

Non-context-Free Languages

The Pumping Lemma

Theorem. (*Pumping Lemma*) Let L be context-free. There exists a positive integer p such that for every $w \in L$ of length at least p , w is divided into five pieces, $w = uvxyz$, such that

- for each $i \geq 0$, $uv^ixy^iz \in L$,
- $|vy| > 0$, and
- $|vxy| \leq p$.

Proof Let $L = L(G)$ for some CNF grammar $G = (V, \Sigma, R, S)$. Let $m = || V ||$ and $p = 2^m$. Let $w, |w| \geq p$, be in L and T be a derivation tree for w .

For any subtree R of T , its non-leaf nodes are all variables and its leaves are symbols with unique parents and form a substring of w .

Proof of Pumping Lemma (cont,d)

Claim. *In every subtree R of T with $\geq 2^{m-1} + 1$ leaves there are two nodes α and β that are labeled by the same variable and are on the same downward path from the root to a leaf.*

Proof of Claim Let R be a subtree of T with $\geq 2^{m-1} + 1$ leaves. Since the complete binary tree of depth $m - 1$ has 2^{m-1} leaves, R has a **downward path of length $\geq m$** . The path has **$\geq m + 1$ nodes**. Since there are **only m variables**, by the pigeonhole principle, the path has **two nodes with the same label**. ■ Claim

Proof of Pumping Lemma (cont,d)

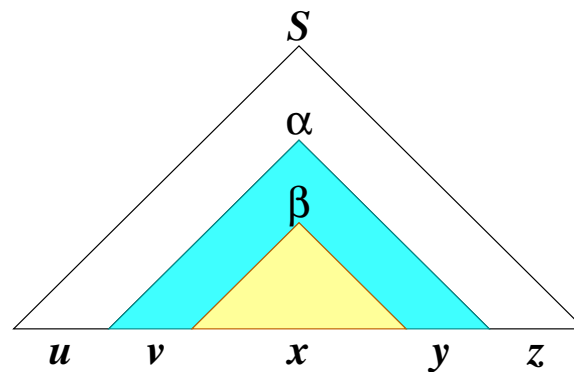
By the Claim there is a node in T whose label coincides with that of a descendant. Let α be one such node that is **the farthest from the root**.

Here **neither the left subtree nor the right subtree of α has more than 2^{m-1} leaves**; otherwise, by the Claim we would find, in one of the two subtrees, a pair of nodes on a downward path labeled by the same variable, which would contradict our assumption that α is the farthest.

Proof of Pumping Lemma (cont,d)

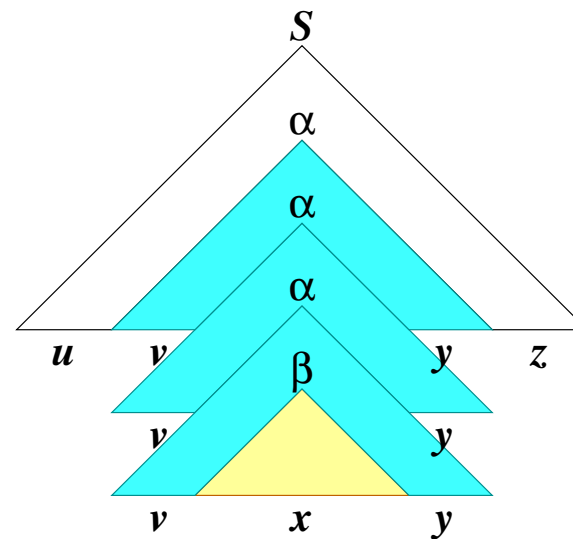
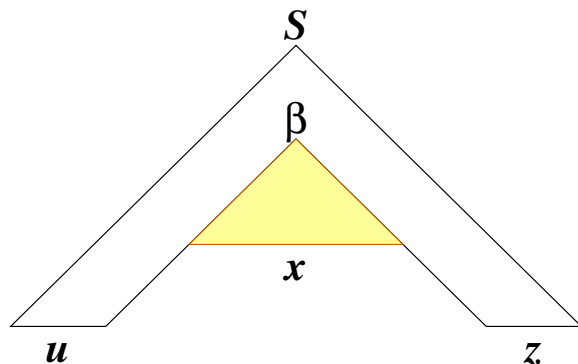
Let β be the descendant of α with the same label as α . **Replacing α by β , or repeatedly replacing β by α , produces a valid derivation tree.**

Let x be the substring under β , $r = vxy$ the substring under α , with v and y to the left and to the right of x , respectively, and $w = urz$ with u and z to the left and to the right of r , respectively.



