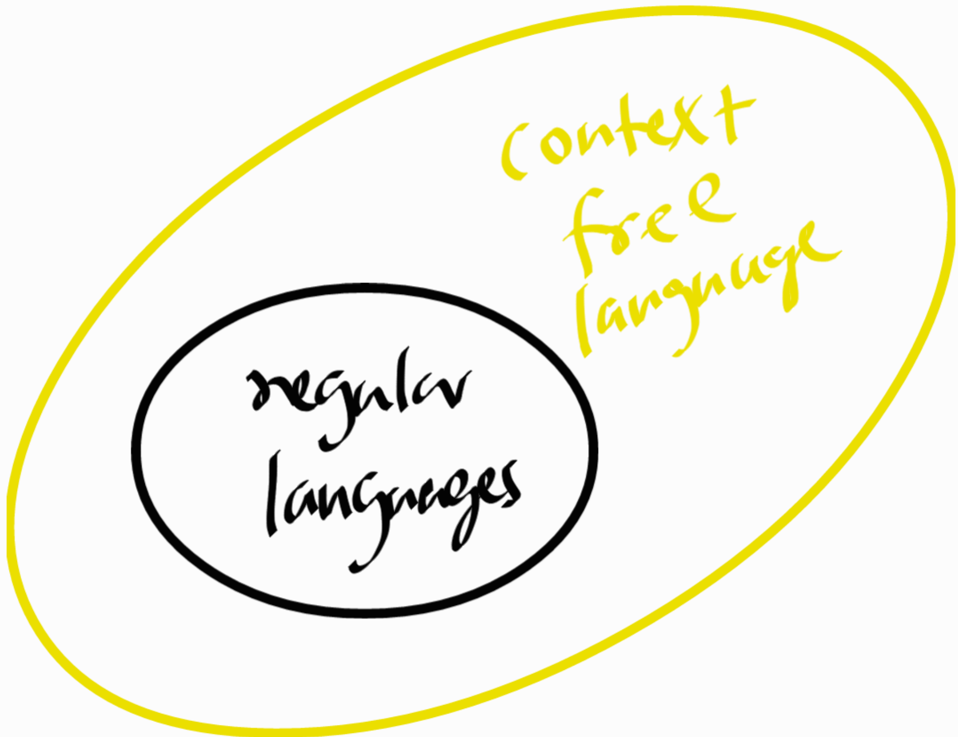
Pumping Lemma for CFL

* Every regular language is a context free language

- NFA is basically a PDA without the stack.



Now we are going look at how to prove a language is not-context free.

For CFL languages we have a slightly different pumping lemma.

* But the basic idea is very simillar.

Example for a non-context free language.

$$A = \{a^n b^n c^n \mid n \geq 0\} \qquad a^n b^n$$

This is not a context-free language.

First I will state the pumping lemma.

Pumping Lemma for context-free Languages.

If A is context-free language, then there is a number p (the pumping length) where, If s is a string in A of length at least p, then s may be divided into five pieces $s = uvxyz$ satisfying the conditions

