CSE 322

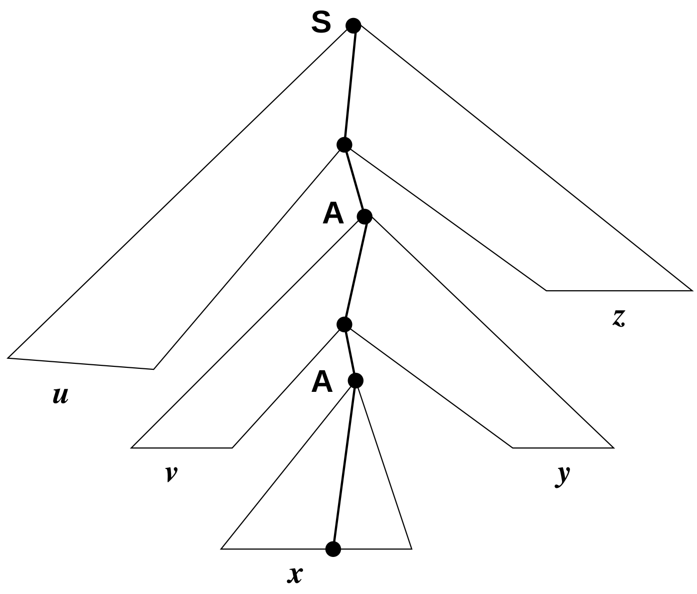
Introduction to Formal Models in Computer Science

Pumping Lemma for CFL's

CFL Pumping Lemma: If L is a context-free language then there is a constant n (depending on L) such that for all $w \in L$ with $|w| \geq n$, w can be written as $w = uvxyz$ such that

1. $|vxy| \leq n$,
2. $|v| + |y| \geq 1$, and
3. for all $i \geq 0$, $uv^ixy^iz \in L$.

Proof: Let $G = (V, \Sigma, R, S)$ be a context-free grammar generating L. Let ℓ be the length of the longest right side of any rule in R. Let $k = |V - \Sigma|$, $n = \ell^{k+1}$ and suppose that $w \in L$ with $|w| \geq n$. Consider the smallest parse tree T_w for w. By the choice of ℓ we know that the fan-out of any node in T_w is at most ℓ. Now any tree of fan-out ℓ and height h has at most ℓ^h leaves. Therefore we know that T_w has height at least $k+1$ and therefore has some root-leaf path with at least $k+1$ interior vertices. Consider the longest such path and consider the labels of the lowest $k+1$ interior vertices on this path. All these labels are non-terminals and since $|V - \Sigma| = k$, one of these labels must appear twice. Call this non-terminal A. Thus we can break up T_w and choose u, v, x, y, and z as in the following figure:



We now need to argue that (1), (2), (3) hold.
(1) is true because the higher A generates vwx and this A has height at most $k+1$ and therefore the subtree below it has at most $\ell^{k+1} = n$ leaves.
(2) says that v and y are not both e. This must be true because T_w was chosen to be minimal to begin with. If $v = y = e$ then we can get a smaller tree by replacing the subtree below the higher A by the subtree below the second one.
To show (3) we note that

$$S \overset{*}{\underset{G}{\Rightarrow}} uAz \overset{*}{\underset{G}{\Rightarrow}} uvAyz \overset{*}{\underset{G}{\Rightarrow}} uvxyz = w$$

and this implies that (a) $A \overset{*}{\underset{G}{\Rightarrow}} vAy$ and (b) $A \overset{*}{\underset{G}{\Rightarrow}} x$. We first show that for all $i \geq 0$, $S \overset{*}{\underset{G}{\Rightarrow}} uv^iAy^iz$ by induction on i and then conclude that $S \overset{*}{\underset{G}{\Rightarrow}} uv^iwy^iz$ using derivation (b).

Base Case: $S \overset{*}{\underset{G}{\Rightarrow}} uAz = uv^0Ay^0z$.

Induction Step: Suppose that $S \overset{*}{\underset{G}{\Rightarrow}} uv^iAy^iz$. Then $S \overset{*}{\underset{G}{\Rightarrow}} uv^iAy^iz \overset{*}{\underset{G}{\Rightarrow}} uv^ivAyy^iz = uv^{i+1}Ay^{i+1}z$ using derivation (a).

As a result we have $uv^iwy^iz \in L$ for all $i \geq 0$ and (3) is proved. $\square$

Slightly Stronger CFL Pumping Lemma: If L is a context-free language then there is a constant n (depending on L) such that for all $w \in L$ with $|w| \geq n$, w can be written as $w = uvxyz$ such that

1. $|vxy| \leq n$,
2. $v \neq e$, and
3. for all $i \geq 0$, $uv^ixy^iz \in L$.

Proof: The only difference here is that the condition $|v|+|y| \geq 1$ is replaced by $v \neq e$. Apply the CFL Pumping Lemma above to get $w = uvxyz$ such that $|vxy| \leq n$, $|v| + |y| \geq 1$ and v and y can be jointly pumped. If $v \neq e$ we are done. Otherwise if $v = e$, then $w = uxyz$ and pumping v and y jointly implies that $uxy^iz \in L$ for all $i \geq 0$. In this case we set $u_{new} = ux$, $v_{new} = y \neq e$, $x_{new} = e$, $y_{new} = e$, and $z_{new} = z$. It is easy to check now that all the conditions are satisfied for these new values. $\square$