Pumping lemma for CFL

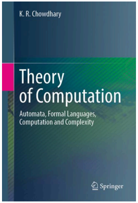
Prof. (Dr.) K.R. Chowdhary
Email: kr.chowdhary@iitj.ac.in

Formerly at department of Computer Science and Engineering
MBM Engineering College, Jodhpur

Saturday 19th July, 2025

K.R. Chowdhary

Theory of Computation



Automata, Formal Languages, Computation and Complexity

Focuses on pedagogy in its writing, that represents a refreshing approach

Ensures comprehensive and enjoyable learning

Undergone a rigorous classroom testing

Pumping Lemma for CFLs

- Given any *CFG G*, $L = (G)$, we can convert it to CNF. The parse-tree creates a *Binary-tree.*
- Let G has $V_1 \dots V_m$ variables. Choose this as the value of longest path in the derivation-tree $(= 2^m)$(for G to be in CNF). A constant p can be chosen such that $p = 2^m$.
- Let $w = uvxyz$, $|w| \geq p$ and $w \in L(G)$. Then followings hold:
  1. A string in L of length m or less has yield length of 2^{m-1} or less.
  2. Since $p = 2^m, \therefore 2^{m-1} = p/2$
  3. $\therefore, w$ is too long to be yielded (obtained) from a parse-tree of length (i.e. height) m.
- **Pumping Lemma for CFLs:** (Bar-Hillel Lemma):

  If a language L is CF, then there exists some integer $p \geq 1$ such that any string in L with $|w| \geq p$ (where p is pumping length) can be written as: $w = uvxyz$, with substrings u, v, x, y, z, such that $|uxy| \geq p$, $|vy| \geq 1$ and,

  $$uv^n xy^n x \in L \text{ for every integer } n \geq 0.$$

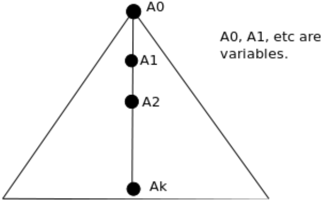
- All CF languages are guaranteed to satisfy this property.

Pumping Lemma for CFLs

For a parse-tree of length $m+1$, choose a path to be $m+1$, yield must be 2^m, and $|w| \leq p$.

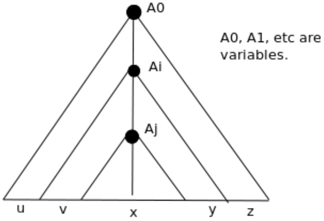
Any parse-tree that yields w must have a path of at least $m+1$.

$w = uvxyz, |w| \geq p$.



- if $k \geq m$ then at least 2 of these variables must be same (since there are m unique variables).
- Suppose the variables are same at $A_i = A_j$, where $1 \leq i < j \leq k$.

Pumping Lemma for CFLs . . .



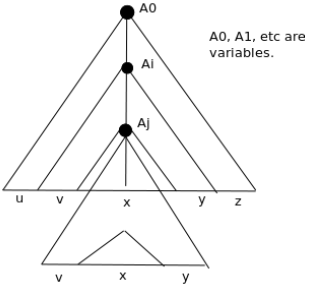
- **Case of** $vy \neq \varepsilon$: The v, y cannot be terminals, otherwise there would not be A_j.

  $\therefore$ we must have two variables, one of them must lead to A_j and other must lead to v or y or both.

- **Case of** $|vxy| \leq p$: (The middle portion is no longer than p). That is yield of subtree rooted at A_i, or for the longest path of $m+1$, $|vxy| \leq p \leq 2^{m+1-1}$. In case of A_i as A_0, vxy is entire tree.

- **Case of** $\forall i \geq 0, uv^i xy^i z \in L$ **:**

  We can show this by $A_i = A_j$. Substituting A_i for A_j, the result is uv^1xy^1z, uv^2xy^2z, etc. The tree is shown below.



To show that a language is not CF:

1.Some p must exist indicating the maximum yield and length of parse-tree.

2. Pickup w, breakup it into $uvxyz$, such that $|vxy| \leq p$, $|vy| \neq \varepsilon$,

3. We win by picking i, and showing that $uv^ixy^iz \notin L$.

Pumping Lemma for CFLs . . .

- Informally, the pumping lemma for CFLs states that for sufficiently long strings, we can find two, short nearby substrings, that we "can pump" in tandom, and resulting string must also be in the language.

- The pumping lemma states that w can be decomposed into five substrings. And, two substrings v, y or one of them can be pumped arbitrary times, and the language strings are still recognized.

- The finite languages, which are regular, and hence context-free, obey the pumping-lemma trivially by having p equal to the maximum string's length in L plus one.

Application of Pumping Lemma for CFL

- **Example:** Show that $L = \{a^p b^p c^p | p \geq 0\} \notin CFL$.
- Assume that L is CFL, to contradict later, Let p is pumping length of L. Let $w = a^p b^p c^p \in L$.
- The pumping lemma tells us that w can be written as $w = uvxyz$, where u, v, x, y, z are substrings of w.
- As per theorem. $|vxy| \leq p$, $|vy| \geq 1$, and $uv^i xy^i z \in L$. for all $i \geq 0$.
- By the fact that $|vxy| \leq p$, it can be seen that vxy can be contain no more than two distinct letters. The possibilities are;
  1. $uxy = a^j$, for some $j \leq p$
  2. $vxy = b^j$, for some $j \leq p$
  3. $vxy = c^j$, for some $j \leq p$
  4. $vxy = a^j b^k$, for some $j + k \leq p$
  3. $vxy = b^j c^k$, for some $j + k \leq p$
- for each case, it can be easily vaerified that $uv^i xy^i z$ does not contain equal numbers of each letter for any $i \neq 1$. Thus, $uv^2 xy^2 z$ does not have the form $a^i b^i c^i$. Hence, the contradiction, and $L \notin CFL$.

Bibliography

Chowdhary, K.R. (2025). Normal Forms of Context-Free Grammars. In: Theory of Computation. Springer, Singapore. https://doi.org/10.1007/978-981-97-6234-7_7