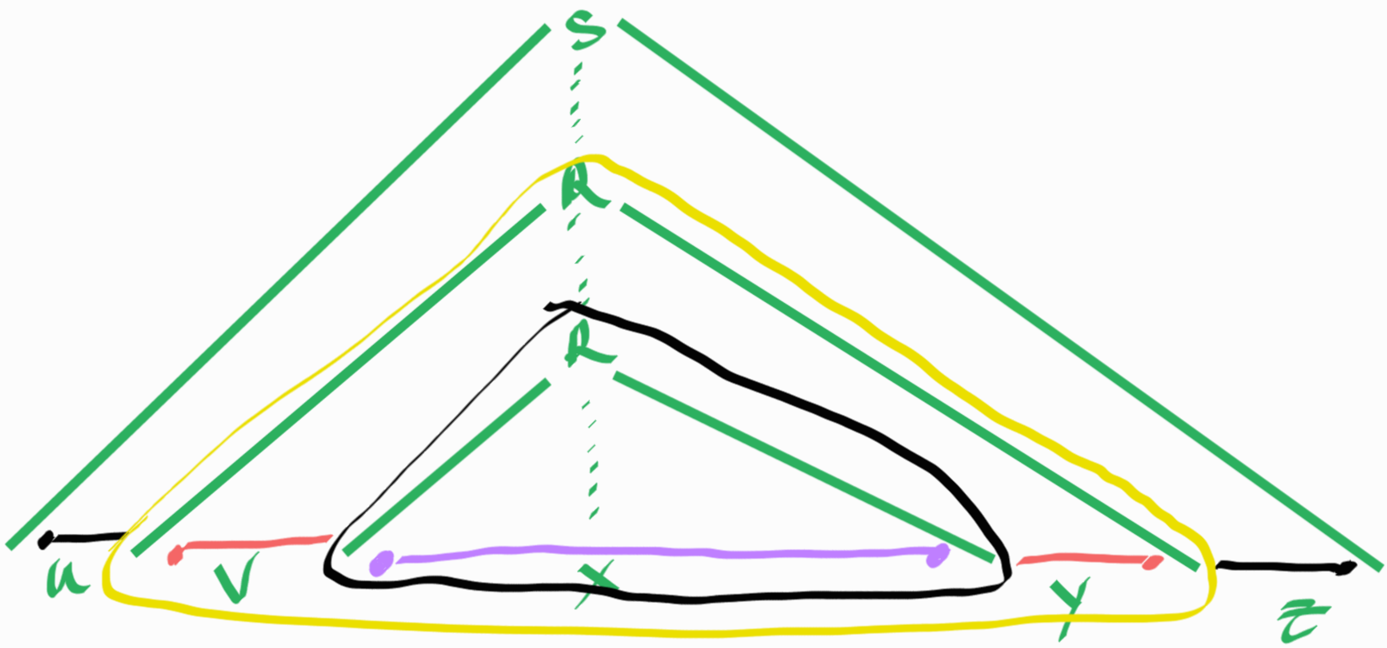
1. $uv^ixy^iz \in A$
2. $|vy| > 0$ either v or y has to be non-empty string
3. $|vxy| \leq p$

