Formal Languages and Automata Theory

Second Short Test

Time: 20 minutes | 26–February–2021 | Maximum marks: 20

---

1. Let $A \triangle B$ denote the symmetric difference of two languages A and B (over the same alphabet). Which of the following statements is true?

(A) If A and B are both CFLs, then $A \triangle B$ must be a CFL.
(B) If A is a CFL and B is not a CFL, then $A \triangle B$ must be a CFL.
(C) If A is a CFL and B is regular, then $A \triangle B$ must be a CFL.
(D) Neither of the other options is true.

Solution (A) and (C): Take $A = \sim\{ww^r \mid w \in \{a,b\}^*\}$ and $B = \{a,b\}^*$, so $A \triangle B = B - A = \{ww^r \mid w \in \{a,b\}^*\}$.
(B): Take $A = \{a^m b^m c^n \mid m,n \geqslant 0\}$ and $B = \{a^m b^m c^m \mid m \geqslant 0\}$, so $A \triangle B = A - B = \{a^m b^m c^n \mid m,n \geqslant 0,\ m \neq n\}$.

2. Which of the following statements is true for a CFG G with start symbol S and with the only productions $S \to aS \mid bS \mid a$?

(A) $\mathscr{L}(G)$ is a CFL but not regular.
(B) $\mathscr{L}(G)$ is regular but not a CFL.
(C) $\mathscr{L}(G)$ is a CFL and regular.
(D) $\mathscr{L}(G)$ is neither a CFL nor regular.

Solution $\mathscr{L}(G) = \mathscr{L}\left((a+b)^* a\right)$. Every regular language is context-free.

3. Consider the languages

$$L_1 = \left\{a^m b^m c^{m+n} \mid m,n \geqslant 1\right\},$$
$$L_2 = \left\{a^m b^n c^{m+n} \mid m,n \geqslant 1\right\}.$$

Which of the following statements is true?

(A) Both L_1 and L_2 are CFLs.
(B) Neither L_1 nor L_2 is a CFL.
(C) L_1 is not a CFL, but L_2 is a CFL.
(D) L_1 is a CFL, but L_2 is not a CFL.

Solution L_1: Supply $a^k b^k c^{k+1}$ to the pumping lemma, where k is a PLC.
L_2: We have $a^m b^n c^{m+n} = a^m (b^n c^n) c^m$, so the following grammar generates L_2: $S \to aSc \mid aTc$, $T \to bTc \mid bc$.

4. Let G be a CFG in the Chomsky normal form of a language L that does not contain ε. For any string $x \in L$ of length l, what is the length of the derivation of x?

(A) $l-1$ **(B)** $2l-1$ **(C)** $3l-1$ **(D)** $4l-1$

Solution You need to use $l-1$ productions of the form $A \to BC$, and l productions of the form $A \to a$.

5. Consider the languages

$$L_1 = \left\{a^m b^n \mid m,n \geqslant 0,\ m = 2n\right\},$$
$$L_2 = \left\{a^m b^n \mid m,n \geqslant 0,\ m \neq 2n\right\}.$$

Which of the following statements is true?

- **(A)** Both L_1 and L_2 are CFLs.
- **(B)** Neither L_1 nor L_2 is a CFL.
- **(C)** L_1 is not a CFL, but L_2 is a CFL.
- **(D)** L_1 is a CFL, but L_2 is not a CFL.

Solution L_1: $a^{2n}b^n = (a^2)^n b^n$. So consider the grammar $S \to aaSb \mid \varepsilon$.

L_2: The following grammar with start symbol T is built on top of the above grammar for L_1. $T \to AS \mid SB$, $A \to aA \mid a$, $B \to bB \mid b$.

Indeed both L_1 and L_2 are deterministic context-free.

6. Consider the two grammars G and G' with the start symbols S and S' and with the only productions:

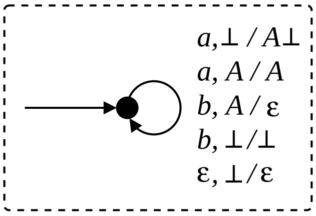
$$\text{Productions of } G \ : \ S \to aS \mid B, \quad B \to bB \mid b.$$
$$\text{Productions of } G' \ : \ S' \to aA' \mid bB', \quad A' \to aA' \mid B', \quad B' \to bB' \mid \varepsilon.$$

Which of the following statements is true?

- **(A)** $\mathscr{L}(G) = \mathscr{L}(G')$.
- **(B)** $\mathscr{L}(G)$ is strictly contained in $\mathscr{L}(G')$.
- **(C)** $\mathscr{L}(G')$ is strictly contained in $\mathscr{L}(G)$.
- **(D)** Neither $\mathscr{L}(G)$ is contained in $\mathscr{L}(G')$ nor $\mathscr{L}(G')$ is contained in $\mathscr{L}(G)$.

