Lecture 20

$A \subseteq \Sigma^*$ is a CFL if $\exists$ a CFG $G = (N, \Sigma, P, S)$ s.t $L(G) = A$.

Question. For all $A \subseteq \Sigma^*$, is A a CFL?
↳ is A a regular set? No - $\{a^n b^n \mid n \geq 0\}$

Observation. For a grammar in Chomsky normal form, any parse tree for a long string should have a long path.
Any long path should have at least two occurences of some nonterminal symbol.

For a grammar in CNF - the number of symbols can atmost double going down a level in the parse tree - RHS of each production contains atmost 2 symbols.

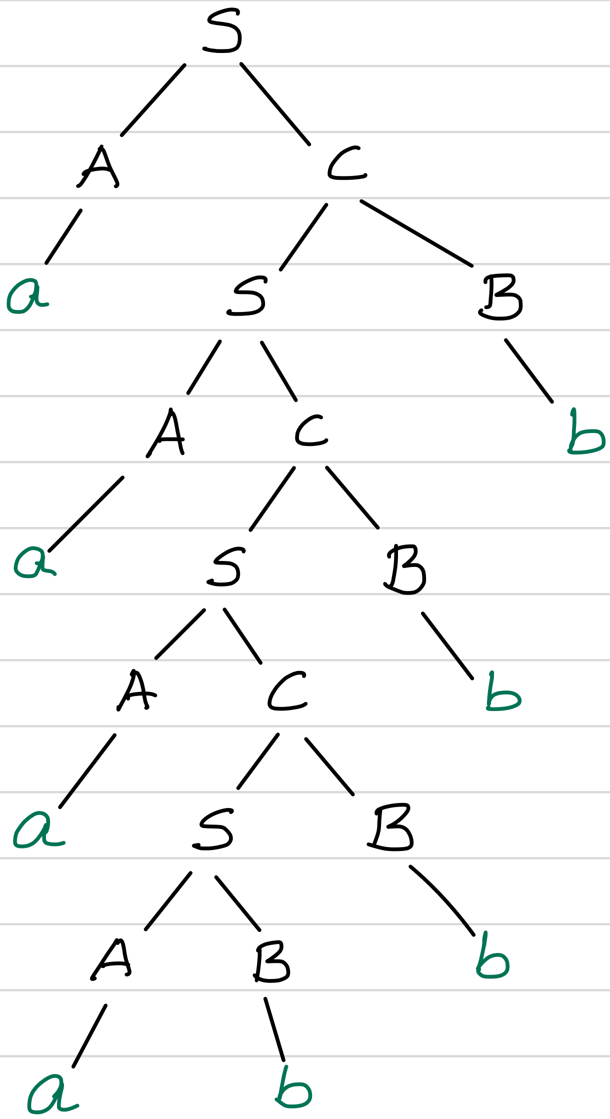
We have 1 symbol at level 0
atmost 2 symbols at level 1.
⋮
2^i symbols at level i.

To have 2^n symbols at the bottom level, the tree must be of depth atleast n - it must have atleast $n+1$ levels.
Depth - number of edges in the longest path from the root to a leaf node.

G: $S \to AC \mid AB$, $A \to a$, $B \to b$, $C \to SB$.