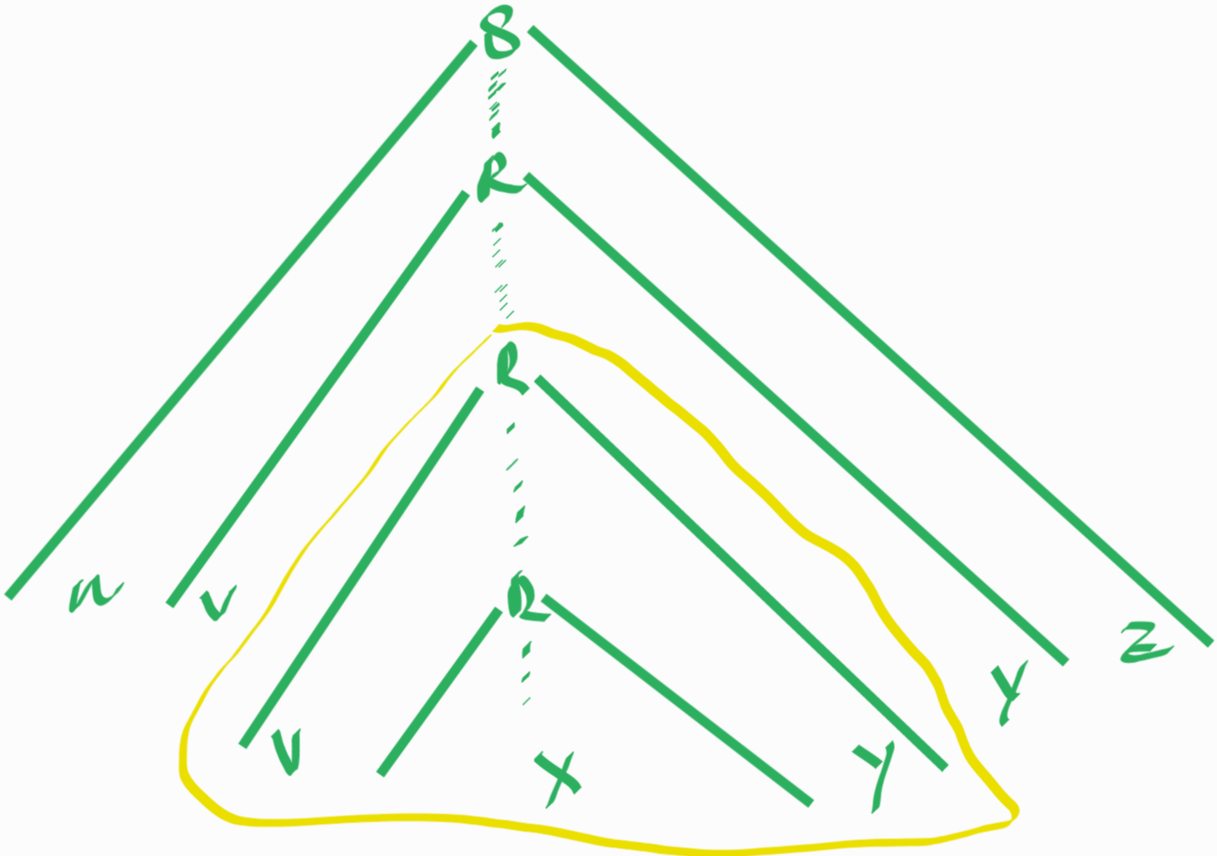
Proof idea: A is a CFL

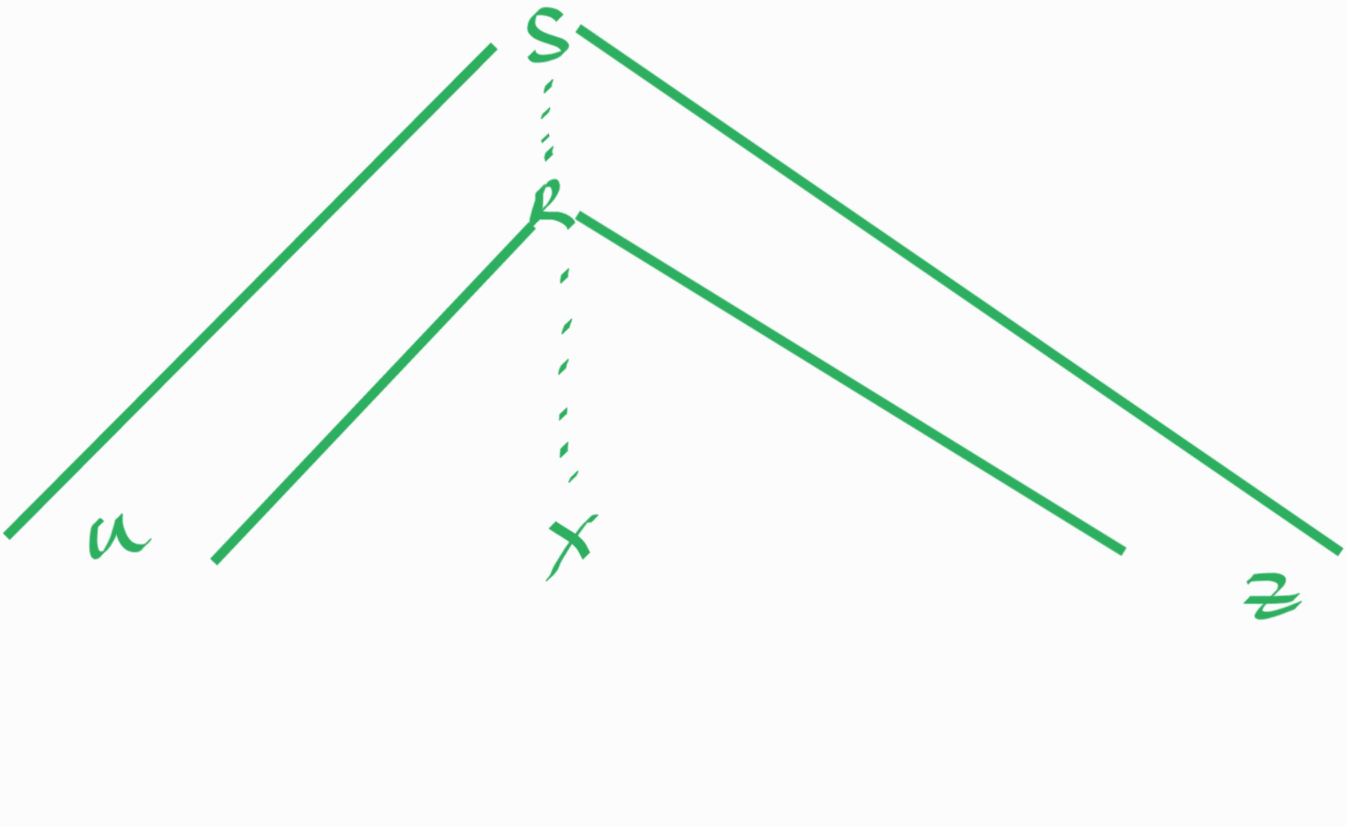
We pick a really large string s in A

