Sesión 17

Forma Normal de Chomsky
(*CNF*)

Forma Normal de Chomsky

- Considere $G = (V, \Sigma, S, P)$ & $S \Rightarrow^* x$ donde $x \in \Sigma^*$ y $|x| = k$
  – Sea l la longitud de la cadena
  – Sea t el número de símbolos terminales
  – Para $S : l + t = 1 + 0 = 1$
  – Para $x : l + t = k + k = 2k$

Forma Normal de Chomsky

- Una propiedad interesante:
  – Si no hay productiones-Λ ni productiones unitarias, en toda derivación $\alpha \Rightarrow \beta$ el valor de $l + t$ se incrementa reescribiendo una variable por una producción de forma:

$$A \to \gamma \quad \text{donde } \gamma \in (V \cup \Sigma)^*$$

Forma Normal de Chomsky

- Una propiedad interesante:
  – En particular, $l + t$ se incrementa en uno si las producciones tienen la siguiente forma:

$$A \to BC \quad \text{(i.e. } l \text{ se incrementa en uno)}$$
$$A \to a \quad \text{(i.e. } t \text{ se incrementa en uno)}$$

  – Por lo tanto, una derivación $S \Rightarrow^* x$ donde $x \in \Sigma^*$ & $|x| = k$ (de $l + t = 1$ a $l + t = 2k$) tiene cuando más $2k - 1$ producciones!

Forma Normal de Chomsky (*CNF*)

- Para toda *GLC* $G = (V, \Sigma, S, P)$ existe una *GLC* $G' = (V', \Sigma, S, P')$ en *CNF* tal que $L(G') = L(G) - \{\Lambda\}$
- Una *GLC* está en *CNF* si toda producción es de alguna de las dos siguientes formas:

$$A \to BC$$
$$A \to a$$

Forma Normal de Chomsky

- Si la gramática G no es ambigua (i.e. es directamente no ambigua o existe una gramática equivalente que genera el mismo lenguaje) su gramática correspondiente en *CNF* también es <u>no ambigua</u>!
- Una gramática en *CNF* es ambigua si y sólo si el lenguaje de la gramática es inherentemente ambiguo!

Removiendo las productiones-Λ

- Definimos a una variable como *nulificable* en una *GLC* $G = (V, \Sigma, S, P)$ como sigue:
  - Si existe una producción de forma $A \rightarrow \Lambda$ en P entonces A es nulificable
  - Si P contiene producciones de forma $A \rightarrow B_1B_2\ldots B_n$ donde $B_1B_2\ldots B_n$ son nulificables (todas) entonces A es nulificable:

    $A \rightarrow B_1B_2\ldots B_n$

    $A \rightarrow \Lambda\Lambda\ldots\Lambda$

    $A \rightarrow \Lambda$
  - Sólo estas variables son nulificables

Removiendo las productiones-Λ

- Dada una *GLC* $G = (V, \Sigma, S, P)$ construir una *GLC* $G_1 = (V, \Sigma, S, P_1)$ sin producciones-Λ como sigue:
  - Sea $P_1 = P$
  - Encontrar todas las variables nulificables en V
  - Para toda producción $A \rightarrow \alpha$ en P, aumentar P_1 con toda producción que se pueda obtenera partir de $A \rightarrow \alpha$ eliminando una o más ocurrencias de las variables nulificables en α
  - Eliminar todas las producciones-Λ de P_1, duplicaciones de una producción y producciones de forma $A \rightarrow A$

Removiendo las productiones-Λ

- Sea una *GLC* $G = (V, \Sigma, S, P)$ donde P contiene las producciones:

  $S \rightarrow AACD$

  $A \rightarrow aAb \mid \Lambda$

  $C \rightarrow aC \mid a$

  $D \rightarrow aDa \mid bDb \mid \Lambda$
- Las variabless nulificabes son A & D:

  $A \rightarrow \Lambda$

  $D \rightarrow \Lambda$

Removiendo las productiones-Λ

- Sea una *GLC* $G = (V, \Sigma, S, P)$ donde P contiene las producciones:

  $S \rightarrow AACD$

  $A \rightarrow aAb \mid \Lambda$

  $C \rightarrow aC \mid a$

  $D \rightarrow aDa \mid bDb \mid \Lambda$
- Agregar producciones en P_1:

  $S \rightarrow AACD \mid ACD \mid AAC \mid CD \mid AC \mid C$

  $A \rightarrow aAb \mid ab \mid \Lambda$

  $C \rightarrow aC \mid a$

  $D \rightarrow aDa \mid bDb \mid aa \mid bb \mid \Lambda$

Removiendo las productiones-Λ

- Producciones en P_1:

  $S \rightarrow AACD \mid ACD \mid AAC \mid CD \mid AC \mid C$

  $A \rightarrow aAb \mid ab \mid \Lambda$

  $C \rightarrow aC \mid a$

  $D \rightarrow aDa \mid bDb \mid aa \mid bb \mid \Lambda$
- Eliminar producciones-Λ (quitar de P_1):

  $S \rightarrow AACD \mid ACD \mid AAC \mid CD \mid AC \mid C$

  $A \rightarrow aAb \mid ab$

  $C \rightarrow aC \mid a$

  $D \rightarrow aDa \mid bDb \mid aa \mid bb$
- Eliminar variables nulificables en *GLC* es como eliminar transiciones-Λ en NFA-Λ