CNF for $\{a^n b^n \mid n \geq 1\}$

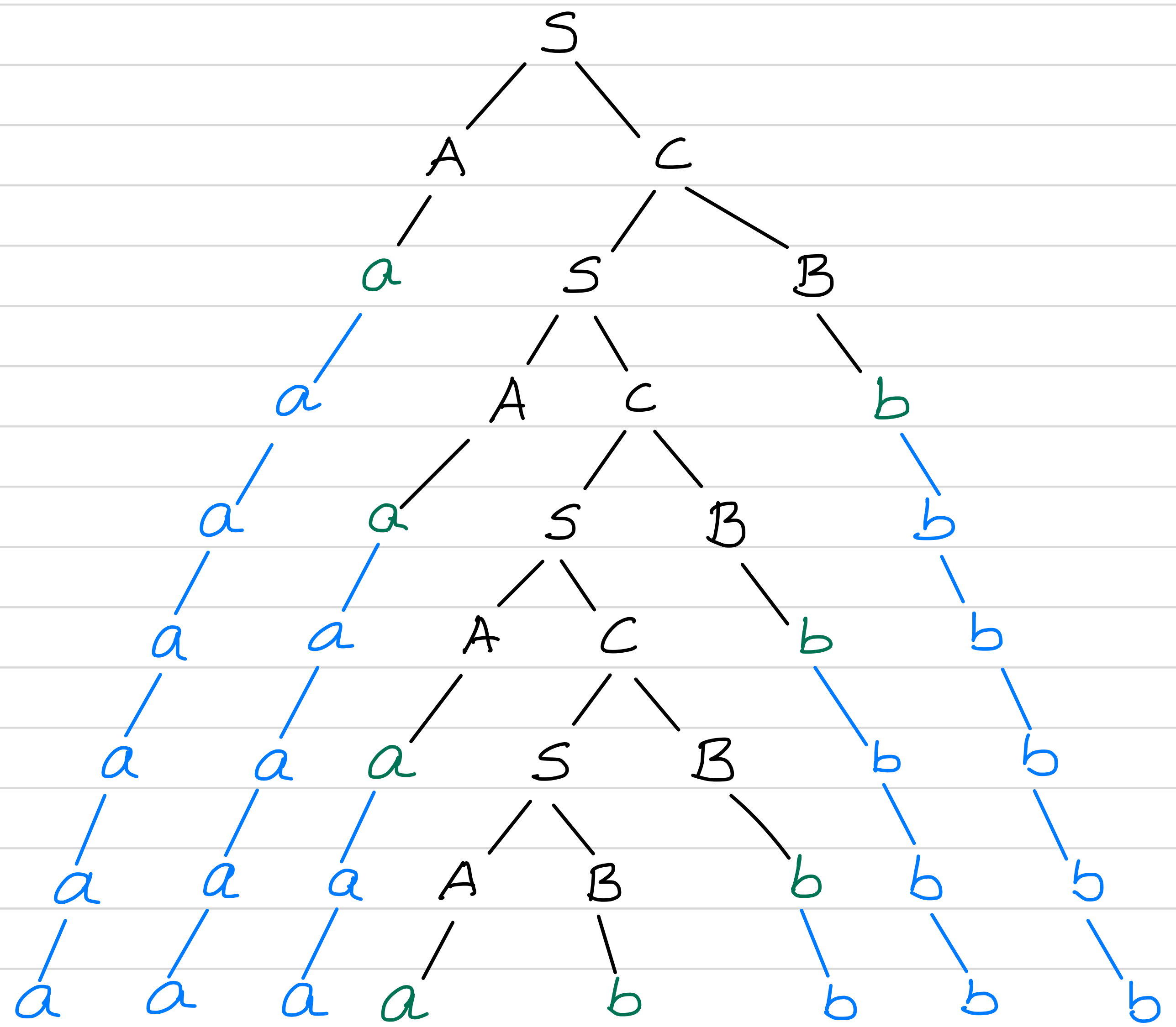
Consider the derivation of $a^4 b^4$.



G: $S \to AC \mid AB$, $A \to a$, $B \to b$, $C \to SB$.

CNF for $\{a^n b^n \mid n \geq 1\}$

Consider the derivation of $a^4 b^4$.



Pumping Lemma for CFLs.

if $A \subseteq \Sigma^*$ is a CFL then there exist $k \geq 0$ such that for every $z \in A$ of length at least k can be split into five substrings $z = u\,v\,w\,x\,y$ such that $vx \neq \epsilon$, $|vwx| \leq k$ and for all $i \geq 0$, $u\,v^i\,w\,x^i\,y \in A$.

Proof.

Let G be a grammar for A in CNF.

Take $k = 2^{n+1}$, where n is the number of nonterminals of G.

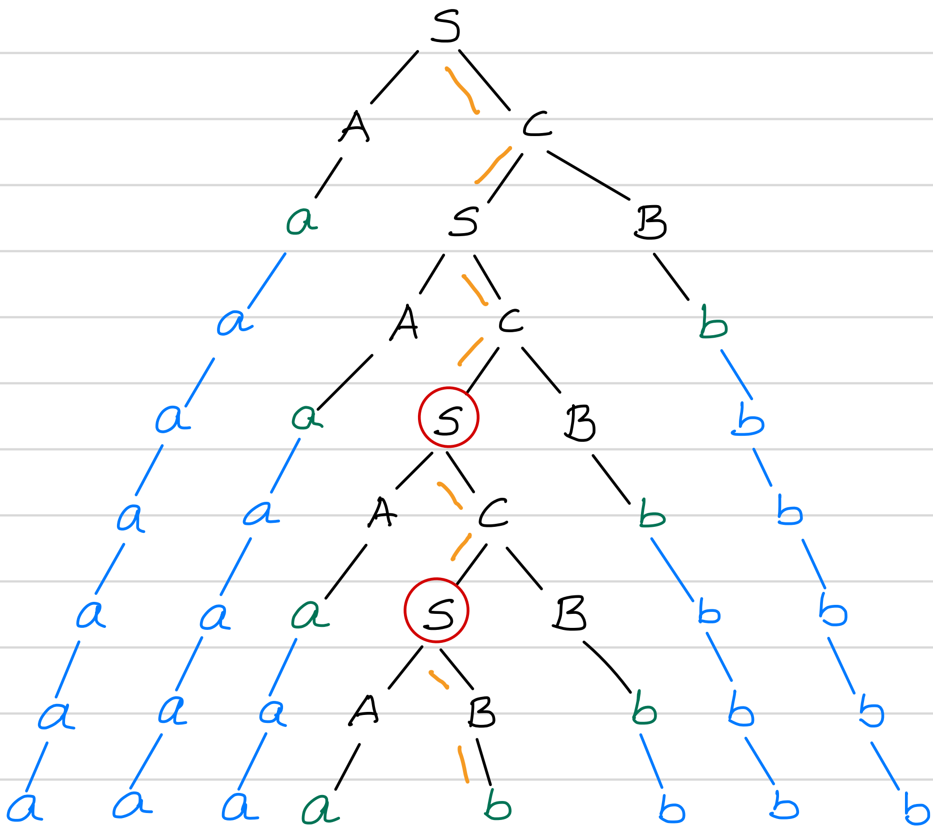
Suppose $z \in A$ and $|z| \geq k$.

