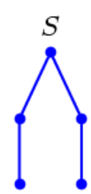
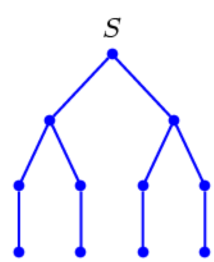
4.12. Lema de bombeo para LIC

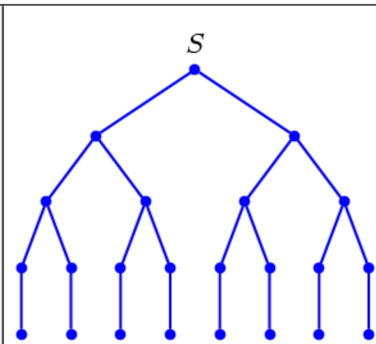
Una de las consecuencias más importantes de la Forma Normal de Chomsky es el lema de bombeo para lenguajes independientes del contexto, el cual es útil, entre muchas aplicaciones, para demostrar que ciertos lenguajes no son LIC.

Nos referiremos a gramáticas en FNC con variable inicial no recursiva. Puesto que las producciones son unitarias ($A \to a$) o binarias ($A \to BC$), en cada nodo el árbol de una derivación se ramifica en dos nodos, a lo sumo. Tales árboles se denominan **binarios**. Si la producción $S \to \lambda$ está presente, su único propósito es generar la cadena λ.

4.12.1 Teorema. *Sea* $G = (V, \Sigma, S, P)$ *una gramática en FNC y* $w \in \Sigma^*$. *Si la longitud de la trayectoria más larga en un árbol de derivación de* $S \stackrel{*}{\Rightarrow} w$ *tiene* k *(o menos) nodos, entonces* $|w| \leq 2^{k-2}$. *Aquí* $k \geq 2$.

Demostración: La siguiente tabla muestra las relaciones obtenidas entre k = número de nodos de la trayectoria más larga de $S \stackrel{*}{\Rightarrow} w$ y la longitud de w, en los casos $k = 2$, $k = 3$, $k = 4$ y $k = 5$. En la tabla se muestran los casos extremos, es decir, los árboles con el mayor número posible de nodos. Se observa que $|w| \leq 2^{k-2}$. Una demostración rigurosa del caso general se hace por inducción sobre k. □

k = número de nodos de la trayectoria más larga	Árbol de derivación	Longitud de w
$k = 2$		$\|w\| = 1 = 2^0 = 2^{k-2}$
$k = 3$		$\|w\| \leq 2 = 2^1 = 2^{k-2}$
$k = 4$		$\|w\| \leq 4 = 2^2 = 2^{k-2}$

$k = 5$	 	$\lvert w \rvert \leq 8 = 2^3 = 2^{k-2}$

4.12.2 Corolario. *Sea $G = (V, \Sigma, S, P)$ una gramática en FNC y $w \in \Sigma^*$.*

(1) Si la longitud de la trayectoria más larga en un árbol de derivación de $S \stackrel{*}{\Rightarrow} w$ tiene $k + 2$ (o menos) nodos, entonces $|w| \leq 2^k$. Aquí $k \geq 0$.

(2) Si $|w| > 2^k$ (con $k \geq 0$) entonces la longitud de la trayectoria más larga en un árbol de derivación de $S \stackrel{*}{\Rightarrow} w$ tiene más de $k + 2$ nodos.

<u>Demostración</u>:

(1) Se sigue inmediatamente del Teorema 4.12.1.

(2) Es la afirmación contra-recíproca de la parte (1). □

4.12.3. Lema de bombeo para LIC. *Dado un LIC L, existe una constante n (llamada constante de bombeo de L) tal que toda $z \in L$ con $|z| > n$ se puede descomponer en la forma $z = uvwxy$ donde:*

(1) $|vwx| \leq n$.

(2) $uv^iwx^iy \in L$ para todo $i \geq 0$.

(3) $v \neq \lambda$ ó $x \neq \lambda$.