Removiendo las productiones unitarias

- Sea una *GLC* $G = (V, \Sigma, S, P)$ donde P no tiene producciones-Λ:

  $S \rightarrow AACD \mid ACD \mid AAC \mid CD \mid AC \mid C$

  $A \rightarrow aAb \mid ab$

  $C \rightarrow aC \mid a$

  $D \rightarrow aDa \mid bDb \mid aa \mid bb$
- Eliminar producciones unitarias: $S \rightarrow C$

  $S \rightarrow AACD \mid ACD \mid AAC \mid CD \mid AC \mid aC \mid a$

  $A \rightarrow aAb \mid ab$

  $C \rightarrow aC \mid a$

  $D \rightarrow aDa \mid bDb \mid aa \mid bb$

Normalizando la forma de las producciones

- Lados derechos sólo con variables o un símbolo terminal:

$$S \rightarrow AACD \mid ACD \mid AAC \mid CD \mid AC \mid aC \mid a$$
$$A \rightarrow aAb \mid ab$$
$$C \rightarrow aC \mid a$$
$$D \rightarrow aDa \mid bDb \mid aa \mid bb$$

- Reemplazar $S \rightarrow aC$ por $S \rightarrow X_aC$ & $X_a \rightarrow a$:

$$S \rightarrow AACD \mid ACD \mid AAC \mid CD \mid AC \mid X_aC \mid a$$
$$A \rightarrow X_aAX_b \mid X_aX_b$$
$$C \rightarrow X_aC \mid a$$
$$D \rightarrow X_aDX_a \mid X_bDX_b \mid X_aX_a \mid X_bX_b$$
$$X_a \rightarrow a$$
$$X_b \rightarrow b$$

Obtener Forma Normal de Chomsky

- Reemplazar $S \rightarrow ABC\alpha$ por $S \rightarrow AT$ & $T \rightarrow BC\alpha$

$$S \rightarrow AACD \mid ACD \mid AAC \mid CD \mid AC \mid X_aC \mid a$$
$$A \rightarrow X_aAX_b \mid X_aX_b$$
$$C \rightarrow X_aC \mid a$$
$$D \rightarrow X_aDX_a \mid X_bDX_b \mid X_aX_a \mid X_bX_b$$
$$X_a \rightarrow a$$
$$X_b \rightarrow b$$

- Obtener forma normal de Chomsky:

$S \rightarrow AT_1$ $\quad T_1 \rightarrow AT_2$ $\quad T_2 \rightarrow CD$

$S \rightarrow AU_1$ $\quad U_1 \rightarrow CD$

$S \rightarrow AV_1$ $\quad V_1 \rightarrow AC$

La Gramática en CNF:

$S \rightarrow AT_1$ $\quad T_1 \rightarrow AT_2$ $\quad T_2 \rightarrow CD$

$S \rightarrow AU_1$ $\quad U_1 \rightarrow CD$

$S \rightarrow AV_1$ $\quad V_1 \rightarrow AC$

$S \rightarrow CD \mid AC \mid X_aC \mid a$

$A \rightarrow X_aW_1$ $\quad W_1 \rightarrow AX_b$ $\quad A \rightarrow X_aX_b$

$C \rightarrow X_aC \mid a$

$D \rightarrow X_aY_1$ $\quad Y_1 \rightarrow DX_a$

$D \rightarrow X_bZ_1$ $\quad Z_1 \rightarrow DX_b$

$D \rightarrow X_aX_a \mid X_bX_b$

$X_a \rightarrow a$

$X_b \rightarrow b$

¡Una analogía!

GLC ambigua	*RE* o NFA-Λ
Variables nulificables	Cerradura-Λ
Sin producciones-Λ o prod. unitarias	NFA
CNF	DFA

¡La ambigüedad como un medio para expresar abstracciones!

GLC ambigua	*RE* o NFA-Λ
Variables nulificables	Cerradura-Λ
Sin producciones-Λ o prod. unitarias	NFA
Implementación: CNF	DFA

Forma Normal de Chomsky

- Forma Normal de Chomsky (*CNF*):

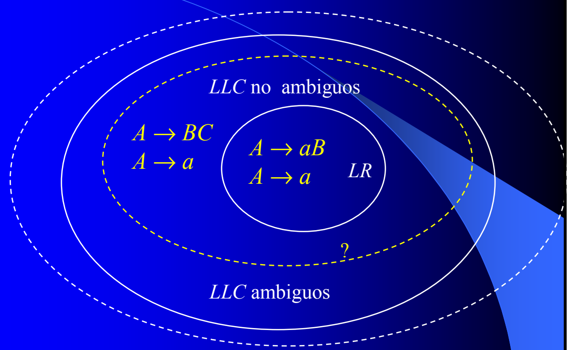
$$A \rightarrow BC$$
$$A \rightarrow a$$

- Gramáticas Regulares:

$$A \rightarrow aB$$
$$A \rightarrow a$$

- ¡De lenguajes regulares a lenguajes no ambiguos (casi)!

¿Existe la clase de los lenguajes no ambiguos?



- No hay un algoritmo para decidir si una gramática es ambigua
- ¡No hay modo de saber si un lenguaje es inherentemente ambiguo!