# Pumping Lemma for Context-Free Languages

Deepak D'Souza

Department of Computer Science and Automation
Indian Institute of Science, Bangalore.

15 February 2024

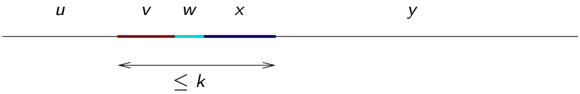
# Outline

# Pumping Lemma for CFL's

## Pumping Lemma

For every CFL $L$ there is a constant $k \geq 0$ such that for any word $z$ in $L$ of length at least $k$, there are strings $u, v, w, x, y$ such that

- $z = uvwxy$,
- $vx \neq \epsilon$,
- $|vwx| \leq k$, and
- for each $i \geq 0$, the string $uv^iwx^iy$ belongs to $L$.



# Parse trees for CFG's

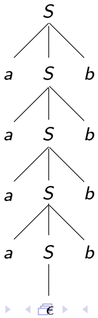
Derivations can be represented as parse trees:

**CFG $G_2$**

$$
\begin{aligned}
S &\rightarrow aSb \\
S &\rightarrow \epsilon.
\end{aligned}
$$

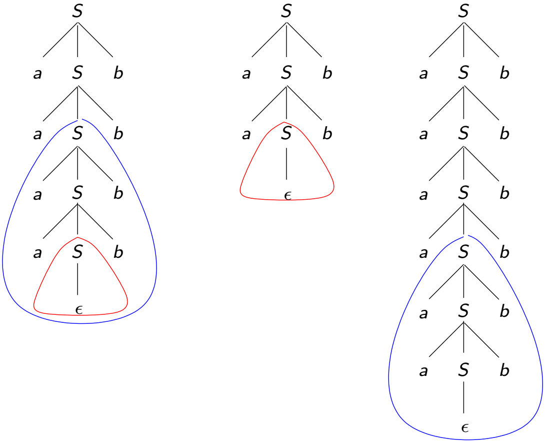
Example derivation:

$$
\begin{aligned}
S &\Rightarrow aSb \\
&\Rightarrow aaSbb \\
&\Rightarrow aaaSbbb \\
&\Rightarrow aaaaSbbbb \\
&\Rightarrow aaaabbbb.
\end{aligned}
$$



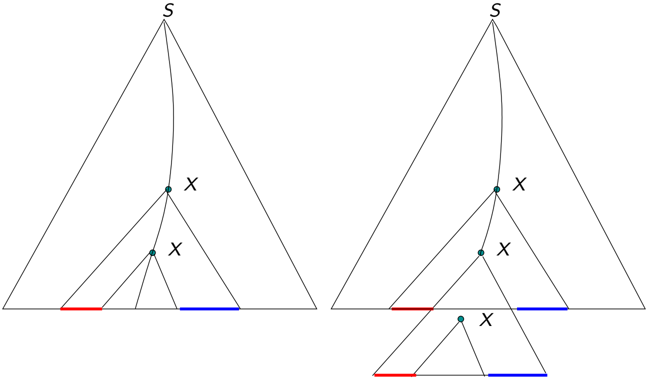
# Cutting and pasting in parse trees

Subtrees hanging at same non-terminal can be replaced for eachother.



# Proof idea

A long string must have a deep parse tree, which in turn means a path with a repeated non-terminal.

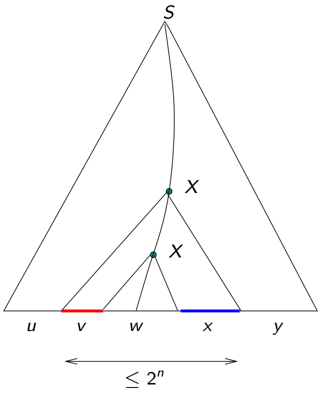


# Proof

- Let $G$ be a CNF grammar for $L$.
- A complete binary tree with $l$ levels has $2^{l-1}$ leaf nodes.
- A parse tree in $G$ with $l$ levels has a terminal string ("yield") of length at most $2^{l-2}$.
- Hence a string of length $2^l$ or more, must have a parse tree of at least $l+2$ levels.
- Take $k = 2^n$ where $n$ is the number of non-terminals in $G$.

