

# The Pumping Lemma for Context-Free Languages

Example. Consider the non-Chomsky grammar

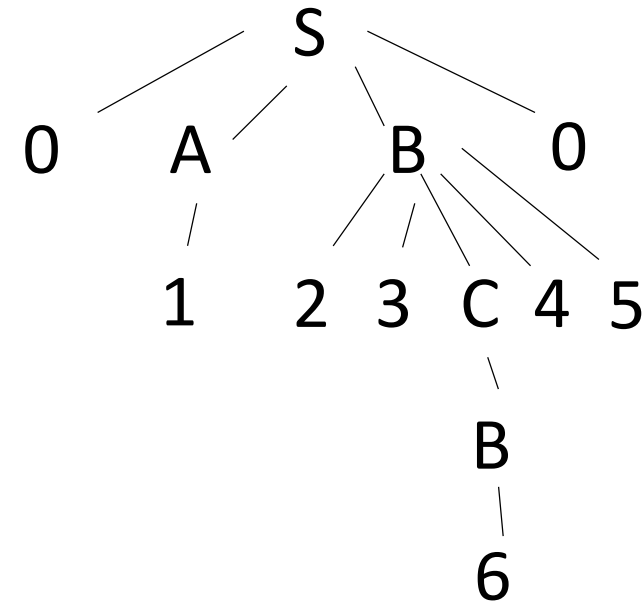
$S \Rightarrow 0AB0$

$A \Rightarrow 1$

$B \Rightarrow 23C45 \mid 6$

$C \Rightarrow B$

Parse 01236450



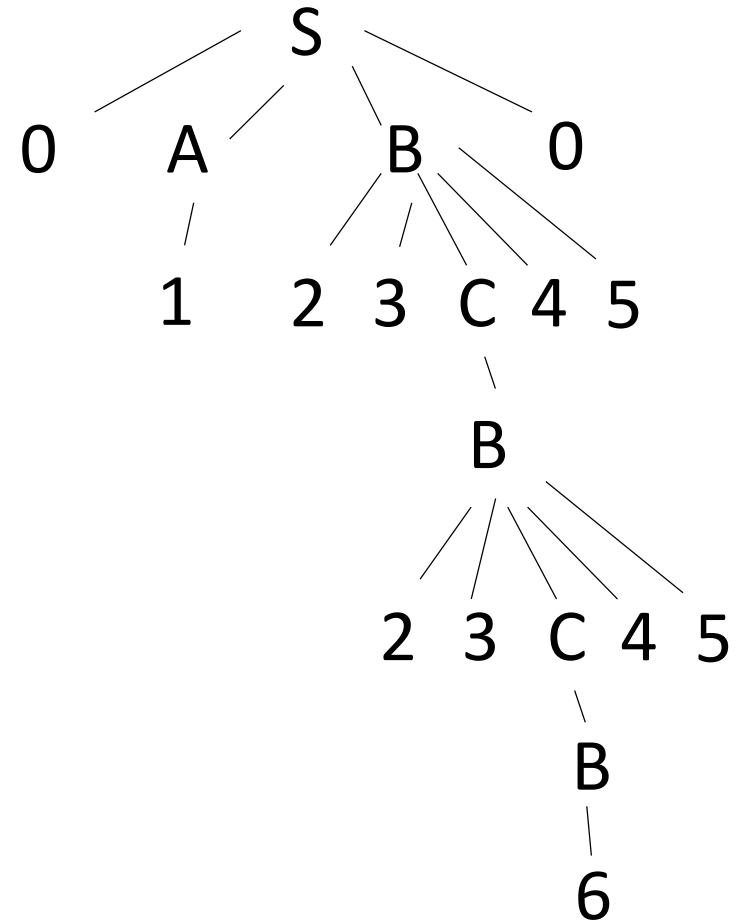
Because of the repeated use of B  
 we could also derive  
 12323645450

In fact, we could derive

$$01(23)^n6(45)^n0$$

for any  $n \geq 0$ .

This looks something like the  
 Pumping Lemma for regular  
 languages.



We need a fact about binary trees that you have seen before:

Lemma: A binary tree with height  $n$  has at most  $2^n$  leaves.

Proof: Simple induction on  $n$ .

So a binary tree with more than  $n$  leaves must have height greater than  $n$ , which means that it must have a path with more than  $n$  non-leaf nodes.

