Lecture 20

Homework #20: 3.5.2 c, 3.5.2 d, 3.5.7, 3.5.8, 3.5.14 b, 3.5.14 a , c
[NOTE: For 3.5.8, use 3.5.7, but don't have to prove it.]

Today we turn to

Periodicity properties:

Like regular languages, CFLs display some periodicity (i.e., repetition).

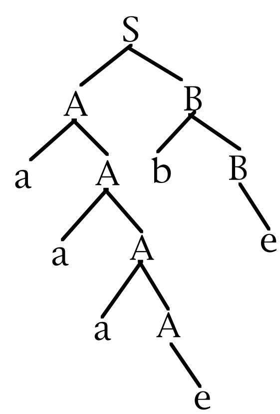
Let's look at an example in preparation for a Pumping Lemma for CFLs.

Consider a CFG with the following rules:

$S \rightarrow AB$

$A \rightarrow aA$ $\quad$ $A \rightarrow e$ $\quad$ $B \rightarrow bB$ $\quad$ $B \rightarrow e$

and now consider the following parse tree



Pumping Thm for CFLs: Let G be a CFG. Then there is a number K, depending on G, such that any string w in L(G) of length > K can be written as w = uvxyz such that either v or y is non-empty and uv^nxy^nz is in L(G) for every $n \geq 0$.

Pf. Let $G = (V, \Sigma, R, S)$

We'll show that there is a number K such that any string of terminals with length > K has a derivation of the form:

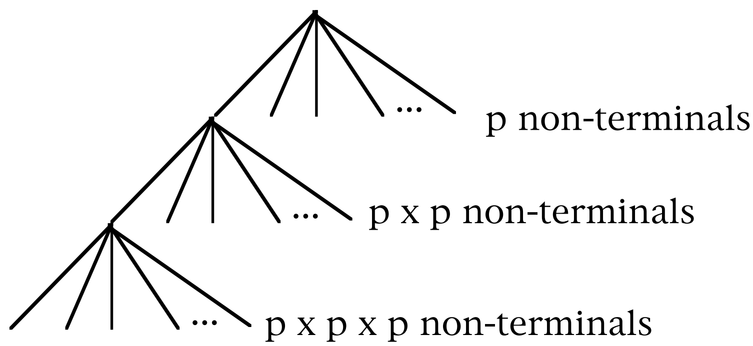
$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz$

$u, v, x, y, z \in \Sigma^*$
$A \in V\text{-}\Sigma$
v or y non-empty

and we can repeat $A \Rightarrow^* vAy$ multiple times to get v^nAy^n.

Let p = largest number of symbols on the right-hand side of a rule in R.

Now, a parse tree of height m can have at most p^m leaves.



So if a parse tree T has yield of length > p^m, then T has some path of length > m.

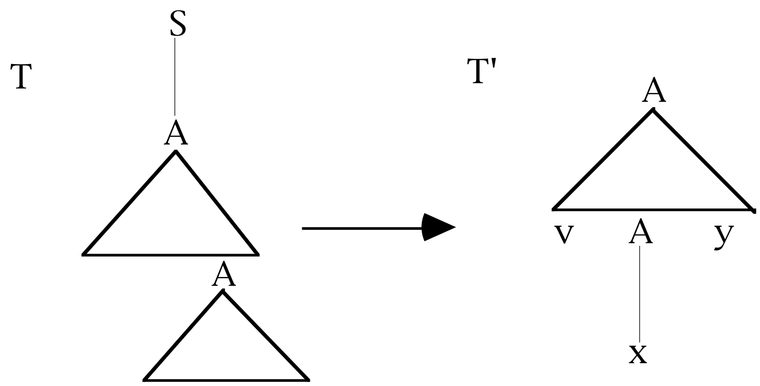
Now, let $m = |V - \Sigma|$ (the number of non-terminals)
and let $K = p^m$

Now suppose w has length > p^m (i.e., K)

And let T be its parse tree.

T has at least 1 path with number of nodes > $|V - \Sigma| + 1$.
So at least that 1 path has 2 nodes labeled with the same non-terminal.

Follow the path being discussed up toward the root of the tree, stopping at the second instance of the repeated non-terminal. Now consider the subtree of T rooted at this node. Call the subtree T'.



Both v and y can't be e. (that is, can't have v = y = e).
if so, you could just remove the part of the path from A to A, with no effect on the yield of the tree. If you removed all such segments you'd get a tree with yield w and height < m.

Now, this defines what u, v, w, x, y are.
Clearly, you can repeat the A-rooted subtree any number of times, including 0 - that is, by "pulling up" the 2nd A-rooted subtree.

Note that the number p we've discussed here is called the **fanout** of G and is denoted $\phi(G)$.

We do not use the Pumping Theorem to show that a language is context free. We do, however, use it to show that a language is not (by showing that the languages violate the theorem).

Classic example. $\{a^n b^n c^n : n \geq 0\}$ is not context free.

Pf. Suppose $L = \{a^n b^n c^n : n \geq 0\}$ is context free.
Then $L = L(G)$ for $G = (V, \Sigma, R, S)$

Let K be the constant specified by the pumping thm, and let $j > K/3$.

Then $w = a^j b^j c^j \in L(G)$ and $|w| > K$.

So w can be written as uvxyz such that v or $y \neq e$, and $uv^i xy^i z \in L(G)$ for all $i \geq 0$.

Now consider the possibilities for v and y:

I. v or y contains 2 different symbols from {a, b, c}. Then they will alternate when pumped, producing a string not in the language.

II. v and/or y contain all a's or all b's or all c's. Pumping on any one or two of the 3 symbols leads to an imbalance.