CS 205
Lecture # 25
16/3/21

Pumping Lemma for CFLs

If L is a CFL then $\exists n$ s.t. $\forall z \in L$, if $|z| \geqslant n$ then $\exists\, u, v, w, x, y$ s.t.

1. $z = uvwxy$
2. $|vwx| \leq n$
3. $|vx| > 0$
4. $\forall i \geqslant 0 .\ uv^i wx^i y \in L$.

Proof: Let L be a CFL and G be a CFG in CNF s.t. $L(G) = L \setminus \{\varepsilon\}$.

Recall: CNF has prodns. are of type $A \to a$ or $A \to BC$, where $A, B, C \in V$, $a \in T$.

Suppose G has K variables.
Take $n = 2^K$.
Suppose $z \in L$ and $|z| \geqslant n = 2^K$.

$\boxed{\text{Claim: A derivation tree for } z \text{ has a path of length at least } K+1.}$

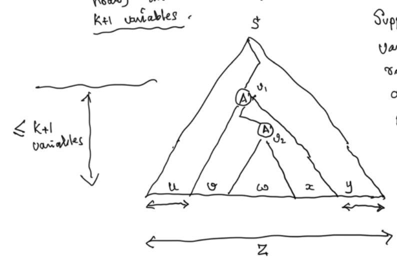
Why? : Since G is in CNF, any derivation tree for z is a binary tree.
Suppose the maximum path length $\leq K$, i.e. the height of the tree $\leq K$.

$\boxed{\text{Fact: The no. of leaves in a binary parse tree for CFG } G \text{ of height } h \text{ is at most } 2^{h-1}.}$

By induction: Base Case $A \to a$; Inductive Case $A \to BC$

So $|z| = \#$ leaves in derivation tree $\leq 2^{K-1}$.
But $|z| \geqslant 2^K = n$ by assumption. Contradiction.
Hence the claim is true.

Now, a path of length $\geqslant K+1$ has $\geqslant K+2$ nodes out of which the last one is a leaf, i.e. a terminal. Thus the path has at least $K+1$ variables (i.e. internal nodes) and so one of them must repeat in the last $K+1$ variables.



Suppose A is a variable which is repeated on such a path.
Here
$A \overset{*}{\Rightarrow} w$
$A \overset{*}{\Rightarrow} vwx$
$S \overset{*}{\Rightarrow} uvwxy = z$

We have $|vwx| \leq n = 2^K$ since the height of the tree rooted at v_1 is $\leq K+1$.
Further $|vx| \geqslant 1$ because v_1 and v_2 are different nodes and the grammar has no ε-productions.

In addition, we have
$S \overset{*}{\Rightarrow} uAy$ (don't expand v_1)
$A \overset{*}{\Rightarrow} vAx$ (start at node v_1)
$A \overset{*}{\Rightarrow} w$ (start at " v_2)

$i = 0$: $S \overset{*}{\Rightarrow} uAy \overset{*}{\Rightarrow} uwy$

Therefore, $A \overset{*}{\Rightarrow} vAx \overset{*}{\Rightarrow} vvAxx \overset{*}{\Rightarrow} v^iAx^i$ for all $i \geqslant 0$

Hence $S \overset{*}{\Rightarrow} uv^iwx^iy \in L$ for all $i \geqslant 0$.

Ex

1. $L = \{a^m b^m c^m \mid m \geqslant 0\}$ is not a CFL.

Proof: Suppose L is a CFL. Then by the PL there is a constant n s.t. ...

Consider $\boxed{z = a^n b^n c^n}$. Since $z \in L$ and $|z| > n$ there are u, v, w, x, y s.t. $z = uvwxy$, $|vwx| \leq n$, $|vx| > 0$ and $uv^iwx^iy \in L$ for all i.

Consider any such u, v, w, x, y. Since $|vwx| \leq n$, vwx cannot contain all three of the symbols a, b, c, since there are n b's.

Then $uv^0wx^0y = uwy$ contains too many symbols of the type that is not present in vwx. Hence $uwy \notin L$. Contradiction.

Ex 2 $L = \{a^i b^j c^i d^j \mid i, j \geqslant 0\}$ is not a CFL.