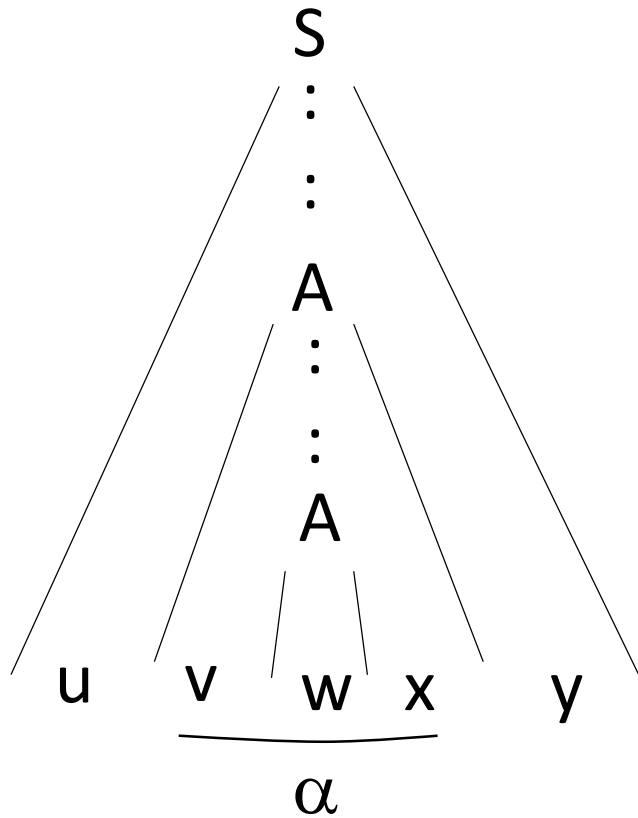
The Pumping Lemma for Context-Free Languages (1961 Bar-Hillel, Perles, Shamir): Let  $\mathcal{L}$  be a context-free language. Then there is a constant  $p$  so that if  $z$  is a string in  $\mathcal{L}$  and  $|z| \geq p$  there must be a decomposition  $z=uvwxy$  such that

- 1)  $|vwx| \leq p$
- 2)  $vx \neq \varepsilon$  (i.e., both  $v$  and  $x$  aren't  $\varepsilon$ )
- 3) For each  $i \geq 0$   $uv^iwx^iy$  is in  $\mathcal{L}$

Proof of the Pumping Lemma:

Let  $G$  be a CNF grammar for  $\mathcal{L} - \{\varepsilon\}$ . Let  $N$  be the number of nonterminal symbols of  $G$ . Let  $p = 2^N + 1$ . Let  $z$  be a string in  $\mathcal{L}$  with  $|z| \geq p$ . Since the parse tree for  $z$  must have more than  $2^N$  leaves, it must have height greater than  $N$  and so it must have a path with more than  $N$  nonterminal symbols. This path must have a repeated nonterminal; call this  $A$ .

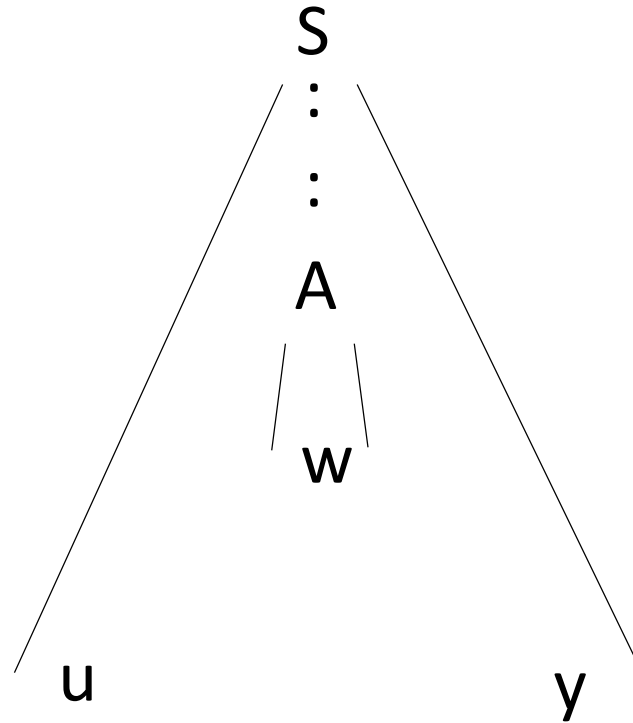
(continued next slide)



Let  $\alpha$  be the portion of  $z$  derived from the upper  $A$  symbol. Let  $u$  be the portion of  $z$  that precedes  $\alpha$ ,  $y$  the portion that follows  $\alpha$ . Similarly, let  $w$  be the portion of  $\alpha$  derived from the lower  $A$ ,  $v$  and  $x$  the portions of  $\alpha$  that precede and follow  $w$ . So  $z = uvwxy$ .

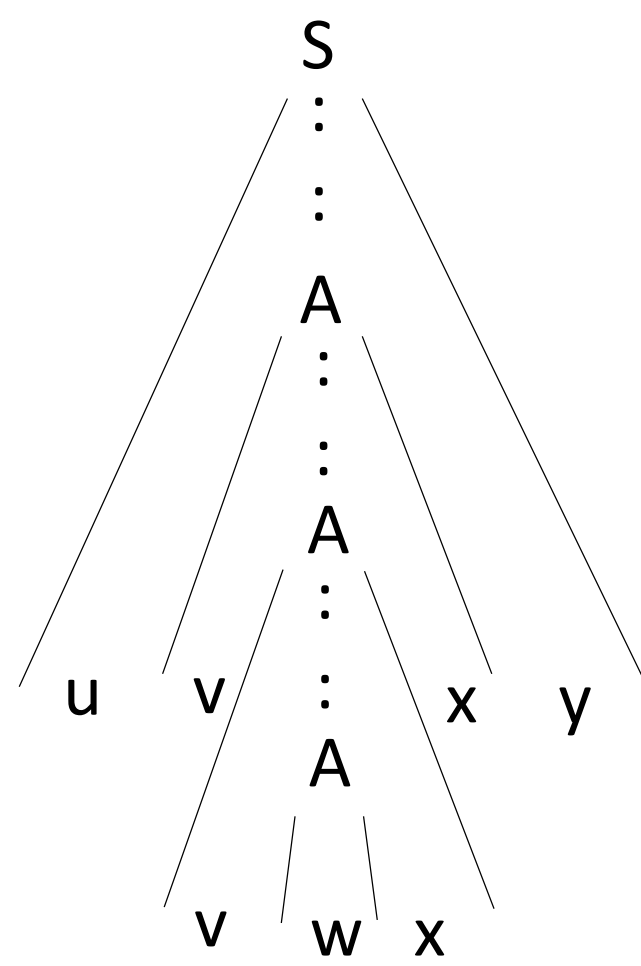
(Continued next slide)

