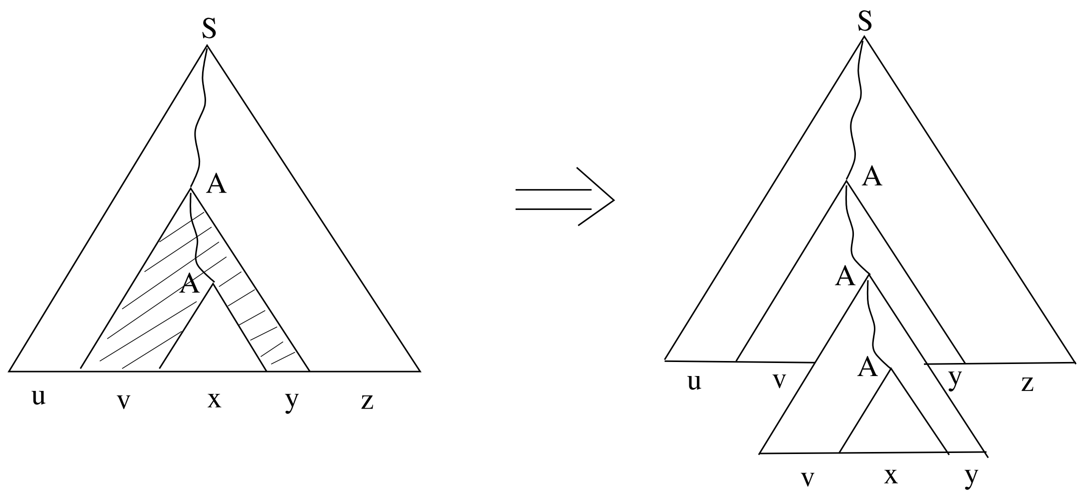
Lecture 13: Pumping Theorem for CFLs

The idea behind the pumping theorem for CF languages is similar to the pumping theorem for regular languages. If the parse tree of a string is deep enough, then at least one nonterminal must be repeated along a certain branch. Some relevant substrings can then either be trimmed away or be pasted repeatedly.



$w = uvxyz$, then

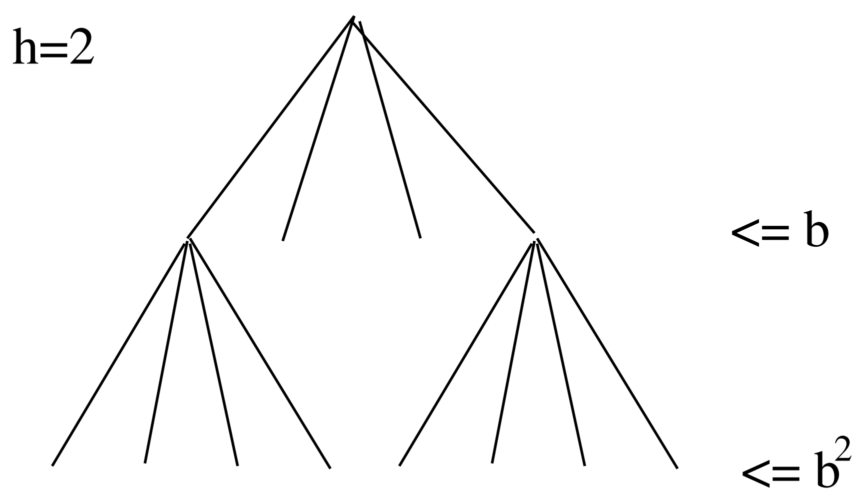
$uxz \in L$, and

$uv^2xy^2z \in L$, and so on.

Height of tree= length of the longest path from the root to some leaf.

Let b be the branching factor of a tree, i.e., the maximum number of branches from any node.

A tree of height h with a branching factor of b can have at most b^h leaves.
Therefore, a tree with b^h leaves must have height $\geq h$.



Let b be the *fanout of a grammar* G, i.e., the maximum number of symbols on the RHS of any rule of G. A string of length b^h must have its parse tree of height $\geq h$.

The pumping theorem for CFLs

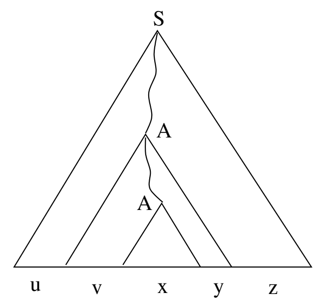
Theorem 1 *Let L be a context free language. Then there is an integer $N \geq 1$ such that for every $w \in L$ and $|w| \geq N$, w can be split into five parts $w = uvxyz$ such that*

1. $vy \neq e$ *(i.e. v and y cannot both be e)*
2. $|vxy| \leq N$
3. $uv^ixy^iz \in L$ *for* $i = 0, 1, 2, ...$

Proof:

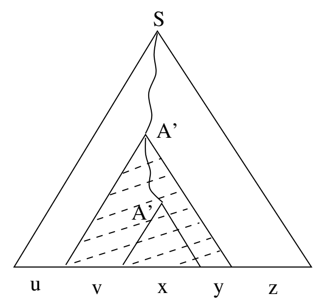
- Let $G = (V, \Sigma, R, S)$ be a CFG that generates the CFL L.
- Let $b = \max\{|\alpha| : A \to \alpha \text{ in } G\}$
- Let $N = b^{|NT|+1} = b^{|V-\Sigma|+1}$
- Let $w \in L$ and $|w| \geq N$.
- Let T be a parse tree for w that has the smallest number of leaves.
- Since $|w| \geq N = b^{|NT|+1}$, T must have height $\geq |NT| + 1$.
- Thus, there is a path (from root to leaf) containing $\geq |NT| + 2$ nodes, with all nodes except the last one being nonterminals.
- By Pigeonhole Principle, since there are only $|NT|$ distinct nonterminals in G and there are $|NT| + 1$ nonterminals along the path, one of the nonterminals must appear at least twice in the path.