Any parse tree in G for z must be of depth at least $n+1$ (i.e. there are $n+2$ levels).

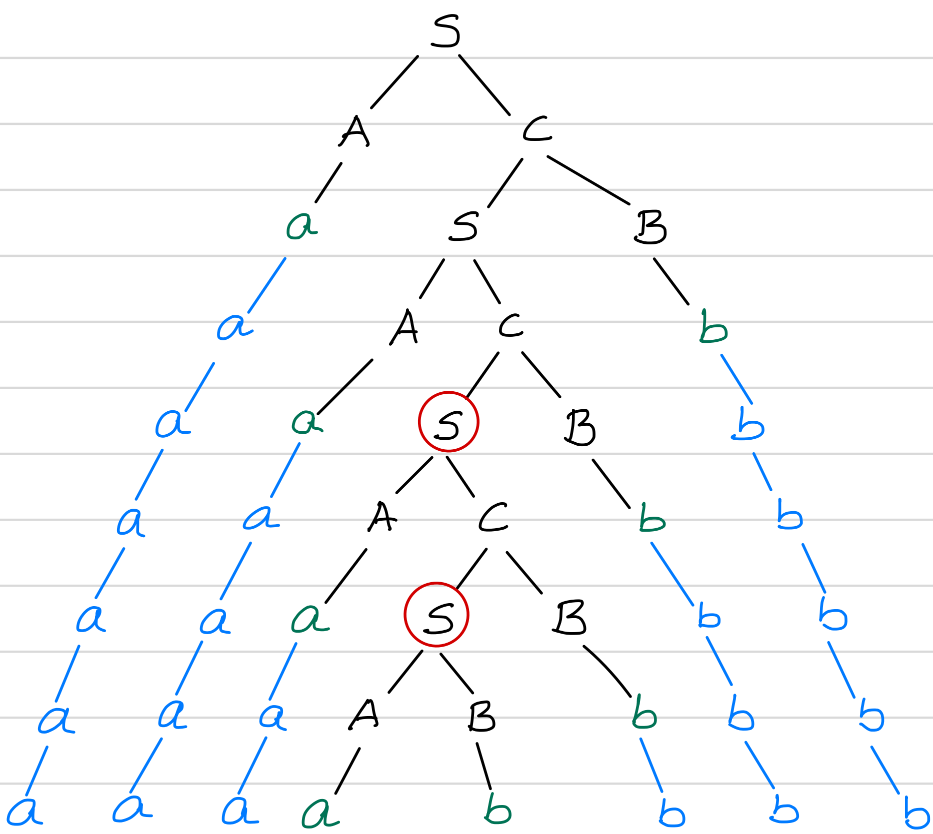
Consider the longest path in the tree (it is of length atleast $n+1$).

The longest path contains atleast $n+1$ occurence of nonterminals.

This implies: Some nonterminal occurs more than once.



Take the first pair of occurences of the same nonterminal along the path - traversing bottom to top.

