Harvard CS 121 and CSCI E-207

Lecture 11: CFL Closure Properties and Non-Context-Free Languages

Harry Lewis

October 8, 2009

Reading: Sipser, pp. 119-128.

Closure Properties of CFLs

- **Thm (last time):** The CFLs are the languages accepted by PDAs
- **Thm:** The CFLs are closed under
  - Union
  - Concatenation
  - Kleene $*$
  - Intersection with a regular set

The intersection of a CFL and a regular set is a CFL

Pf sketch: Let L_1 be CF and L_2 be regular

$L_1 = L(M_1)$, M_1 a PDA

$L_2 = L(M_2)$, M_2 a DFA

$Q_1 =$ state set of M_1

$Q_2 =$ state set of M_2

Construct a PDA with state set $Q_1 \times Q_2$ which keeps track of computation of both M_1 and M_2 on input.

Q: Why doesn't this argument work if M_1 and M_2 are both PDAs?

In fact, the intersection of two CFLs is not necessarily CF.

And the complement of a CFL is not necessarily CF (PS4).

Q: How to prove that languages are not context free?

Pumping Lemma for CFLs

Lemma: If L is context-free, then there is a number p (the pumping length) such that any $s \in L$ of length at least p can be divided into $s = uvxyz$, where

1. $uv^ixy^iz \in L$ for every $i \geq 0$,
2. $v \neq \varepsilon$ or $y \neq \varepsilon$, and
3. $|vxy| \leq p$.

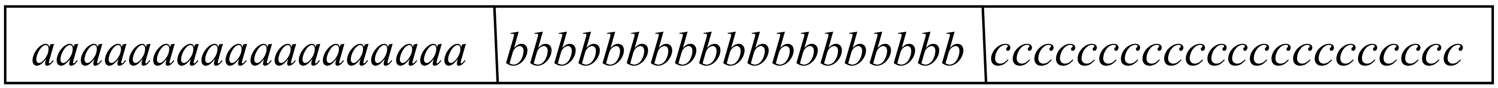
Pumping Lemma for CFLs (aka Yuvecksy'sTheorem ;)

Lemma: If L is context-free, then there is a number p (the pumping length) such that any $s \in L$ of length at least p can be divided into $s = uvxyz$, where

1. $uv^ixy^iz \in L$ for every $i \geq 0$,
2. $v \neq \varepsilon$ or $y \neq \varepsilon$, and
3. $|vxy| \leq p$.

Using the Pumping Lemma to Prove Non-Context-Freeness

$\{a^n b^n c^n : n \geq 0\}$ is not CF.



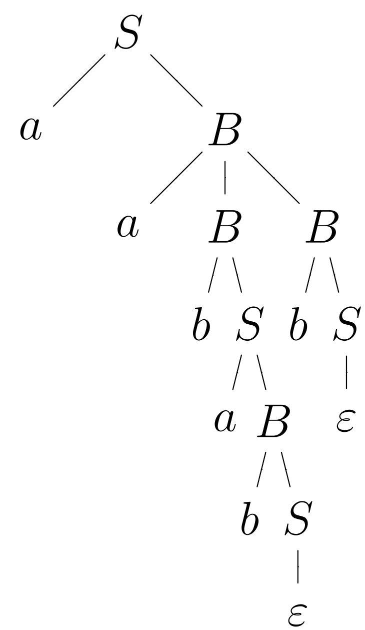
What are v, y?

- Contain 2 kinds of symbols
- Contain only one kind of symbol

$\Rightarrow$ **Corollary:** CFLs not closed under intersection.

Recall: Parse Trees

$$
\begin{array}{rcl}
S & \rightarrow & aB \mid bA \mid \varepsilon \\
A & \rightarrow & aS \mid bAA \\
B & \rightarrow & bS \mid aBB
\end{array}
$$

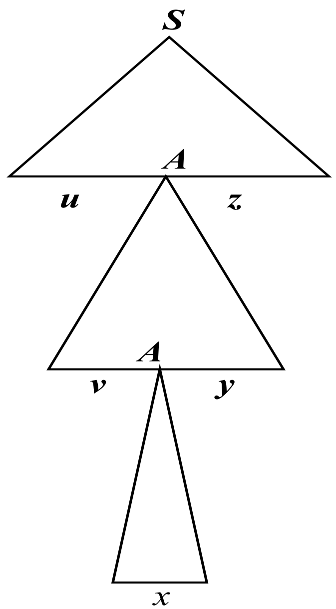


Parse tree for $aabab$, the "yield" of the tree

Height = max length path from S to a terminal symbol = 6 in above example.

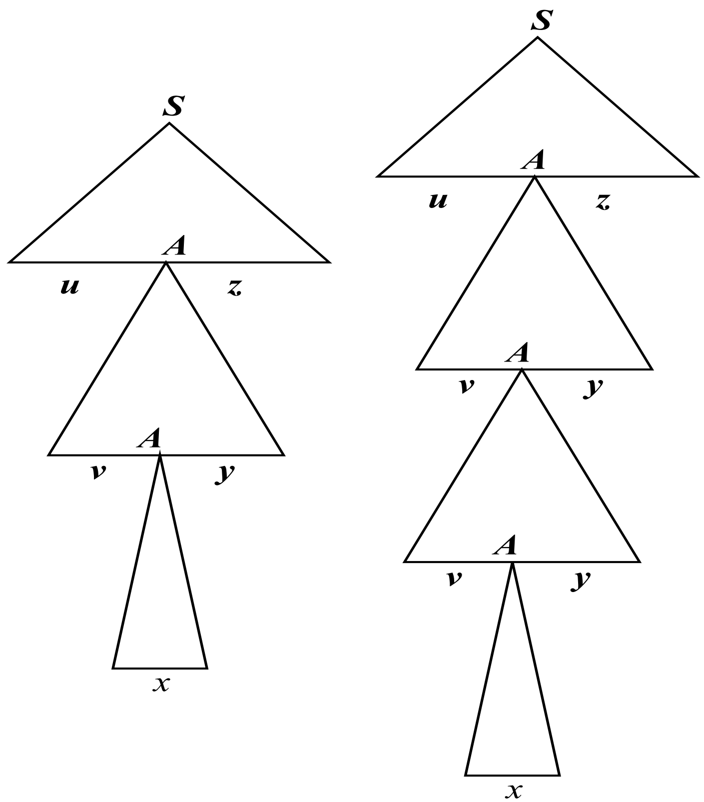
Proof of Pumping Theorem

Show that there exists a p such that any string s of length $\geq p$ has a parse tree of the form:



Proof of Pumping Theorem

Show that there exists a p such that any string s of length $\geq p$ has a parse tree of the form:



Finding "Repetition" in a big parse tree

- Since RHS of rules have bounded length, long strings must have tall parse trees.
- A tall parse tree must have a path with a repeated nonterminal.
- Let $p = b^m + 1$, where:

  $b =$ max length of RHS of a rule

  $m =$ # of variables

- Suppose T is the smallest parse tree for a string $s \in L$ of length at least p. Then

  Let $h =$ height of T. Then $b^h \geq p = b^m + 1$,

  $\Rightarrow h > m$,

  $\Rightarrow$ Path of length h in T has a repeated variable.

Final annoying details

- **Q:** Why is v or y nonempty?

- **Q:** How to ensure $|vxy| \leq p$?