The Pumping Lemma for CFLs, and Closure Properties

Prof. Shrisha Rao
IIIT Bangalore

srao@iiitb.ac.in

2025-09-04

Not all Languages are Context-Free

- We know that context-free languages are generated by context-free grammars a/k/a Type 2 grammars, and are accepted by PDAs.
- Languages generated by context-sensitive grammars (Type 1 grammars) or unrestricted grammars (Type 0 grammars) cannot be accepted by PDAs.
- We need a test that can determine if a language is context-free.

The Pumping Lemma for CFLs

Let L be a context-free language. Then there is a constant k such that for every $z \in L$ of length at least k, it is possible to split z as $uvwxy$, where:

- vx is not ϵ, i.e., $|vx| > 0$
- $|vwx| \leq k$
- $uv^iwx^iy \in L$, for all $i \geq 0$

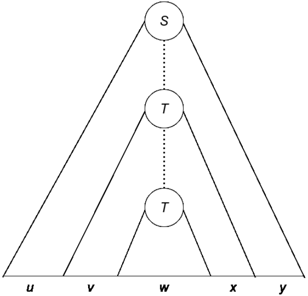
Reasoning

- As with the Pumping Lemma for regular languages, the similarly-named Pumping Lemma for context-free languages does not prove that a language is context-free; it proves that a language is *not* context-free.
- The idea behind this Pumping Lemma is again that any CFG is finite (has a finite number of production rules), and thus any sufficiently long string in a CFL must involve repetition of some production rule.
- As with the previous Pumping Lemma for regular languages, this allows us to formulate the result.

Proof Sketch

- Assume that a CFL L is generated by a CFG G. Consider a very long string z in L. Then, any derivation tree for z has $|z|$ leaves.
- If z is long enough, then there must be a path in the tree from the start symbol S to a leaf that contains the same non-terminal twice.
- Suppose that non-terminal is T. The leaves of the subtree under the second T form a string generated by T; let this be w. The leaves of the subtree under the first T form a string containing w; let v be the substring before w and x the one after.
- Finally, all the leaves together form z, which contains vxw; let u be the substring before vwx and y the substring after.

Proof Sketch—cont'd



It follows that $T \rightsquigarrow v^i w x^i$, $\forall i \geq 0$, and therefore $S \rightsquigarrow u v^i w x^i y$, $\forall i \geq 0$. Thus, $u v^i w x^i y \in L$, $\forall i \geq 0$, as in the lemma.

Using the Pumping Lemma for CFLs

- The classic example of a language that is not context-free is $\{0^n 1^n 2^n \mid n \geq 0\}$.
- We can use the Pumping Lemma to prove that this language is not context-free.
- Choose a sufficiently large k, and consider the string $z = 0^k 1^k 2^k$. Consider the split of z into $uvwxy$. As vwx combined has length at most k, vx cannot contain both 0s and 2s.
- Therefore, uwy cannot have equal numbers of 0s, 1s, and 2s, and therefore is not in the language.

Exercises

Prove that the following languages are not context-free.

(1) $\{w\#w \mid w \in (0+1)^*\}$

(2) $\{0^n 1^{2n} 0^n \mid n \geq 0\}$

(3) $\{0^n 1^n 0^n 1^n \mid n \geq 0\}$

Exercises—cont'd

(4) Prove that the complement of $\{0^n1^n2^n \mid n \geq 0\}$ is context-free.
Lesson learned: *The CFLs are not closed under complementation.*

(5) Prove that the context-free languages are closed under union, concatenation, and Kleene-star.

(6) Prove that the context-free languages are not closed under intersection.

Proving that CFLs are closed under union

- Let L_1 and L_2 be languages generated by CFGs $G_1 = \langle \Sigma_1, N_1, P_1, S_1 \rangle$ and $G_2 = \langle \Sigma_2, N_2, P_2, S_2 \rangle$ respectively.
- Construct a new grammar G to generate $L_1 \cup L_2$, as follows:

$$G = \left\langle \Sigma_1 \cup \Sigma_2, N_1 \cup N_2 \cup \{S\}, P_1 \cup P_2 \cup \{S \to S_1 \mid S_2\}, S \right\rangle$$

- S is a new start symbol for G, and from there, derivations either go to S_1 or S_2.
- Every derivation using G produces a string either in L_1 or L_2, and all of L_1 and L_2 can be derived using G. Hence G is a grammar for $L_1 \cup L_2$.

Proving that CFLs are closed under concatenation

- Let L_1 and L_2 be languages generated by CFGs $G_1 = \langle \Sigma_1, N_1, P_1, S_1 \rangle$ and $G_2 = \langle \Sigma_2, N_2, P_2, S_2 \rangle$ respectively.
- Construct a new grammar G to generate $L_1 L_2$, as follows:

$$G = \left\langle \Sigma_1 \cup \Sigma_2, N_1 \cup N_2 \cup \{S\}, P_1 \cup P_2 \cup \{S \to S_1 S_2\}, S \right\rangle$$

- S is a new start symbol for G, and from there, derivations go to $S_1 S_2$, with S_1 and S_2 in turn producing strings in L_1 and L_2 respectively.
- Every derivation using G produces a string in L_1 followed by a string in L_2, and all of $L_1 L_2$ can be derived using G. Hence G is a grammar for $L_1 L_2$.

Proving that CFLs are closed under Kleene-star

- Let L be a language generated by the CFG $G = \langle \Sigma, N, P, S \rangle$.
- Construct a new grammar G^* to generate L^*, as follows:

$$G^* = \left\langle \Sigma, N \cup \{S^*\}, P \cup \{S^* \to SS^* \mid \epsilon\}, S^* \right\rangle$$

- Every derivation using G^* produces ϵ or some sequence of strings in L. Therefore, $L*$ is context-free.

Proving that CFLs are not closed under intersection

- Consider the languages $L_1 = \{0^i 1^i 2^j \mid i, j \geq 0\}$ and $L_2 = \{0^j 1^i 2^i \mid i, j \geq 0\}$.
- L_1 and L_2 are both context-free, being generated by:

$$
\begin{aligned}
G_1 : S &\rightarrow BC \\
B &\rightarrow 0B1 \mid \epsilon \\
C &\rightarrow 2C \mid \epsilon \\
G_2 : S &\rightarrow AB \\
A &\rightarrow 0A \mid \epsilon \\
B &\rightarrow 1B2 \mid \epsilon
\end{aligned}
$$

- However, the intersection of L_1 and L_2 is $\{0^i 1^i 2^i \mid i \geq 0\}$, which is not context-free as previously seen.

Proving that CFLs are not closed under complementation–again

- Proof by contradiction: Let L_1 and L_2 be two CFLs. If the CFLs are closed under complementation, then $\overline{L_1}$ and $\overline{L_2}$ must also be context-free.
- Since we know that CFLs are closed under union, $\overline{L_1} \cup \overline{L_2}$ must also be context-free.
- Further, since the CFLs are assumed to be closed under complement, $\overline{\overline{L_1} \cup \overline{L_2}}$ must also be context-free.
- However, by De Morgan's Law of set theory, $\overline{\overline{L_1} \cup \overline{L_2}} = L_1 \cap L_2$, which implies that CFLs are closed under intersection, which is false.
- Therefore, the CFLs are not closed under complementation.

Exercises

Give CFGs for the following languages:

(7) The language of all strings over {0,1} that are either of the form 0^n1^n or are palindromes.

(8) The language of all strings over {0,1} that consist of a palindrome followed by 0^n1^n for some n.

(9) The language of all strings of the form $(0^n1^n)^*$.