Theory of Formal Languages and Automata

Lecture 13

Mahdi Dolati

Sharif University of Technology

Fall 2023

November 22, 2023

Closure

Union

Theorem

CFLs are closed under union.

Proof.

- Consider $G_1 = (V_1, \Sigma, R_1, S_1)$ such that $L(G_1) = A$
- Consider $G_2 = (V_2, \Sigma, R_2, S_2)$ such that $L(G_2) = B$
- Construct $G = (V, \Sigma, R, S)$ by
  - $V = V_1 \cup V_2 \cup \{S\}$
  - $R = R_1 \cup R_2 \cup \{S \to S_1 \mid S_2\}$
- If $w \in A$ then $S_1 \overset{*}{\Rightarrow} w$. So, $S \Rightarrow S_1 \overset{*}{\Rightarrow} w$
- If $w \in B$ then $S_2 \overset{*}{\Rightarrow} w$. So, $S \Rightarrow S_2 \overset{*}{\Rightarrow} w$
- If $w \in L(G)$ then $S \Rightarrow S_1 \overset{*}{\Rightarrow} w$ or $S \Rightarrow S_2 \overset{*}{\Rightarrow} w$. Thus, $w \in A \cup B$.

□

Theorem

CFLs are closed under concatenation.

Proof.

- Consider $G_1 = (V_1, \Sigma, R_1, S_1)$ such that $L(G_1) = A$
- Consider $G_2 = (V_2, \Sigma, R_2, S_2)$ such that $L(G_2) = B$
- Construct $G = (V, \Sigma, R, S)$ by
  - $V = V_1 \cup V_2 \cup \{S\}$
  - $R = R_1 \cup R_2 \cup \{S \to S_1 S_2\}$

$\square$

Theorem

CFLs are closed under Kleene star.

Proof.

- Consider $G_1 = (V_1, \Sigma, R_1, S_1)$ such that $L(G_1) = A$.
- Construct $G = (V, \Sigma, R, S)$ by
  - $V = V_1 \cup \{S\}$
  - $R = R_1 \cup \{S \to SS_1 \mid \varepsilon\}$

$\square$

Theorem

CFLs are closed under prefix.

Proof.

- Consider a CFG that generates the language L. Convert it to Chomsky normal form $G = (V, \Sigma, R, S)$.
  - If $L = \emptyset$, then $\text{Prefix}(L) = \emptyset$.
- Consider the derivation and parse tree for any string in the language. Any prefix groups the variables in the parse tree into three types.
  - The prefix includes all terminals in subtree of the variable,
  - The prefix includes some terminals in subtree of the variable (split path),
  - The prefix includes no terminals in subtree of the variable,

□

Proof Cont.

- We want to construct a CFG that keeps track of a split path such that:
  - Terminals left to the split path are produced,
  - Terminals right to the split are replaced with ε,
  - The terminal connected to the last variable in the split path is produced.
- We introduce three variables $\langle A, L\rangle$, $\langle A, S\rangle$, and $\langle A, R\rangle$ for every variable A in the grammar.
- Note, we have three types of rules:
  - $S \to \varepsilon$
  - $A \to BC$
  - $A \to t$

□

Proof Cont.

- Construct $G' = (V', \Sigma, R', S')$:
  - $V' = \{\langle A, D\rangle \mid A \in V$ and $D \in \{L, S, R\}\}$,
  - $S' = \langle S, S\rangle$,
  - R':

Since $L \neq \emptyset$	$\langle S, S\rangle \rightarrow \varepsilon$
For every $A \rightarrow BC$	$\langle A, L\rangle \rightarrow \langle B, L\rangle\langle C, L\rangle$ $\langle A, S\rangle \rightarrow \langle B, L\rangle\langle C, S\rangle \mid \langle B, S\rangle\langle C, R\rangle$ $\langle A, R\rangle \rightarrow \langle B, R\rangle\langle C, R\rangle$
For every $A \rightarrow t$	$\langle A, L\rangle \rightarrow t$ $\langle A, S\rangle \rightarrow t$ $\langle A, R\rangle \rightarrow \varepsilon$

□

Proof Cont.