Solution $\mathscr{L}(G) = \mathscr{L}(a^*b^+)$, whereas $\mathscr{L}(G') = \mathscr{L}(a^+b^* + b^+)$. Take a string x of the form a^*b^+. If x does not contain a, then it is of the form b^+, and is covered by G'. If x contains a, then it is of the form a^+b^+, and is covered by a^+b^* in $\mathscr{L}(G')$. On the other hand, all strings of the form a^+ are generated by G' but not by G.

7. What is the language over the alphabet $\{a, b\}$, that is accepted by the following PDA? The PDA accepts by empty stack. Here, $\perp$ is the initial bottom marker for the stack.

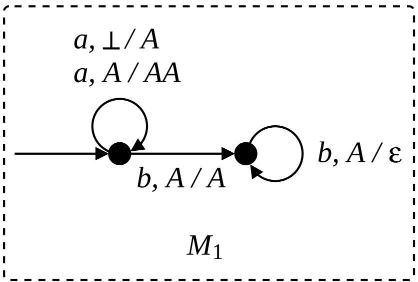


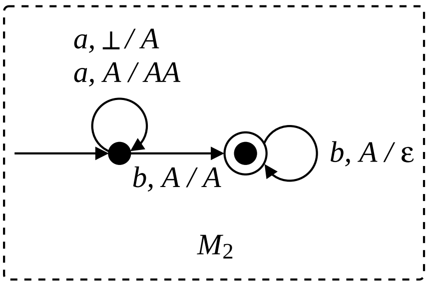
(A) $\{a^n b^n \mid n \geqslant 0\}$ **(B)** $\{a^m b^n \mid m, n \geqslant 0\}$ **(C)** $\{a^m b^n \mid m, n \geqslant 1\}$ **(D)** $\mathscr{L}\left((a+b)^* b\right)$

Solution There exist strings of the form $(a+b)^*b$ other than those specified in (A), (B), and (C).

Note: The correct answer is $\mathscr{L}\left(\varepsilon + (a+b)^* b\right)$.

8. Consider the following two PDA. M_1 accepts by empty stack, whereas M_2 accepts by final state. Which of the following strings is accepted by M_2 but not by M_1? The initial stack-bottom marker is $\perp$ for both the machines.





(A) ε **(B)** a^3b^3 **(C)** a^3b^4 **(D)** a^3b^5

Solution (A): Both M_1 and M_2 get stuck at the start state.

(B): After consuming the entire input a^3b^3, an A stays in the stack.

(C): After consuming the entire input a^3b^4, the stack is empty, so both M_1 and M_2 accept a^3b^4.

(D): Both M_1 and M_2 get stuck when the stack gets empty after consuming a^3b^4.

9. Let L be the language of a right-linear grammar over some alphabet Σ. Which of the following statements is *false* about the complement $\sim L = \Sigma^* - L$?

(A) $\sim L$ can be generated by a CFG where every production is of the form $A \to aB$ or $A \to a$ for non-terminal symbols A, B and for $a \in \Sigma \cup \{\varepsilon\}$.

(B) $\sim L$ can be generated by a CFG where every production is of the form $A \to Ba$ or $A \to a$ for non-terminal symbols A, B and for $a \in \Sigma \cup \{\varepsilon\}$.

[(C)] It is possible that no PDA can have the language $\sim L$.

(D) A PDA with only one state can have the language $\sim L$.

Solution (C): L is regular, so $\sim L$ is regular too. Every regular language is context-free.

10. Let $L = \left\{x \in \{a,b,c,d\}^* \mid \#a(x) = \#b(x) \text{ and } \#c(x) = \#d(x)\right\}$. Which of the following languages is context-free but not deterministic context-free?

[(A)] $\{a,b,c,d\}^* \setminus L$ **(B)** $\{a,b,c\}^* \setminus L$ **(C)** L **(D)** L^*

Solution (A): $\sim L = \{a,b,c,d\}^* \setminus L = \left\{x \in \{a,b,c,d\}^* \mid \#a(x) \neq \#b(x) \text{ or } \#c(x) \neq \#d(x)\right\}$. An NPDA can guess which inequality to verify. Intuitively, this is not possible for a DPDA. For a proof, note that $\sim L$ is context-free, whereas L is not (see Part (C)), so $\sim L$ cannot be deterministic context-free.

(B): A string $x \in \{a,b,c\}^*$ is not in L if and only if either

(1) $\#a(x) \neq \#b(x)$

or

(2) $\#c(x) > 0$.

A DPDA can (deterministically) verify the inequality (1). If, during this process, it ever encounters a c, it can accept after reading the rest of the input, regardless of whether the inequality (1) holds or not.

(C): L is not context-free. Supply $a^kc^kb^kd^k$ to the pumping lemma, where k is a PLC.

(D): $L^* = L$.