

Sesión 24

Lema de bombeo para *LLC*

Dr. Luis Pineda, IIMAS, 2010

¿Cómo podemos decir si un lenguaje es libre del contexto?

- Definir una GLC o diseñar un AP para el lenguaje
- Pero que tal si el lenguaje se describe por otros medios:
 - $L = \{a^i b^i c^i \mid i \geq 1\}$
 - ¿Es este lenguaje un *LLC*?

Dr. Luis Pineda, IIMAS, 2010

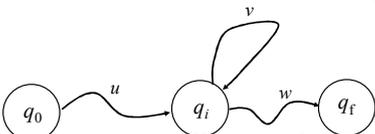
¿Cómo podemos decir si un lenguaje es libre del contexto?

- Usar el Lema del bombeo para *LLC*
 - Asumir que L es un *LLC*
 - Si se llega a una contradicción L no lo es
- Antecedentes:
 - Forma Normal de Chomsky (1959)
 - Debido a Bar-Hillel, Perles & Shamir (1961)
 - El lema de bombeo para los *LRs* es una simplificación del lema de bombeo para *LLC*

Dr. Luis Pineda, IIMAS, 2010

El lema de bombeo para *LR*

- Supongamos que L es un *LR* reconocido por un FA con n estados; entonces, para todo $x \in L$ con $|x| \geq n$, $x = uvw$ para las cadenas u, v, w que satisfacen:
 - $|uv| \leq n$
 - $|v| > 0$
 - For any $m \geq 0$, $uv^m w \in L$



Dr. Luis Pineda, IIMAS, 2010

El "ciclo" en cadenas de *LLC*

- Si una derivación es suficientemente larga, alguna variable se tiene que repetir:

$$S \Rightarrow^* vAz \Rightarrow^* vwAyz \Rightarrow^* vwxyz$$
 donde $v, w, x, y, z \in \Sigma^*$
- Los contextos que están antes y después de las variables en el lado derecho de una producción (e.g. w & y en $A \rightarrow wAy$) se bombea con la repetición de dicha variable en la derivación:

$$S \Rightarrow^* vAz \Rightarrow^* vwAyz \Rightarrow^* vw^2Ay^2z \Rightarrow^* vw^3Ay^3z \Rightarrow^* \dots$$

Dr. Luis Pineda, IIMAS, 2010

El "ciclo" en cadenas de *LLC*

- Bombeando w & y
- Si x se deriva de A ($A \Rightarrow^* x$):

$$vAz \Rightarrow^* vxz \in L$$

$$vwAyz \Rightarrow^* vwxyz \in L$$

$$vw^2Ay^2z \Rightarrow^* vw^2xy^2z \in L$$

$$vw^3Ay^3z \Rightarrow^* vw^3xy^3z \in L$$

$$\dots$$

Dr. Luis Pineda, IIMAS, 2010

El lema de bombeo para LLC

- Sea $G = (V, \Sigma, S, P)$ una GLC en FNC con un total de p variables. Toda cadena u en $L(G)$ con $|u| \geq 2^{p+1}$ pueden escribirse como $u = vwxyz$, para las cadenas v , w , x , y & z que satisfacen:
 - $|wy| > 0$
 - $|wxy| \leq 2^{p+1}$
 - Para toda $m \geq 0$, $vw^mxy^mz \in L$

Dr. Luis Pineda, IIMAS, 2010

El lema de bombeo para LLC

- Las condiciones no vienen de Marte: para toda $u = vwxyz$
 - $|u| \geq 2^{p+1}$
 - $|wy| > 0$
 - $|wxy| \leq 2^{p+1}$
 - para un parámetro p (i.e. el número de variables en V)

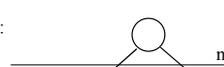
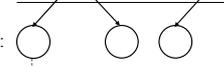
Dr. Luis Pineda, IIMAS, 2010

La historia en breve...

