \varnothing$
- **while** $G_0 \neq G$ **do**
  - $G := G_0$
  - **foreach** $(X \to \alpha) \in P$ **do**
    - **if** *todas las variables en* α *estan en* G **then**
      - $G_0 := G_0 \cup \{X\}$
- **return** G

¿cómo determinamos si una variable es **generadora**?

¿cuáles son las variables generadoras?

$$
\begin{array}{lcl}
S & \rightarrow & aAa \mid aBD \\
A & \rightarrow & aB \mid bD \\
B & \rightarrow & aBa \mid b \\
C & \rightarrow & abb \mid DD \\
D & \rightarrow & aDCa
\end{array}
$$

Eliminación de variables inútiles

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

Teorema

Sea $\mathcal{G}''$ una gramática creada a partir de $\mathcal{G}$ después de:

- eliminar todos la variables y reglas **NO** **generadoras**.
- eliminar todas las variables y reglas **NO** **alcanzables**.

Entonces, $\mathcal{L}(\mathcal{G}'') = \mathcal{L}(\mathcal{G})$ y $\mathcal{G}''$ no contiene variables ínutiles.

¿qué falla al eliminar primero las no alcanzables y después las no generadoras?

Ejemplo

$$\mathcal{G}: \quad \begin{array}{lcl} S & \to & AB \mid b \\ A & \to & a \\ B & \to & B \end{array}$$

Al eliminar variables **no alcanzables** en $\mathcal{G}$:

$$\mathcal{G}': \quad \begin{array}{lcl} S & \to & AB \mid b \\ A & \to & a \\ B & \to & B \end{array}$$

Al eliminar variables **no generadoras** en $\mathcal{G}'$:

$$\mathcal{G}'': \quad \begin{array}{lcl} S & \to & b \\ A & \to & a \end{array}$$



Eliminación de variables inútiles

Demostración

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

Sea $\mathcal{G}' = (V', \Sigma, P', S)$ al eliminar las variables **no generadoras** de $\mathcal{G}$:

$$\begin{aligned} V' &= \{ X \in V \mid \exists w.\ X \overset{\star}{\underset{\mathcal{G}}{\Rightarrow}} w \} \\ P' &= \{ X \to \alpha \in P \mid X \in V' \wedge \alpha \in (V' \cup \Sigma)^* \} \end{aligned}$$

Sea $\mathcal{G}'' = (V'', \Sigma, P'', S)$ al eliminar las variables **no alcanzables** de $\mathcal{G}'$:

$$\begin{aligned} V'' &= \{ X \in V' \mid \exists \alpha, \beta.\ S \overset{\star}{\underset{\mathcal{G}'}{\Rightarrow}} \alpha X \beta \} \\ P'' &= \{ X \to \alpha \in P' \mid X \in V'' \wedge \alpha \in (V'' \cup \Sigma)^* \} \end{aligned}$$

Eliminación de variables inútiles

Demostración

Sea $\mathcal{G}' = (V', \Sigma, P', S)$ tal que $V' = \{ X \in V \mid \exists w.\ X \overset{\star}{\underset{\mathcal{G}}{\Rightarrow}} w \}$.

Sea $\mathcal{G}'' = (V'', \Sigma, P'', S)$ tal que $V'' = \{ X \in V' \mid \exists \alpha, \beta.\ S \overset{\star}{\underset{\mathcal{G}'}{\Rightarrow}} \alpha X \beta \}$.

Considere las siguientes propiedades de $\mathcal{G}$, $\mathcal{G}'$ y $\mathcal{G}''$.

1. Para todo $\alpha \in (V \cup \Sigma)^*$, si $\alpha \overset{\star}{\underset{\mathcal{G}}{\Rightarrow}} w$ entonces $\alpha \overset{\star}{\underset{\mathcal{G}'}{\Rightarrow}} w$.
2. Para todo $\alpha \in (V' \cup \Sigma)^*$, si $S \overset{\star}{\underset{\mathcal{G}'}{\Rightarrow}} \alpha$ entonces $S \overset{\star}{\underset{\mathcal{G}''}{\Rightarrow}} \alpha$.
3. Para todo $\alpha \in (V'' \cup \Sigma)^*$, si $\alpha \overset{\star}{\underset{\mathcal{G}'}{\Rightarrow}} w$ entonces $\alpha \overset{\star}{\underset{\mathcal{G}''}{\Rightarrow}} w$.

Ejercicio: demuestre las propiedades.

Eliminación de variables inútiles

Demostración

PD: $\mathcal{L}(\mathcal{G}'') = \mathcal{L}(\mathcal{G})$.

Como $V'' \subseteq V$ y $P'' \subseteq P$, entonces $\mathcal{L}(\mathcal{G}'') \subseteq \mathcal{L}(\mathcal{G})$.

PD: $\mathcal{L}(\mathcal{G}) \subseteq \mathcal{L}(\mathcal{G}'')$.

Sea $w \in \mathcal{L}(\mathcal{G})$ tal que $S \overset{\star}{\underset{\mathcal{G}}{\Rightarrow}} w$.

- Por la propiedad 1. tenemos que $S \overset{\star}{\underset{\mathcal{G}'}{\Rightarrow}} w$.
- Por la propiedad 2. tenemos que $S \overset{\star}{\underset{\mathcal{G}''}{\Rightarrow}} w$.

Por lo tanto $w \in \mathcal{L}(\mathcal{G}'')$ y concluimos que $\mathcal{L}(\mathcal{G}) \subseteq \mathcal{L}(\mathcal{G}'')$.

Eliminación de variables inútiles

Demostración

PD: Para todo $X \in V''$, X es **útil en** $\mathcal{G}''$.

Como $X \in V''$, entonces $S \overset{\star}{\underset{\mathcal{G}'}{\Rightarrow}} \alpha X \beta$ para algún $\alpha, \beta \in (V' \cup \Sigma)^*$.

Por la propiedad 2. se tiene que: $S \overset{\star}{\underset{\mathcal{G}''}{\Rightarrow}} \alpha X \beta$ y $\alpha, \beta \in (V'' \cup \Sigma)^*$.

Como $X \in V'$ y $\alpha, \beta \in (V' \cup \Sigma)^*$, entonces existen u, v, w tal que:

$$\alpha \overset{\star}{\underset{\mathcal{G}}{\Rightarrow}} u \ , \quad X \overset{\star}{\underset{\mathcal{G}}{\Rightarrow}} v \ , \quad \beta \overset{\star}{\underset{\mathcal{G}}{\Rightarrow}} w$$

Por la propiedad 1. se tiene que: $\alpha \overset{\star}{\underset{\mathcal{G}'}{\Rightarrow}} u \ , \ X \overset{\star}{\underset{\mathcal{G}'}{\Rightarrow}} v \ , \ \beta \overset{\star}{\underset{\mathcal{G}'}{\Rightarrow}} w$.

Por la propiedad 3. se tiene que: $\alpha \overset{\star}{\underset{\mathcal{G}''}{\Rightarrow}} u \ , \ X \overset{\star}{\underset{\mathcal{G}''}{\Rightarrow}} v \ , \ \beta \overset{\star}{\underset{\mathcal{G}''}{\Rightarrow}} w$.

Juntando todo $S \overset{\star}{\underset{\mathcal{G}''}{\Rightarrow}} \alpha X \beta \overset{\star}{\underset{\mathcal{G}''}{\Rightarrow}} uvw$ y X es útil en $\mathcal{G}''$. ■

Cierre de clase

En esta clase vimos:

1. Árboles de derivación.
2. Eliminación de variables inútiles.

Próxima clase: Forma normal de Chomsky