Pumping Lemma for Context-Free Languages

Deepak D'Souza

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

23 September 2019

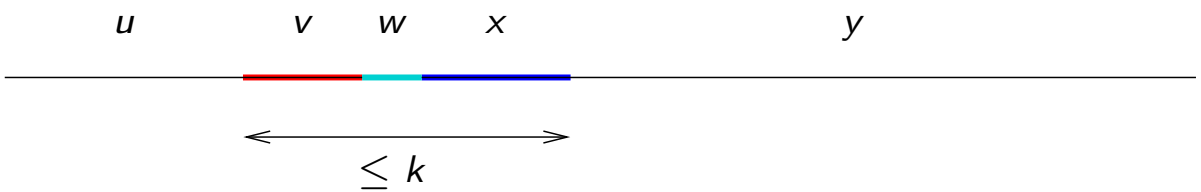
Outline

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 $uv^i wx^i y$ belongs to L.



Parse trees for CFG's

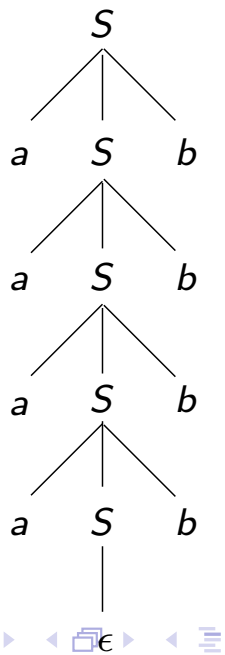
Derivations can be represented as parse trees:

CFG G_2

$$S \rightarrow aSb$$
$$S \rightarrow \epsilon.$$

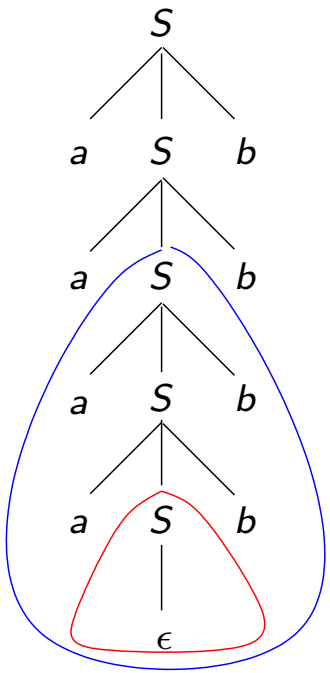
Example derivation:

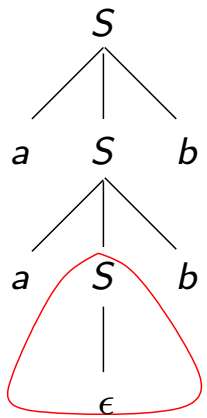
$$\begin{aligned} S &\Rightarrow aSb \\ &\Rightarrow aaSbb \\ &\Rightarrow aaaSbbb \\ &\Rightarrow aaaaSbbbb \\ &\Rightarrow aaaabbbb. \end{aligned}$$