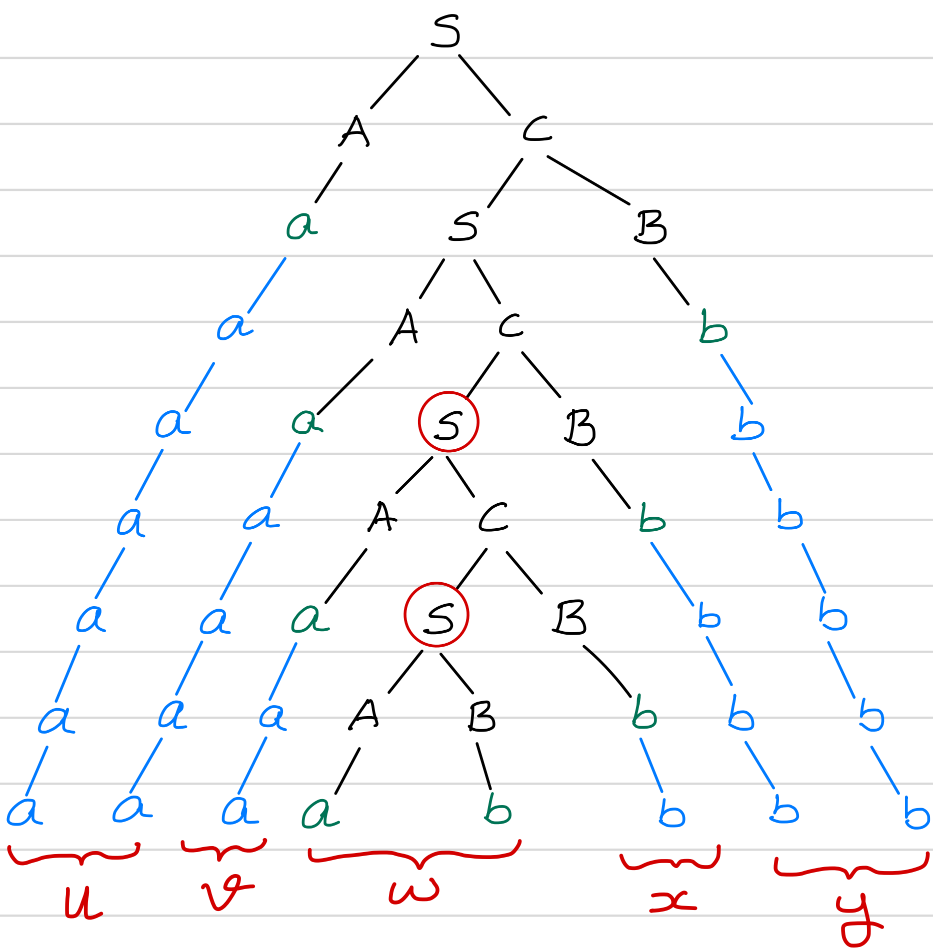


Take the first pair of occurences of the same nonterminal along the path - traversing bottom to top.

Suppose X is the nonterminal with two occurences. Split $z = uvwxy$ such that.

w - string (of terminals) generated by lower occurence of X

& vwx - string generated by the upper occurence of X

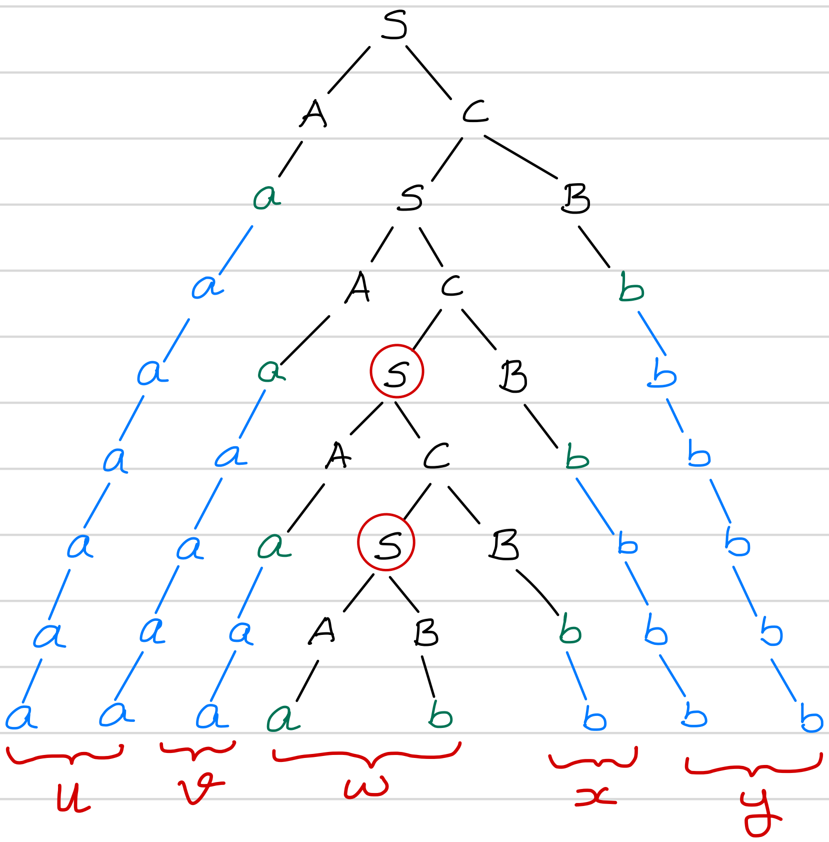


Take the first pair of occurences of the same nonterminal along the path - traversing bottom to top.

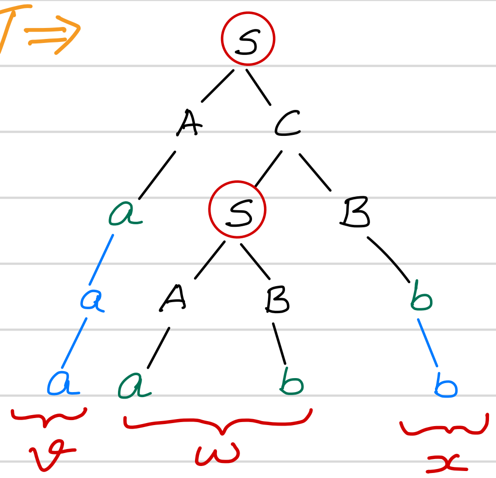
Suppose X is the nonterminal with two occurences. Split $z = uvwxy$ such that.

w - string (of terminals) generated by lower occurence of X
vwx - string generated by the upper occurence of X

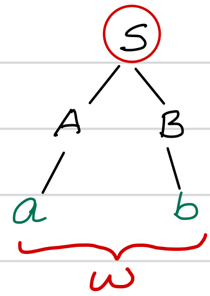
Let T - subtree rooted at upper occurence of X.
t - subtree rooted at the lower occurence of X