- Las gramáticas en FNC producen estructuras sintácticas en forma de árboles binarios; sólo los nodos en el último nivel (categorías léxicas) tienen un solo hijo (o hija)
- Un árbol binario de altura h tiene una cosecha $\leq 2^{h-1}$; por lo tanto, un árbol binario con más de 2^{h-1} hojas tiene una altura mayor que h

Dr. Luis Pineda, IIMAS, 2010

Propiedades de los árboles binarios

- Altura de una trayectoria: número de nodos en la trayectoria
 - En un árbol binario completo: número de nodos en el nivel h es 2^{h-1}
- altura = 1:  nodos = $2^{1-1} = 2^0 = 1$
nivel 0
- altura = 2:  nodos = $2^{2-1} = 2^1 = 2$
nivel 1
- altura = 3:  nodos = $2^{3-1} = 2^2 = 4$
nivel 2
- Derivación u con más de 2^{h-1} símbolos tiene una altura mayor que h
 - Nodos en nivel $l = 2^l$ (nivel = altura - 1)

Dr. Luis Pineda, IIMAS, 2010

La historia en breve...

- Si una gramática tiene p variables, una derivación con 2^{p+1} símbolos tiene cuando menos una trayectoria con altura $p + 2$, por lo que cuando menos una variable tiene que repetirse (el último nodo es un símbolo terminal)

Dr. Luis Pineda, IIMAS, 2010

El lema de bombeo para LLC

- Las condiciones no vienen de Marte: para toda $u = vwxyz$
 - $|u| \geq 2^{p+1}$
 - $|wy| > 0$
 - $|wxy| \leq 2^{p+1}$
 - para un parámetro p (i.e. el número de variables en V)
 - (i) garantiza que u sea suficientemente grande para que una variable se repita
 - (ii) & (iii) son condiciones necesarias en cadenas con variables repetidas

Dr. Luis Pineda, IIMAS, 2010

Estructuras en FNC

- Altura de un árbol: altura de la trayectoria más larga
- Una estructura producida por una GLC en FNC puede tener menos pero no más de 2^{h-1} símbolos en el nivel h

Producción: $S \rightarrow AB$

Producción: $A \rightarrow CD$

Producción: $B \rightarrow b$

No. de variables h : $\leq 2^{h-1}$ nodos

- Una cadena con más de 2^{h-1} símbolos tiene una altura mayor que h

Dr. Luis Pineda, IIMAS, 2010

Número de variables y profundidad

- Si el número de hojas es $> 2^p$ entonces la altura es cuando menos $p+2$

$h=1$: nodos = $2^{1-1} = 2^0 \leq 1$

$h=2$: nodos = $2^{2-1} = 2^1 \leq 2$

$h=p+1$: nodos $\leq 2^p$

Entonces $h \geq p+2$ Si el número de nodos $> 2^p$

Dr. Luis Pineda, IIMAS, 2010

Tamaño de cadenas y longitud de la derivación

- If $|u| > 2^{p+1}$ entonces la altura es cuando menos $p+2$

$h=1$: nodos = $2^{1-1} = 2^0 \leq 1$

$h=2$: nodos = $2^{2-1} = 2^1 \leq 2$

$h \geq p+2$: nodos = $|u| \leq 2^{p+1}$

- Si hay p variables diferentes en la gramática, una cadena u tal que $|u| > 2^{p+1}$ tienen un árbol sintáctico cuya altura es cuando menos $p+2$