<u>Demostración</u>: Sea $G = (V, \Sigma, S, P)$ una gramática en FNC, con variable inicial no recursiva, tal que $L(G) = L$. Tal gramática existe por el Teorema 4.10.1 y el Teorema 4.10.2. Sea $k = |V| =$ número de variables de G y $n = 2^k$. Sea $z \in L$ con $|z| > n = 2^k$. Por la parte (2) del Corolario 4.12.2, la trayectoria más larga en el árbol de una derivación $S \stackrel{*}{\Rightarrow} z$ tiene más de $k + 2$ nodos. Consideremos los últimos $k + 2$ nodos de tal trayectoria (siguiendo el orden que va desde la raíz S hasta las hojas del árbol). El

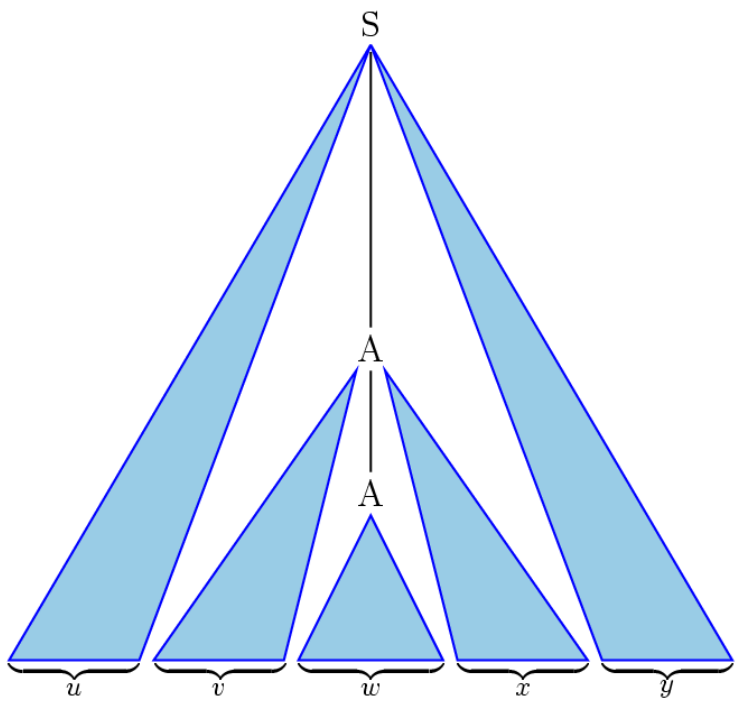
último nodo de esa trayectoria es un terminal $a \in \Sigma$ y los restantes $k+1$ nodos son variables. Como hay sólo k variables en la gramática, entonces hay por lo menos una variable $A \neq S$ repetida en la trayectoria. Por lo tanto, existen cadenas $u, v, w, x, y \in \Sigma^*$ tales que

$$S \stackrel{*}{\Rightarrow} uAy, \qquad A \stackrel{*}{\Rightarrow} vAx, \qquad A \stackrel{*}{\Rightarrow} w.$$

Así que

$$S \stackrel{*}{\Rightarrow} uAy \stackrel{*}{\Rightarrow} uvAxy \stackrel{*}{\Rightarrow} uvwxy = z.$$

La siguiente gráfica ilustra la situación:



Puesto que la trayectoria más larga en el árbol de derivación de $A \stackrel{*}{\Rightarrow} vAx \stackrel{*}{\Rightarrow} vwx$ tiene $k+2$ nodos o menos, por la parte (1) del Corolario Corolario 4.12.2, podemos concluir que $|vwx| \leq 2^k = n$. Además:

$$S \stackrel{*}{\Rightarrow} uAy \stackrel{*}{\Rightarrow} uvAxy \stackrel{*}{\Rightarrow} uv^iAx^iy \stackrel{*}{\Rightarrow} uv^iwx^iy, \quad \text{para todo } i \geq 0.$$

Obsérvese que el caso $i = 0$ corresponde a la derivación $S \stackrel{*}{\Rightarrow} uAy \stackrel{*}{\Rightarrow} uwy$.

Finalmente, la derivación $A \stackrel{*}{\Rightarrow} vAx$ se puede escribir como

$$A \Rightarrow BC \stackrel{*}{\Rightarrow} vAx$$

utilizando una producción de la forma $A \to BC$ como primer paso. Se deduce que u y x no pueden ser ambas λ porque se tendría $BC \stackrel{*}{\Rightarrow} A$, lo cual es imposible en una gramática en FNC (recuérdese que la única producción λ en la gramática es, posiblemente, $S \to \lambda$; pero S no aparece en el cuerpo de ninguna producción de G ya que S no es recursiva). Se deduce entonces que $v \neq \lambda$ ó $x \neq \lambda$. Esto demuestra las propiedades (1), (2) y (3) del lema de bombeo. $\square$

Ejemplo Demostrar que el lenguaje $L = \{a^i b^i c^i : i \geq 0\}$ sobre $\Sigma = \{a, b, c\}$ no es un LIC.