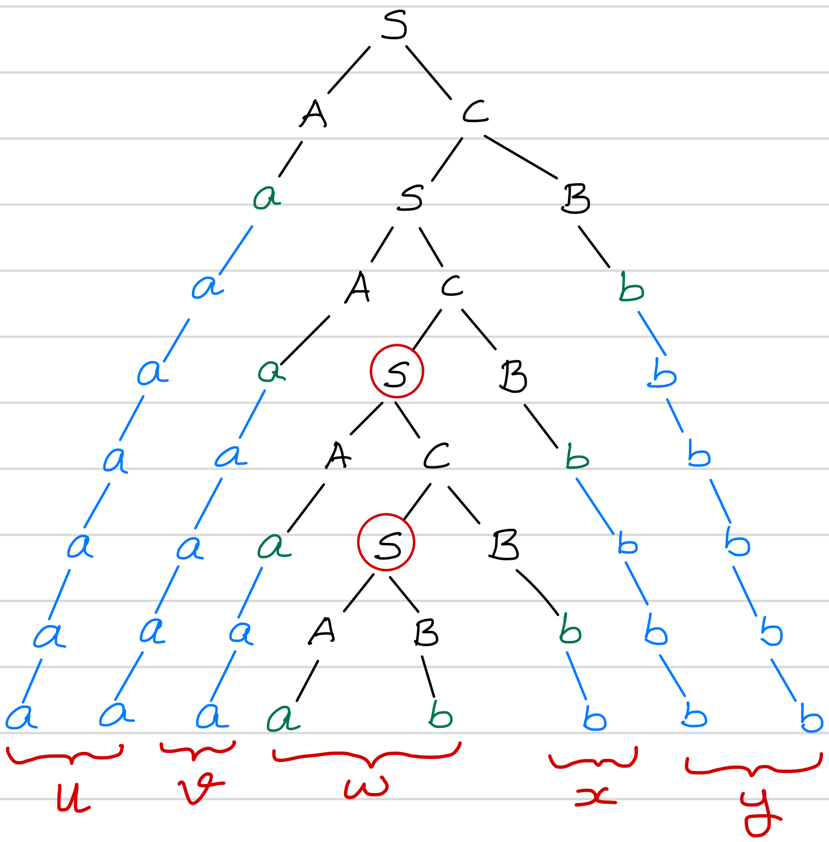
$T \Rightarrow$



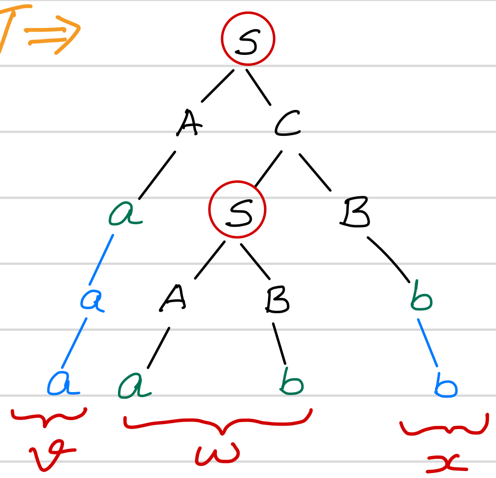
$t \Rightarrow$



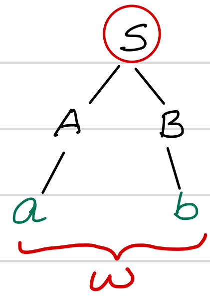
Consider the string uv^2wx^2y

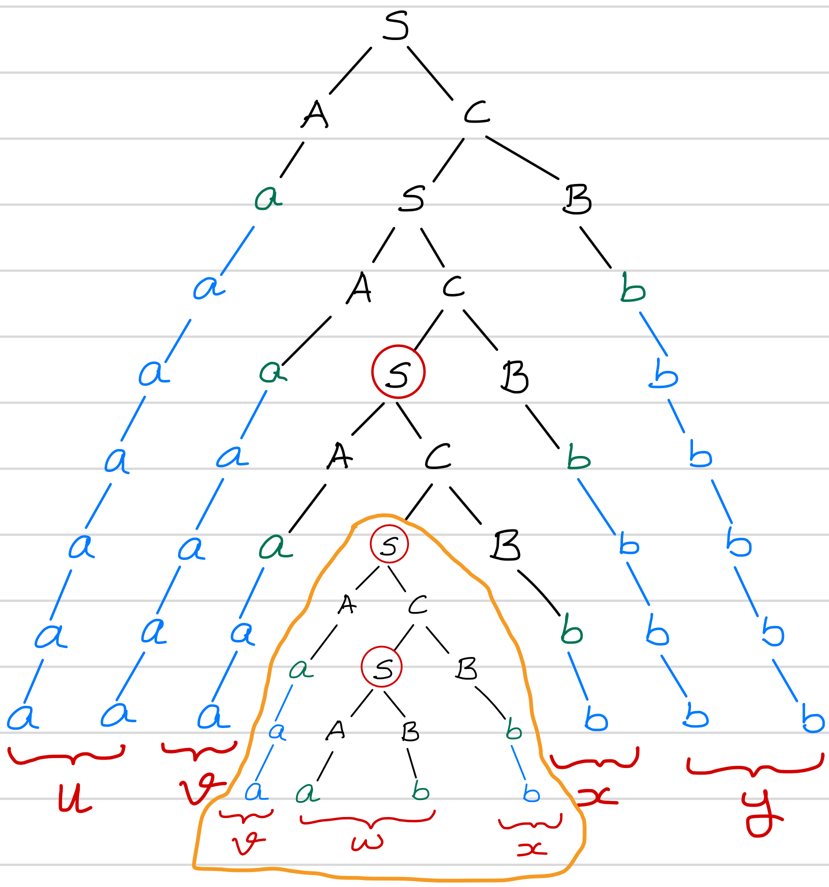


$T \Rightarrow$



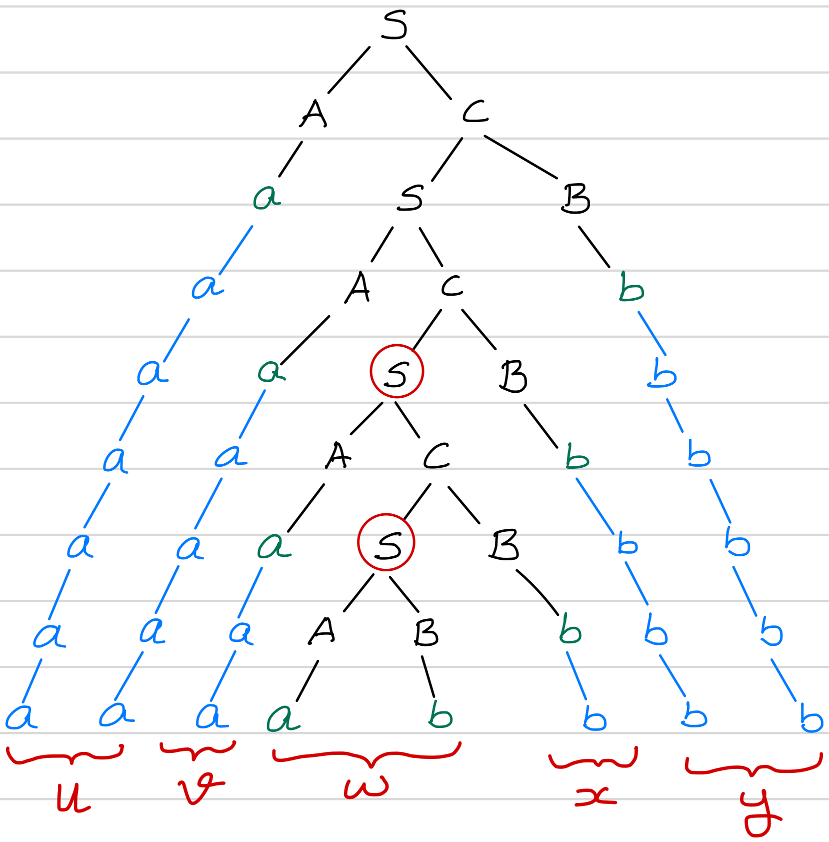