Dr. Luis Pineda, IIMAS, 2010

Estructura en FNC con repetición de variable

$u = abbabaa$
 Altura = 7
 Variable repetida = A

Dr. Luis Pineda, IIMAS, 2010

$u = vwxyz$

$v = ab$
 $wxy = baba$
 $z = a$

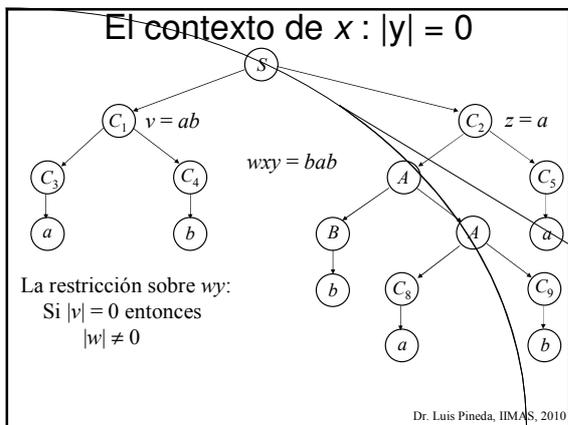
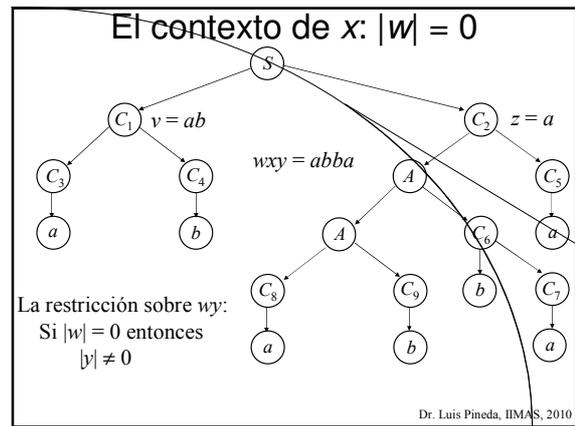
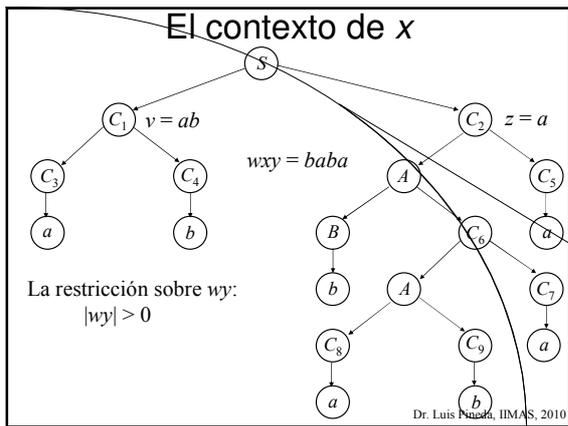
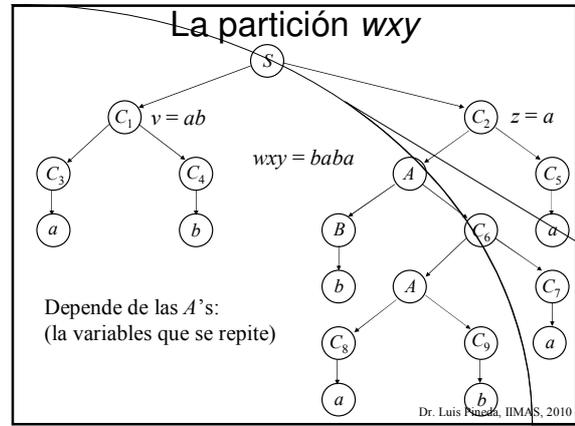
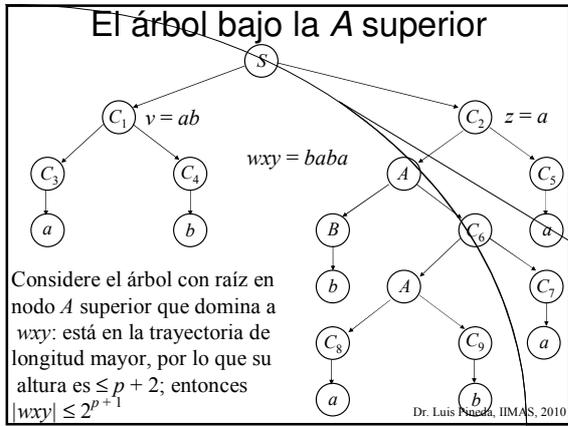
$u = abbabaa$ and $|u| = 7$

Dr. Luis Pineda, IIMAS, 2010

Si $|u| \geq 2^{p+1}$ una variable se repite!