We can pump 0 times by replacing the upper A with the lower one:

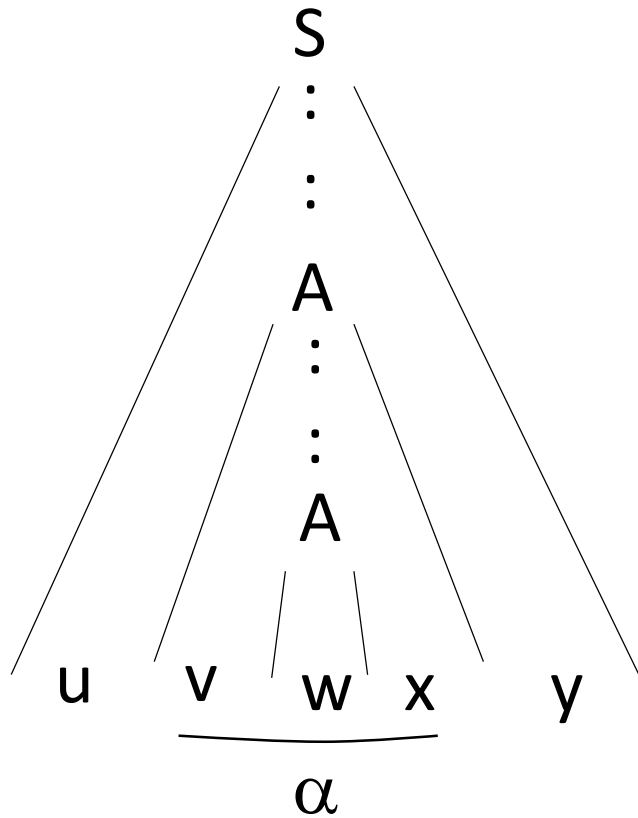




We can pump twice by replacing the lower A with a copy of the upper A:



And so forth. We can pump  $n$  times for every  $n \geq 0$ .



If we choose the upper A to be the lowest repeated nonterminal in the original parse tree for z, then the subtree rooted at A has height n or less so it has no more than  $2^N$  leaves. So  $|a| \leq 2^N < p$ .

Since we can't have  $A \overset{*}{\Rightarrow} A$ ,  $vx \neq \epsilon$ .

This completes the proof.

Example: Show  $\{1^n 2^n 3^n \mid n \geq 0\}$  is not context-free.

Suppose it is context-free; let  $p$  be its pumping constant. Let  $z = 1^p 2^p 3^p$ . Consider any decomposition  $z = uvwxy$  where  $|vwx| \leq p$ . Then  $vwx$  contains at most 2 of the digits  $\{1, 2, 3\}$ , so  $uv^2wx^2y$  cannot have the same numbers of all 3 digits. So  $z$  can't be pumped, contradicting the Pumping Lemma.

Example: Show that  $\{ww \mid w \in (0+1)^*\}$  is not context-free

Suppose it is context-free. Let  $p$  be its pumping constant. Let  $z=0^p1^p0^p1^p$ . Consider any decomposition  $z=uvwxy$  where  $|vwx| \leq p$ .  $vwx$  can't contain both 0's in the first half and 0's in the second, and it can't contain both 1's in the first half and 1's in the second half. Either way,  $uv^2wx^2y$  must have different numbers of 0's or different numbers of 1's between the two halves. So  $uv^2wx^2y$  is not in the language and  $z$  can't be pumped, contradicting the pumping lemma.

Example: Show that  $\{0^{n^2} \mid n \geq 0\} = \{\varepsilon, 0, 0^4, 0^9, 0^{16}, \dots\}$  is not context-free.

Suppose it is context-free; let  $p$  be its pumping constant. Let  $z = 0^{p^2}$

Consider any decomposition  $z=uvwxy$  where  $|vwx| \leq p$ . Then

$|uv^2wx^2y| \leq |z| + p \leq p^2 + p < (p+1)^2$ . So

$p^2 \leq |uv^2wx^2y| < (p+1)^2$ , so  $uv^2wx^2y$  is not in the language.

This means  $z$  is not pumpable, contradicting the pumping lemma.