$t \Rightarrow$

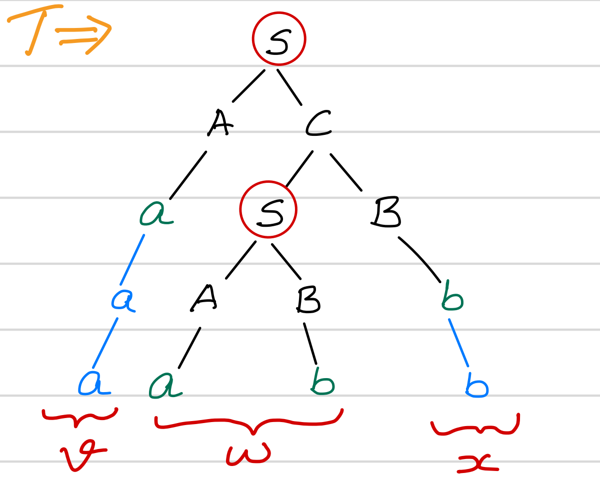


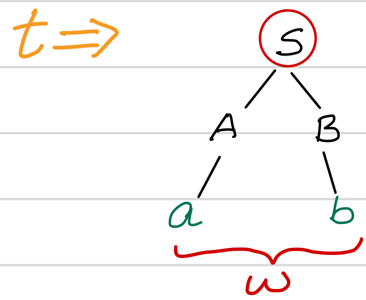


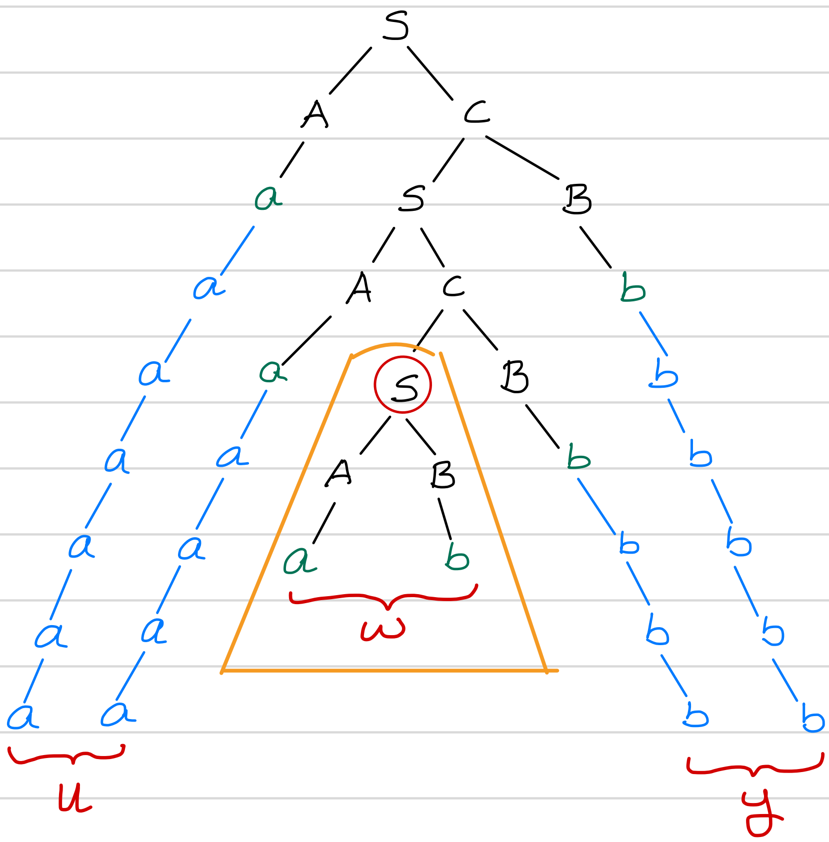
Consider the string uv^2wx^2y

We can replace t with as many copies of T to get a parse tree for uv^iwx^iy for all $i \geq 1$.