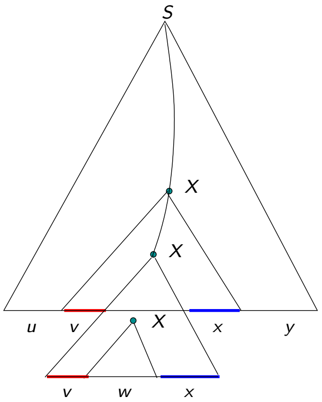
# Proof

- Consider parse tree in $G$ of a string $z$ of length at least $k = 2^n$.
- Consider longest path from root to leaf.
- Choose the first repeated non-terminal $X$ starting from bottom of path.
- Path from upper $X$ down to leaf is at most $n + 2$ levels. Also it must be the longest path in the subtree rooted at $X$. Hence length of $vwx$ is at most $2^n$.
- Also $vx \neq \epsilon$, as $G$ is a CNF grammar.



# Proof

- Consider parse tree in $G$ of a string $z$ of length at least $k = 2^n$.
- Consider longest path from root to leaf.
- Choose the first repeated non-terminal $X$ starting from bottom of path.
- Path from upper $X$ down to leaf is at most $n+2$ levels. Also it must be the longest path in the subtree rooted at $X$. Hence length of $vwx$ is at most $2^n$.
- Also $vx \neq \epsilon$, as $G$ is a CNF grammar.
- Each $uv^iwx^iy$ also belongs to $L(G)$.



# Applications

Argue that the following languages are not CFL's:

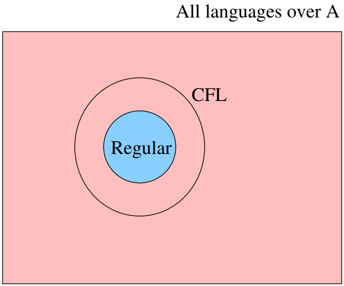
- $\{a^n b^n c^n \mid n \geq 0\}$.

# Applications

Argue that the following languages are not CFL's:

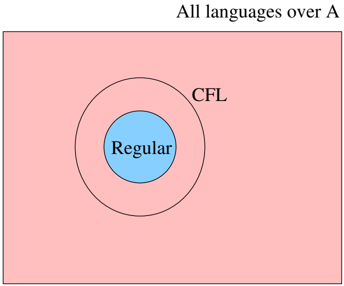
- $\{a^n b^n c^n \mid n \geq 0\}$.
- $\{ww \mid w \in \{a, b\}^*\}$.

# Closure Properties of CFL's



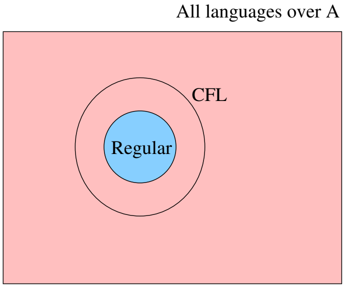

| | Closed? |
|---|---|
| Union | |

# Closure Properties of CFL's



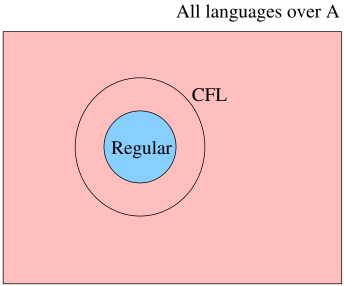

| | Closed? |
|---|---|
| Union | $\surd$ |
| Intersection | |

# Closure Properties of CFL's




| | Closed? |
|---|---|
| Union | $\surd$ |
| Intersection | X |
| Complementation | |

# Closure Properties of CFL's




| | Closed? |
|---|---|
| Union | $\surd$ |
| Intersection | X |
| Complementation | X |