- Consider string $w = w_1 w_2 \dots w_n \in L$, therefore there is derivation:

$$S \stackrel{*}{\Rightarrow} A_1 A_2 \dots A_n \text{ where } A_i \Rightarrow w_i,$$

- By construction, for each $1 \leq i \leq n$, we have:

$$\begin{aligned} \langle S, S \rangle &\stackrel{*}{\Rightarrow} \langle A_1, L \rangle \dots \langle A_{i-1}, L \rangle \langle A_i, S \rangle \langle A_{i+1}, R \rangle \dots \langle A_n, R \rangle \\ &\stackrel{*}{\Rightarrow} w_1 w_2 \dots w_i \end{aligned}$$

- Therefore, $G^{'}$ derives any prefix of any string in L.
- A similar argument shows that any string in $L^{'}$ is a prefix of a string in L.

□

Theorem

CFLs are closed under reversal.

Proof.

Consider grammar $G = (V, \Sigma, R, S)$ where $B = L(G)$. Construct CFG G' where

$$R^{'} = \{A \to u^{\mathcal{R}} \mid A \to u \text{ is a rule in } R\}.$$

Show that each for variable $A \in V$ and $u \in (V \cup \Sigma)^*$

$$A \overset{n}{\Rightarrow}_G u \leftrightarrow A \overset{n}{\Rightarrow}_{G'} u^{\mathcal{R}}.$$

$\overset{n}{\Rightarrow}$ means $\overset{*}{\Rightarrow}$ in exactly k steps. □

Proof Cont.

- Base case $n = 0$: $A \overset{0}{\Rightarrow}_G u$, then $u = A = u^{\mathcal{R}}$. So $A \overset{0}{\Rightarrow}_{G'} u^{\mathcal{R}}$ and vice versa.

□

Proof Cont.

- Inductive step: Assume for all $n > 0$, $A \in V$ and $u \in (V \cup \Sigma)^*$:

$$A \xRightarrow{n-1}_G u \leftrightarrow A \xRightarrow{n-1}_{G'} u^{\mathcal{R}}.$$

- If $A \overset{n}{\Rightarrow}_G u$, then there is $C \in V$ and $x, y, z \in (V \cup \Sigma)^*$ such that $u = xyz$, $A \xRightarrow{n-1}_G xCz$, and $C \Rightarrow_G y$.
- By induction hypothesis: $A \xRightarrow{n-1}_{G'} z^{\mathcal{R}} C x^{\mathcal{R}}$.
- By construction $C \Rightarrow_{G'} y^{\mathcal{R}}$.
- Thus, $A \overset{n}{\Rightarrow}_{G'} z^{\mathcal{R}} y^{\mathcal{R}} x^{\mathcal{R}} = (xyz)^{\mathcal{R}} = u^{\mathcal{R}}$.
- Repeat for the reverse direction. Then:

$$A \overset{n}{\Rightarrow}_G u \leftrightarrow A \overset{n}{\Rightarrow}_{G'} u^{\mathcal{R}}.$$

□

Proof Cont.

- Therefore, for $w \in B$

$$S \overset{*}{\Rightarrow}_G w \leftrightarrow S \overset{*}{\Rightarrow}_{G'} w^{\mathcal{R}},$$

which implies $L(G') = B^{\mathcal{R}}$

□

Theorem

CFLs are closed under Suffix*.*

Proof.

Note that $\text{Suffix}(A) = \text{Prefix}(A^{\mathcal{R}})^{\mathcal{R}}$. CFLs are closed under prefix and reversal. Thus, CFLs are closed under suffix. □

Theorem

CFLs are closed under intersection with a regular language.

Proof.

- A a CFL recognized by the PDA $M_1 = (Q_1, \Sigma, \Gamma, \delta_1, q_1, F_1)$
- B a regular language recognized by the NFA $M_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$
- Construct PDA $M = (Q, \Sigma, \Gamma, \delta, q_0, F)$
  - $Q = Q_1 \times Q_2$
  - $q_0 = (q_1, q_2)$
  - $F = F_1 \times F_2$
  - The transition function for all $a \in \Sigma_\varepsilon, b, c \in \Gamma_\varepsilon$:

$$\delta((q,r),a,b) = \{((s,t),c) \mid (s,c) \in \delta_1(q,a,b) \text{ and } t \in \delta_2(r,a)\},$$

  - As M runs on input w, its stack and the first element of its state change according to q_1 whereas the second element of its state changes according to q_2.

□

- Union
- Concatenation
- Kleene star
- Prefix
- Reversal
- Suffix
- Intersection with a regular language

- CFLs are not closed under intersection.
- Consider two CFLs:
  - $A = \{a^m b^m c^n \mid m, n \geq 0\}$
  - $B = \{a^m b^n c^n \mid m, n \geq 0\}$
- However, we will prove (later) that their intersection is not context free:

$$A \cap B = \{a^n b^n c^n \mid n \geq 0\}$$