Solución: Argumento por contradicción. Si L fuera LIC, por el lema de bombeo, existiría una constante de bombeo n. Sea $z = a^n b^n c^n$; se tiene que $z \in L$ y $|z| > n$. Por lo tanto, z se puede descomponer como $z = a^n b^n c^n = uvwxy$ con las propiedades (1), (2) y (3) del lema de bombeo. Puesto que $|vwx| \leq n$, en la cadena vwx no pueden aparecer los tres terminales a, b y c simultáneamente (para que aparezcan los tres terminales simultáneamente, una subcadena de $a^n b^n c^n$ debe tener longitud $\geq n+2$). Como $v \neq \lambda$ ó $x \neq \lambda$, se distinguen dos casos:

- Caso 1. Alguna de las cadenas v ó x contiene dos tipos de terminales. Entonces en uv^2wx^2y aparecen algunas $b$es seguidas de $a$es o algunas $c$es seguidas de $b$es. En cualquier caso, $uv^2wx^2y \notin L$.

- Caso 2. Las cadenas v y x contienen un sólo tipo de terminal cada una (o sólo $a$es o sólo $b$es o sólo $c$es). Como en vwx no aparecen los tres terminales a, b y c simultáneamente, en la cadena bombeada uv^2wx^2y se altera el número de dos de los terminales a, b, c, a lo sumo, pero no de los tres. Por lo tanto, $uv^2wx^2y \notin L$.

Pero el lema de bombeo afirma que $uv^2wx^2y \in L$. Esta contradicción muestra que L no es un LIC.

Ejemplo Demostrar que el lenguaje $L = \{a^i : i \text{ es primo}\}$ sobre $\Sigma = \{a\}$ no es un LIC.

Solución: Argumento por contradicción. Si L fuera LIC, por el lema de bombeo, existiría una constante de bombeo n. Sea $z = a^m$ con m primo $m > n$ y $m > 2$ (m existe porque el conjunto de los números primos es infinito). Entonces $z \in L$ y $|z| > n$. Por lo tanto, z se puede descomponer como $z = a^m = uvwxy$ con las propiedades (1), (2) y (3) del lema de bombeo.

Sea $|u| + |w| + |y| = k$; entonces $|v| + |x| = m - k \geq 1$. Por el lema de bombeo, $uv^iwx^iy \in L$ para todo $i \geq 0$; es decir, $|uv^iwx^iy|$ es primo para todo $i \geq 0$. Pero

$$|uv^iwx^iy| = k + |v^i| + |x^i| = k + i|v| + i|x| = k + i(|v| + |x|) = k + i(m - k).$$

(i) Si $k = 0$, tomando $i = m$ se obtiene que $k + i(m - k) = 0 + i(m - 0) = im = mm$ que no es primo, pues $m > 2$.

(ii) Si $k = 1$, tomando $i = 0$ se obtiene que $k+i(m-k) = 1+i(m-1) = 1+0(m-1) = 1$ que no es un número primo.

(iii) Si $k > 1$, tomando $i = k$ se obtiene que

$$|uv^k wx^k y| = k + k(m - k) = k(1 + m - k),$$

el cual no es un número primo pues $k > 1$ y como $m-k \geq 1$, entonces $1+m-k \geq 2$.

Por (i), (ii) y (iii) se puede escoger i de tal manera que $|uv^i wx^i y|$ no sea un número primo, lo cual contradice que $uv^i wx^i y \in L$ para todo $i \geq 0$. Esta contradicción muestra que L no es un LIC.

Ejercicios de la sección 4.12

Utilizar el lema de bombeo para demostrar que los siguientes lenguajes no son LIC:

1. $L = \{a^i b^i c^j : j \geq i\}$, sobre $\Sigma = \{a, b, c\}$.
2. $L = \{a^i b^j c^k : 1 \leq i \leq j \leq k\}$, sobre $\Sigma = \{a, b, c\}$.
3. $L = \{a^i b^{2i} a^i : i \geq 1\}$, sobre $\Sigma = \{a, b\}$.
4. $L = \{a^i b^i c^i d^i : i \geq 0\}$, sobre $\Sigma = \{a, b, c, d\}$.
5. $L = \{a^i b^j c^i d^j : i, j \geq 0\}$, sobre $\Sigma = \{a, b, c, d\}$.
6. $L = \{ww : w \in \{a, b\}^*\}$.
7. $L = \{a^i : i \text{ es un cuadrado perfecto}\}$, sobre $\Sigma = \{a\}$.