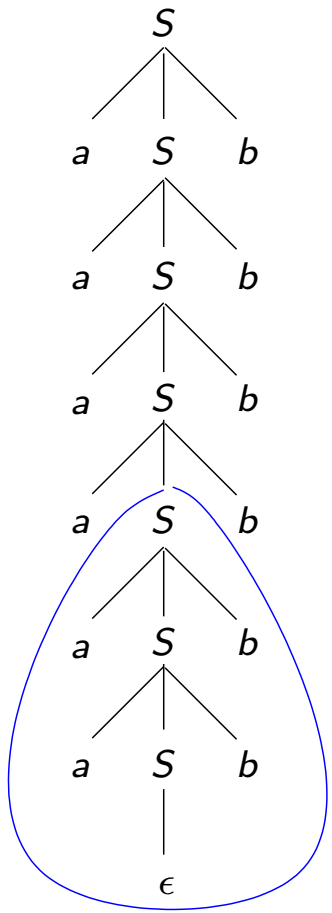


Cutting and pasting in parse trees

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

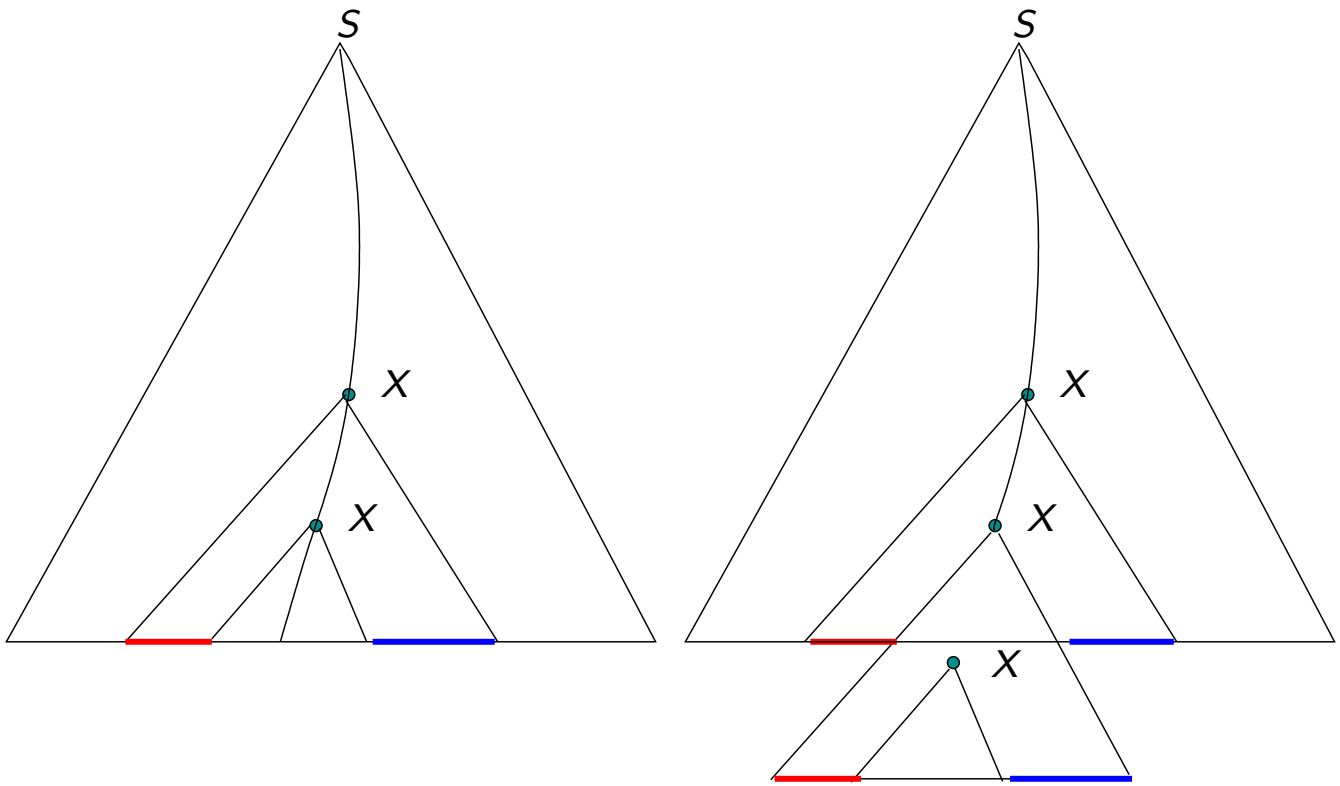






Proof idea

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

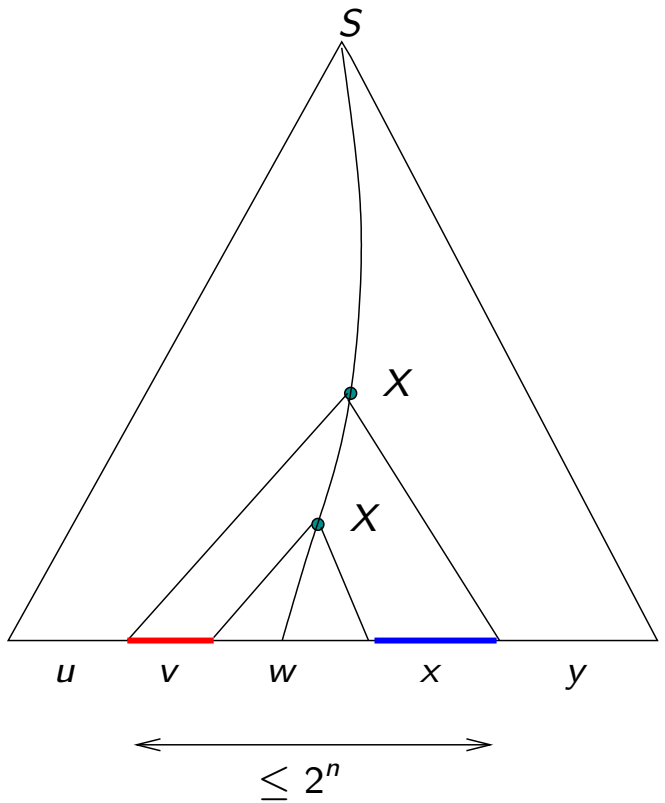


Proof