G is the grammar of A

A $\rightarrow$ xyz




S
R
R
R
u
v
v
x
y
y
z


S
R
u
x
z

Proof:

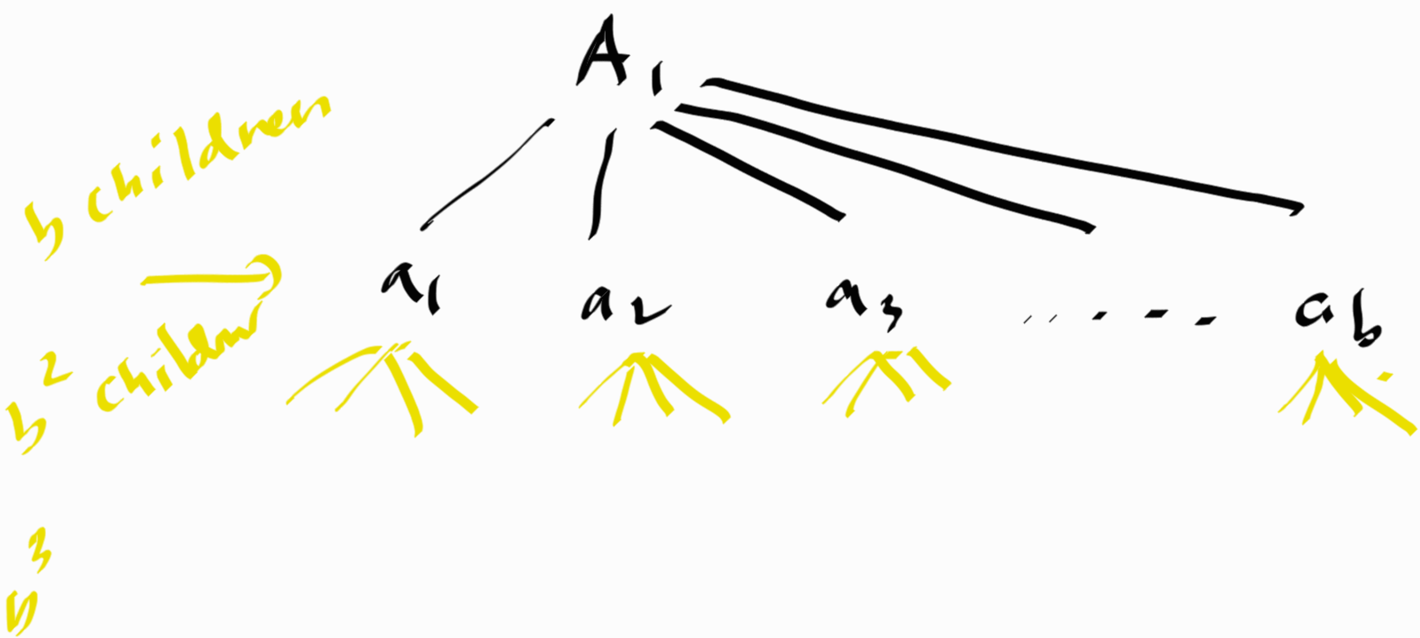
Suppose A is a CFL.