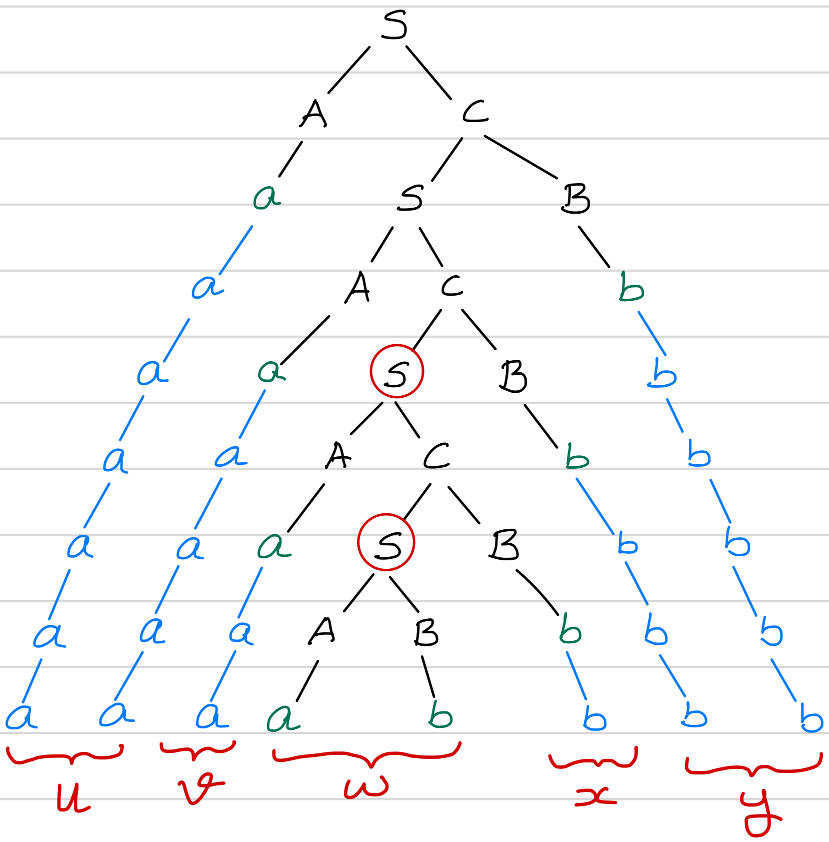




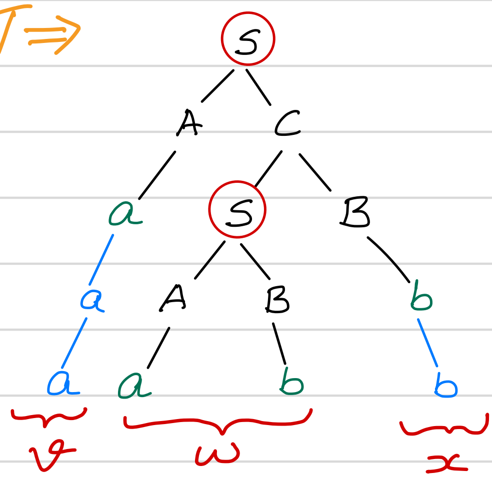
Consider the string uv^2wx^2y

Note: $vx \neq \epsilon$

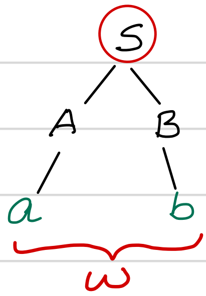
We can cut T and replace it with t to get a parse tree for $uv^0wx^0y = uwy$



T ⇒



t ⇒



Note.