Proof of Pumping Lemma (cont,d)

Then replacing α by β corresponds to eliminating v and y ; and replacing β by α corresponds to inserting a v before x and a y after x . So, for every $i \geq 0$, $uv^i xy^i z \in L$.




Proof of Pumping Lemma (cont,d)

Since G does not have ϵ rules, either v or y is nonempty, so $|vy| > 0$. Since both left and right subtrees of α have at most 2^{m-1} leaves, α has at most 2^m leaves, and so $|vxy| \leq p$. This proves the lemma. ■

Example 1

$A = \{0^n 1^n 2^n \mid n \geq 0\}$ is not context free.

Proof Assume, on the contrary, that A is context free. By the Pumping Lemma there exists a constant p such that every $w \in A$ of length $\geq p$ is divided into $w = uvxyz$ such that $|vxy| \leq p$, $|vy| \geq 1$, and for every $i \geq 0$, $uv^i xy^i z \in A$.

Let $w = 0^p 1^p 2^p$. Since $|vxy| \leq p$, vxy is either in $0^* 1^*$ or $1^* 2^*$. So it is not the case $uv^2 xy^2 z$ has the same number of 0s, 1s, as 2s. 

Example 2

$B = \{a\#b\#c \mid a, b \text{ and } c \text{ are binary numbers such that } a + b = c\}$ is not context free.

Proof Assume, on the contrary, that B is context free. Let p be the constant from the Pumping Lemma for B . Let $w = 10^p\#10^p\#10^{p+1}$, where $a = b = 2^p$ and $c = 2^{p+1}$. Let $uvxyz$ be the decomposition of w as in the lemma.

For “pumping” to be possible, v has to be a nonempty part of a or of b , and y a nonempty part of c . If v either is a part of a or contains the ‘1’ of b , since $|vxy| \leq p$, y cannot contain a part of c . Thus, v is a part of b and $v \in 0^*$.

Proof Continued

If y contains the first symbol of c , then (pumping down) uxz is not in B because now c is 0 while $a = 2^p$.

If $y \in 0^*$, then (pumping up) $uv^2xy^2z \notin B$ because now the equation becomes $2^p + 2^q = 2^r$ for some $q > p$.

Thus, B is not context-free. 


Example 3

$C = \{ww \mid w \in \{0,1\}^*\}$ is not context free.

Proof Assume C is context free. Let p the constant from the pumping lemma for C .

Let $w = 0^p 1^p 0^p 1^p$, which is in C .

Let $w = uvxyz$ be the decomposition of w such that $|vy| > 0$, $|vxy| \leq p$, and for every $i \geq 0$, $uv^i xy^i z \in C$.

If v contains a symbol from the first 0^p then y cannot contain one from the second 0^p , so pumping (up or down) doesn't work. If v contains symbols from the first 1^p then y cannot contain one from the second 1^p , so pumping doesn't work. If v contains only symbols from the second $0^p 1^p$ then pumping does not work. 

Application

Corollary. *The class of context-free languages is not closed under intersection.*

Proof Let $L_1 = \{0^i 1^j 2^k \mid i = j\}$ and $L_2 = \{0^i 1^j 2^k \mid j = k\}$. Then L_1 and L_2 are both context-free. If the class were closed under intersection then $L_1 \cap L_2 = \{0^n 1^n 2^n \mid n \geq 0\}$ would be context-free. ■

Corollary. *The class of context-free languages is not closed under complement.*