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

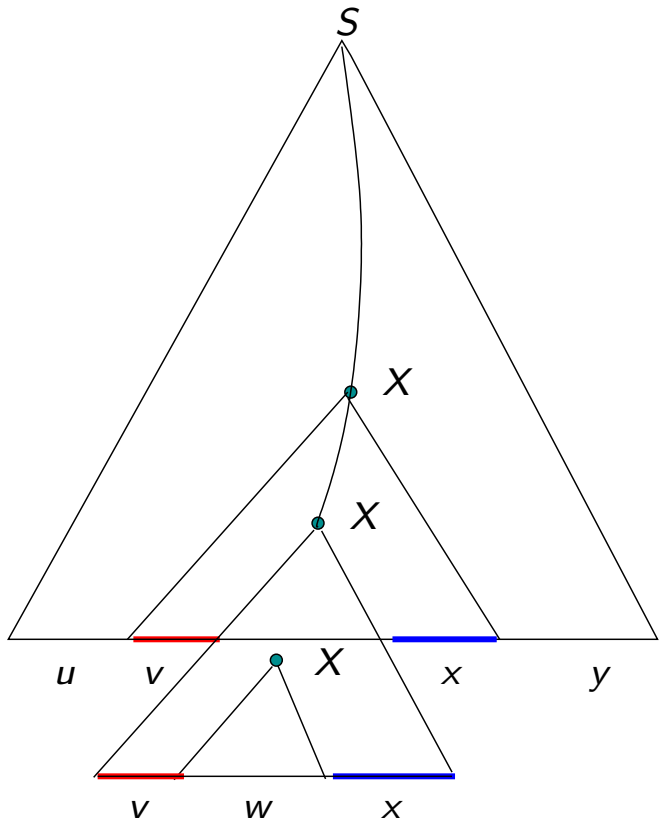
Proof - II

- 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 - II

- 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:

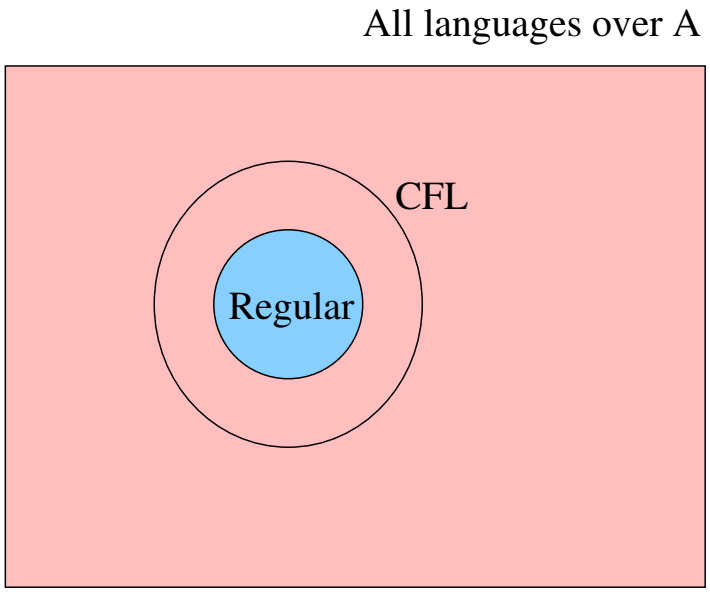
- $\{a^n b^n c^n \mid n \geq 0\}$.