Suppose max size of any rule in the grammar G of A is b

$$S \rightarrow \underline{\hspace{3cm}}$$

$$A_1 \rightarrow a_1 a_2 a_3 \cdots a_b$$

This means in the worst case we can use this rule, and in the parse tree we could have b children.



If you use h steps in the generation process max length of string that you can produce is b^h

If the string s is length $b^h + 1$

Let # of variables in G is $|V|$, so let $p = b^{|V|} + 1$

suppose s is a string in A and $|s| \geq p$

Let τ be one of the parse trees of s. pick the parse tree that uses the minimum number of variables.

pick the longest path in τ from root to a leaf.

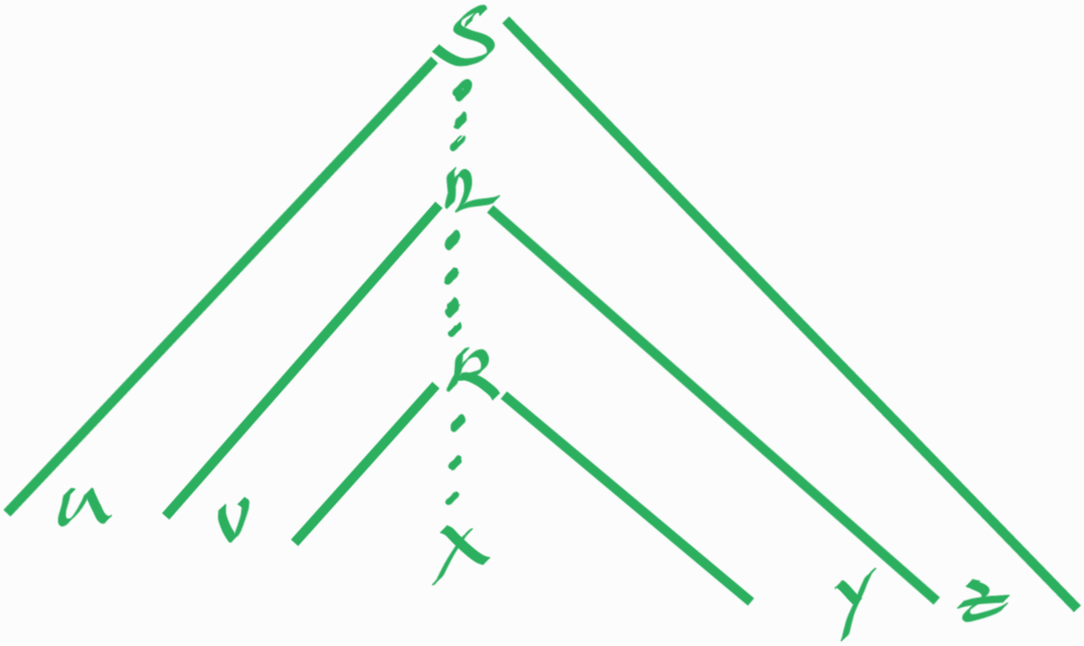
and it must be at least of length $|V|+1$

# of nodes in path $|V|+2$

This means in this path we must have used at least 1 variable twice.

Let R be that variable that is repeated at least twice.

using previous figure I can decompose s into $uvxyz$



we can substitute either the small subtree R or large subtree R to create larger parse trees.

But we pick the larger subtree to replace the smaller subtree

$u\ v \ldots v\ x\ y \ldots y\ z$

① $uv^i x y^i z \in A$

② $|vy| > 0$

This is true because we could get the string s by substituting the smaller subtree with larger subtree. But since we already picked the smallest parse tree for S this is not possible.

③ $|vxy| \leq p$

look at the bottom most |V|+1 nodes in the path. Note that one variable must be repeated (pigeon-hole principal). Then consider the two repeated variable in bottom most |V|+1 nodes. consider this |V|+1 nodes, then consider the repeated variable, Choose the uppermost subtree of this variable in the last |V|+1 nodes. If we use this subtree we can generate the largest vxy. but length of vxy is bounded by $b^{|V|+1}$ because the max height of the selected subtree is |V|+1.

$\therefore |vxy| \leq p$.

$B = \{ a^n b^n c^n \mid n \geq 0 \}$ is not context-free.