- For the pair of repeated nonterminal, denote A, we can define u, v, x, y, z as shown in the figure below



$vy \neq e$, since otherwise there is a tree for w that has a smaller number of leaves than T. Contradicting the assumption.

- Next, we show that $|vxy| \leq b^{|NT|+1} = N$.
  Among all the repeated nonterminal, pick the 'lowest' nonterminal A' such that the shaded portion has no other repeated nonterminal.



  Since there is only one repetition in the shaded subtree, the longest path within the shaded subtree must have at most $|NT| + 1$ nonterminals (all the distinct ones plus the repeated A') and one terminal (the leaf in that path)

  - Hence, the height of the shaded subtree is at most $|NT| + 1$.
  - Hence, the yield of the subtree is at most $b^{|NT|+1}$. Hence, $|vxy| \leq b^{|NT|+1} = N$

- Replacing the bigger expansion of A' by the smaller, we get uxz; replacing the smaller by the bigger, we get uv^2xy^2z, and so on. Hence, $uv^ixy^iz \in L(G)$ for $i = 0, 1, 2, ...$

Proving a language is not CF

Prove that $L = \{a^i b^i c^i : i \geq 0\}$ is not context-free.
Proof:

- Suppose L is context free.
- Then there exists a CFG $G = (V, \Sigma, R, S)$ such that $L = L(G)$. Let N be the integer in the Pumping Theorem.
- Choose $w = a^N b^N c^N$.
- By the Pumping Theorem, w can be rewritten as $w = uvxyz$ such that $vy \neq e$, $|vxy| \leq N$, and $uv^i xy^i z \in L$ for all $i = 0, 1, 2, ...$
- Several cases:
  1. vxy contain only one type of symbols (i.e., only a's, only b's, or only c's). Then $uv^2xy^2z \notin L$ since the number of a, b and c will not be equal.
  2. vxy contains both a's and b's. Then $uv^2xy^2z \notin L$ since the number of a or the number of b's will be more than the number of c's. The resulting string also may not be in the form $a^*b^*c^*$.
  3. vxy contains both b's and c's. Similar to case 2. (Note: since $|vxy| \leq N$, it is not possible for vxy to contain both a's and c's.)

That is, for all possible ways of splitting, we have $uv^ixy^iz \notin L$ for some i. Contradiction to the Pumping Theorem.

- Hence, L is not context-free.

Proving a language is not CF

Prove that

$L = \{a^n b^m a^n b^m : m, n \geq 0\}$ is not context-free.

Proof:

- Suppose L is context free.
- Then there exists a CFG $G = (V, \Sigma, R, S)$ such that $L = L(G)$. Let N be the integer in the Pumping Theorem.
- Choose $w = a^N b^N a^N b^N$.
- By the Pumping Theorem, w can be rewritten as $w = uvxyz$ such that $vy \neq e$, $|vxy| \leq N$, and $uv^i xy^i z \in L$ for all $i = 0, 1, 2, ...$
- Two cases:
  1. Either v or y contains both types of symbols. Then uv^2xy^2z will not be in the form $a^*b^*a^*b^*$, hence it cannot be in L.
  2. Both v and y contain only one type of symbol.

     (i) if v consists of a's from the first block of a's in w, then since $|vxy| \leq N$, y cannot be from the second block of a's. Thus, uv^2xy^2z will not have two equal blocks of a's, thus it is not in L.

(ii) if v consists of b's from the first block of b's in w, then since $|vxy| \leq N$, y cannot be from the second block of b's. Thus, uv^2xy^2z will not have two equal blocks of b's, thus it is not in L.

(iii) if v consists of a's from the second block of a's in w, then pumping vy would increase the number of a's in the second block while the number of a's in the first block would remain the same; thus, the resulting string is not in L.

(iv) if v consists of b's from the second block of b's in w, then pumping vy would increase the number of b's in the second block while the number of b's in the first block would remain the same; thus, the resulting string is not in L.

That is, for all ways of splitting, we have $uv^ixy^iz \notin L$ for some i. Contradiction to the Pumping Theorem.

- Hence, L is not context-free.

Proving a language is not CF

Show that

$L = \{w \in \{a, b, c\}^* : w$ has equal number of a's, b's, and c's $\}$

is not context-free.

We can prove this using Pumping Theorem. In fact, the proof can be exactly the same as the proof of $\{a^i b^i c^i : i \geq 0\}$. But, here, we prove it by contradiction using closure property.

Proof:

- Suppose L is context-free.
- We have $L \cap L(a^* b^* c^*) = \{a^i b^i c^i : i \geq 0\}$
- Since $L(a^* b^* c^*)$ is regular, by earlier theorem the intersection of a regular language and a context-free language would be context-free.
- But $\{a^i b^i c^i : i \geq 0\}$ is known to be non context-free. Contradiction.
- Hence L is not context-free.

Theorem 2 *The CFLs are not closed under intersection.*

Proof: (by counter example)

$L_1 = \{a^n b^n c^m : m, n \geq 0\}$ is a CFL.

$$S \to TC, T \to aTb, T \to e, C \to cC, C \to e$$

$L_2 = \{a^m b^n c^n : m, n \geq 0\}$ is a CFL.

$$S \to AT, T \to bTc, T \to e, A \to aA, A \to e$$

But $L = L_1 \cap L_2 = \{a^n b^n c^n : n \geq 0\}$ is not a CFL.

Theorem 3 *The CFLs are not closed under complementation.*

Proof:
Suppose CFLs were closed under complementation. Then since

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$

and CFLs is also closed under union, CFLs would be closed under intersection! A contradiction. $\Box$