1. $vx \neq \epsilon$. — v and x are not both ϵ.

2. $|vwx| \leq k$ - Since we chose the first repeated occurence of a nonterminal from the bottom.
   Depth of the subtree under the upper occurence of the repeated nonterminal X is atmost $n+1$

$\therefore$ it can have atmost $2^{n+1} = k$ terminals $\square$

To show that a set is not a CFL - use pumping lemma in the contrapositive form.

For all $k \geq 0$, $\exists z \in A$ s.t $|z| \geq k$ and for all split of z into substrings $z = uvwxy$ with $vx \neq \epsilon$ and $|vwx| \leq k$, there exists an $i \geq 0$ s.t $uv^iwx^iy \notin A$

To show that a set is not a CFL - use pumping lemma in the contrapositive form.

For all $k \geq 0$, $\exists z \in A$ s.t $|z| \geq k$ and for all split of z into substrings $z = uvwxy$ with $vx \neq \epsilon$ and $|vwx| \leq k$, there exists an $i \geq 0$ s.t $uv^i w x^i y \notin A$

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

Proof. Given k, let $z = a^k b^k a^k$. We have $z \in A$, $|z| = 3k$

Now consider any split $z = uvwxy$, $vx \neq \epsilon$ and $|vwx| \leq k$.
Let $i = 2$. Consider the string uv^2wx^2y.

Case 1. v or x contains atleast one "a" and at least one "b". then uv^2wx^2y is not of the form $a^* b^* a^*$

Case 2. v and x contains only a's. Then uv^2wx^2y has more a's than b's.

Case 3. v and x contains only b's. Then the number of b's is greater than the number of a's.

Case 4. One of v or x contains only a's and the other only b's. Then uv^2wx^2y is not of the form $a^m b^m a^m$.

$\therefore$ in all cases, the resulting string $uv^2wx^2y \notin A$.
By pumping lemma A is not a CFL.

Closure properties of CFLs.

Union. Suppose A and B are CFLs where $L(G_1)=A$, $L(G_2)=B$ and the start symbols are S_1 for G_1 and S_2 for G_2.

Construct a grammar G s.t $L(G)= A \cup B$ as follows:

Ensure that G_1 and G_2 have disjoint set of nonterminals [rename the nonterminals if required]

- Combine the productions of G_1 & G_2.
- Add a new start symbol S and the productions: $S \rightarrow S_1$, $S \rightarrow S_2$

$S \rightarrow S_1 \mid S_2$.

Closure properties of CFLs.

Union. Suppose A and B are CFLs where
$L(G_1)=A$, $L(G_2)=A$ and the start symbols are
S_1 for G_1 and S_2 for G_2.

Construct a grammar G s.t $L(G) = A \cup B$ as follows:

Ensure that G_1 and G_2 have disjoint set
[rename the nonterminals if required]
Combine the productions of G_1 & G_2.
Add a new start symbol S and the productions: $S \to S_1$, $S \to S_2$

Concatenation. if A and B are CFLs with $L(G_1)=A$
and $L(G_2)=B$ construct G s.t
$L(G) = AB = \{xy \mid x \in A, y \in B\}$ as follows:

- Combine the grammars G_1 & G_2.

- Add a new start symbol S with production $S \to S_1S_2$

Closure properties of CFLs.

Union. Suppose A and B are CFLs where $L(G_1)=A$, $L(G_2)=A$ and the start symbols are S_1 for G_1 and S_2 for G_2.

Construct a grammar G s.t $L(G) = A \cup B$ as follows:

Ensure that G_1 and G_2 have disjoint set [rename the nonterminals if required]

Combine the productions of G_1 & G_2.

Add a new start symbol S and the productions: $S \to S_1$, $S \to S_2$

Concatenation. if A and B are CFLs with $L(G_1)=A$ and $L(G_2)=B$ with start symbols S_1 & S_2. Construct G such that $L(G) = AB = \{xy \mid x \in A, y \in B\}$ as follows:

- Combine the grammars G_1 & G_2.
- Add a new start symbol S with production $S \to S_1 S_2$

Kleene star. if A is a CFL with $L(G_1)=A$ and start symbol S_1, Construct G s.t $L(G) = A^*$ as follows:

- Take G_1 along with a new start symbol S along with the production

$$S \to S_1 S \mid \epsilon.$$

Intersection. CFLs are not closed under intersection.

$$\{a^n b^m c^n \mid m,n \geq 0\} \cap \{a^m b^n c^n \mid m,n \geq 0\}$$

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

we know that this set is not a CFL.