Considere una trayectoria de longitud máxima $p+2$: Los nodos de hasta abajo son terminales y los nodos $p+1$ superiores contienen variables!
 Si sólo hay p variables, una variable se tiene que repetir!

Dr. Luis Pineda, IIMAS, 2010



El lema de bombeo para LLC

- Pero si no sabemos si L es LLC, no sabemos cual sería su gramática, ni cuantas variables tendría!
- Sea L una CFL. Existe un entero n tal que para toda u que satisface $|u| \geq n$ existen cadenas v, w, x, y & z que satisfacen:
 - $u = vwxyz$
 - $|wy| > 0$
 - $|wxy| \leq n$
 - Y para todo $m \geq 0, vw^mxy^mz \in L$

Dr. Luis Pineda, IIMAS, 2010

El lema de bombeo para LLC

- Prueba:
 - Encontrar una GLC en FNC que genere $L - \{\Lambda\}$.
 - Sea p el número de variables en esta gramática & $n = 2^{p+1}$

Dr. Luis Pineda, IIMAS, 2010

Uso del El lema de bombeo para LLC

- Determinar si un lenguaje es un LLC
 - $L = \{a^i b^j c^i \in \Sigma^* \mid i \geq 1\}$
- Estrategia:
 - Asumir que el lema de bombeo para los LLC se satisface para L
 - Si se llega a una contradicción, L no es un LLC!

Dr. Luis Pineda, IIMAS, 2010

$$L = \{a^i b^j c^i \in \Sigma^* \mid i \geq 1\}$$

- Sea n la constante & $u = a^n b^n c^n$
 - $|u| = 3n$ (Ok: $n = 2^{p+1}$)
- Particionar u en $vwxyz$ tal que $|wxy| \leq n$ & $|wy| > 0$;
- Dado que $|wxy| \leq n$, esta subcadena tiene cuando más dos tipos de símbolos (sólo a 's o sólo b 's o sólo c 's, o a 's y b 's o b 's y c 's)

Dr. Luis Pineda, IIMAS, 2010

$$L = \{a^i b^j c^i \in \Sigma^* \mid i \geq 1\}$$

- Escoger $m = 0$ en $vw^m xy^m z$
 - Dado que $|wy| > 0$, entonces $|w| > 0$ o $|y| > 0$ (o ambos)
 - Los segmentos de dos símbolos (e.g. a & b) que contienen a w & y tienen menos símbolos (i.e. $m = 0$) que el segmento que incluye al símbolo restante, que no está en wxy (i.e. c)
 - L no es un LLC!

Dr. Luis Pineda, IIMAS, 2010

$$L = \{a^i b^j c^i \in \Sigma^* \mid i \geq 1\}$$

Escoger $m = 0$ en $vw^m xy^m z$:

- Caso 1: wxy en el bloque de a 's:
 - $w = a^i, y = a^j$ & $a^{n-i-j} b^n c^n \notin L$ ya que $i > 0$ o $j > 0$ & $n - i - j < n$
- Caso 2: wxy tiene a 's & b 's:
 - $a^i b^j c^n \notin L$ ya que $i < n$ o $j < n$ (o ambos) & $i + j < 2n$
- Caso 3: wxy está en las b 's:
 - $w = b^i, y = b^j$ & $a^n b^{n-i-j} c^n \notin L$ ya que $i > 0$ o $j > 0$ & $n - i - j < n$

Dr. Luis Pineda, IIMAS, 2010

$$L = \{a^i b^j c^i \in \Sigma^* \mid i \geq 1\}$$

Escoger $m = 0$ en $vw^m xy^m z$:

- Caso 4: wxy tiene b 's & c 's:
 - $a^n b^i c^j \notin L$ ya que $i < n$ o $j < n$ (o ambos) & $i + j < 2n$
- Caso 5: wxy está en las c 's:
 - $w = c^i, y = c^j$ & $a^n b^n c^{n-i-j} \notin L$ ya que $i > 0$ o $j > 0$ & $n - i - j < n$

Dr. Luis Pineda, IIMAS, 2010

$L = \{abc^i \in \Sigma^* \mid i \geq 1\}$

- La abstracción:
 - Los segmentos de dos símbolos que contienen w & y tienen menos símbolos que el segmento que contiene al símbolo que no está en wxy (i.e. con $m = 0$)
- Caso 1: wxy está en las a 's:
- Caso 2: wxy tiene a 's & b 's:
 - $a^i b^j c^n \notin L$
- Caso 3: wxy está en las b 's:
- Caso 2 incluye a los casos 1 & 3!
 - El segmento $|a^i b^j| < 2n$

Dr. Luis Pineda, IIMAS, 2010

$L = \{abc^i \in \Sigma^* \mid i \geq 1\}$

- La abstracción:
 - Los segmentos de dos símbolos que contienen w & y tienen menos símbolos que el segmento que contiene al símbolo que no está en wxy (i.e. con $m = 0$)
- Caso 3: wxy está en las b 's:
- Caso 4: wxy tiene b 's & c 's:
 - $a^n b^i c^j \notin L$
- Caso 5: wxy está en las c 's:
- El caso 4 incluye a los casos 3 & 5!
 - El segmento $|b^i c^j| < 2n$

Dr. Luis Pineda, IIMAS, 2010

$L = \{abc^i \in \Sigma^* \mid i \geq 1\}$

- La abstracción:
 - Los segmentos de dos símbolos que contienen w & y tienen menos símbolos que el segmento que contiene al símbolo que no está en wxy
- La abstracción: Sea $m = 0$
 - p : El segmento $|a^i b^j| < 2n$ & $a^i b^j c^n \notin L$
 - q : El segmento $|b^i c^j| < 2n$ & $a^n b^i c^j \notin L$
 - L no es un LLC ya que p o q

Dr. Luis Pineda, IIMAS, 2010

$L = \{x \in \{a, b, c\}^* \mid n_a(x) < n_b(x) \text{ \& } n_a(x) < n_c(x)\}$

- Sea n la constante & $u = a^n b^{n+1} c^{n+1}$
 - $|u| = 3n + 2$ (ok: $n = 2^{p+1}$)
- Partición de u en $wxyz$ tal que $|wxy| \leq n$ & $|wy| > 0$
 - Otra vez wxy tiene cuando más dos tipos de símbolos
- Caso 1:
 - w o y tienen cuando menos una a
 - sea $m = 2$ & $a^i b^j c^{n+1} \notin L$ ya que $i \geq n + 1$ y por lo tanto $a^i \geq c^{n+1}$

Dr. Luis Pineda, IIMAS, 2010

$L = \{x \in \{a, b\}^* \mid n_a(x) < n_b(x) \text{ \& } n_a(x) < n_c(x)\}$

- Sea n la constante & $u = a^n b^{n+1} c^{n+1}$
 - $|u| = 3n + 2$ (ok: $n = 2^{p+1}$)
- Partición de u en $wxyz$ tal que $|wxy| \leq n$ & $|wy| > 0$
 - Otra vez wxy tiene cuando más dos tipos de símbolos
- Caso 2:
 - w o y no tienen a 's
 - sea $m = 0$ & $a^i b^j c^k \notin L$ ya que $i < n + 1$ o $j < n + 1$ (o ambos) & $n_a(u) \geq n_b(u)$ o $n_a(u) \geq n_c(u)$

Dr. Luis Pineda, IIMAS, 2010