Applications

Argue that the following languages are not CFL's:

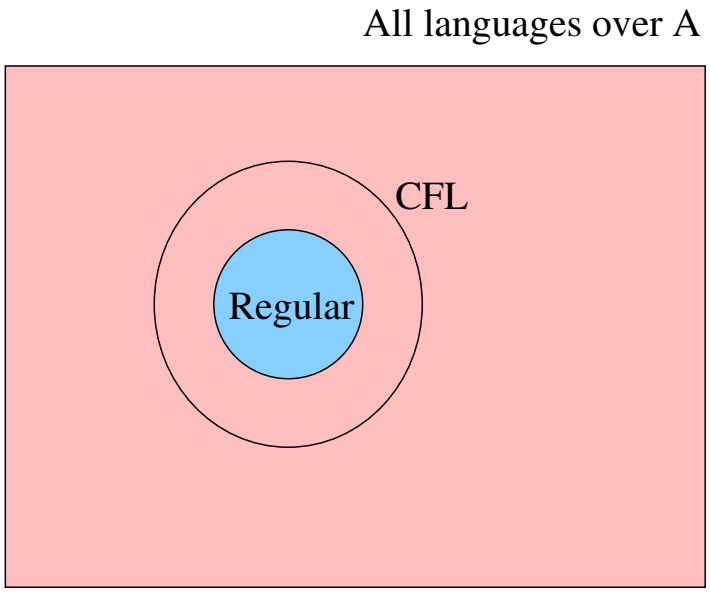
- $\{a^n b^n c^n \mid n \geq 0\}$.
- $\{ww \mid w \in \{a, b\}^*\}$.

Closure Properties of CFL's



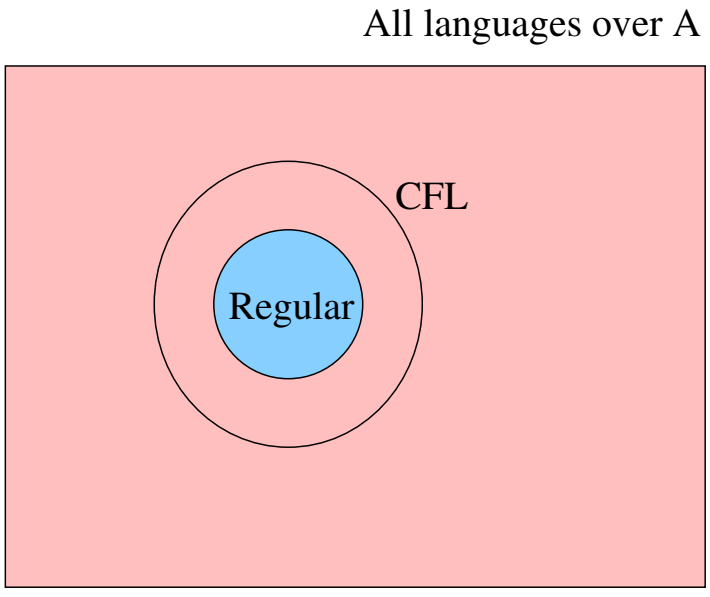
	Closed?
Union	

Closure Properties of CFL's



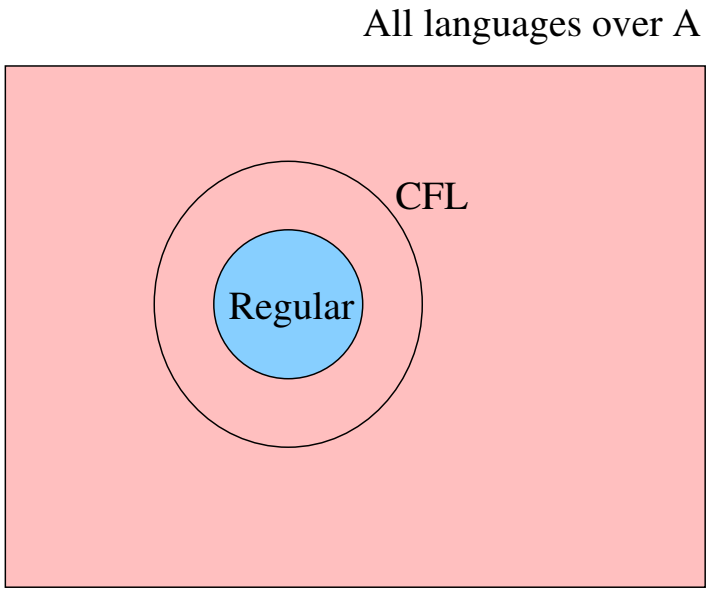
	Closed?
Union	√
Intersection	

Closure Properties of CFL's



	Closed?
Union	√
Intersection	X
Complementation	

Closure Properties of CFL's



	Closed?
Union	√
Intersection	X
Complementation	X