CS320: Problems and Solutions for Day 7, Winter 2023

Problem 1 Let L be the set of all strings over $\{a, b\}$ whose length is not a prime number. Draw a state-transition graph of a finite automaton that accepts L. If such automaton does not exist, prove it.

Answer: Such a finite automaton does not exist, since L is not a regular language. To prove this, observe that if L was regular, so would be its complement $\overline{L}$, which is the set of all strings over $\{a, b\}$ whose length is a prime number. To prove that $\overline{L}$ is not regular, assume the opposite. Let k be the constant as in the Pumping Lemma. Let $n > k$ be a prime number; then $a^n \in \overline{L}$. In the "pumping" decomposition: $a^n = uvx$, let $j = |v|$; recall that $j > 0$, since the "pumping" substring cannot be empty. By the pumping, every word of the form uv^ix, $i \geq 0$, belongs to $\overline{L}$. However, $uv^ix = a^{n+(i-1)j}$. Select $i = n + 1$. Then:

$$uv^ix = a^{n+(i-1)j} = a^{n+nj} = a^{n(1+j)}$$

Since $n > k \geq 1$ and $j > 0$, the length of this word is a product of two integer numbers, each greater than 1. Hence, this length is not prime, and this word does not belong to $\overline{L}$, whence a contradiction.

Problem 2 Let L be the language generated by the context-free grammar $G = \{V, \Sigma, P, S\}$, where $\Sigma = \{a, b\}$, $V = \{S, A, B\}$, and the production set P is:

$$\begin{aligned} &S \to AB \\ &A \to aaAbbb \mid \lambda \\ &B \to aB \mid \lambda \end{aligned}$$

Draw a state-transition graph of a finite automaton that accepts L. If such automaton does not exist, prove it.

Answer: Such a finite automaton does not exist, since L is not a regular language. By inspection of the grammar, we conclude that:

$$L = \{a^{2n}b^{3n}a^m \mid m, n \geq 0\}$$

To prove that L is not regular, assume the opposite. Let k be the constant as in the Pumping Lemma. Let $n > k$; then $a^{2n}b^{3n} \in L$. In the "pumping" decomposition: $a^{2n}b^{3n} = uvx$, we have that $|uv| \leq k < n < 2n$, hence the "pumping" substring v consists entirely of a's, say $v = a^j$. Recall that $j > 0$, since the "pumping" substring cannot be empty. By the pumping, every word of the form uv^ix, $i \geq 0$, belongs to L. However, such a word has $3n$ occurrences of b and $2n + (i-1)j$ occurrences of a, whereas it should have $2n$ occurrences of a for $3n$ occurrences of b. Since $2n + (i-1)j > 2n$ whenever $i > 1$, this is a contradiction.

Problem 3 Let L be the set of all palindromes over alphabet $\{a, b, c, d\}$. (A palindrome is a string that is equal to its reversal.)

(a) Draw a state-transition graph of a finite automaton that accepts L. If such automaton does not exist, prove it.

Answer: Such automaton does not exist, since L is not regular.

To prove this, assume the opposite, that L is regular. Let k be the constant as in the Pumping Lemma for L. Let $m > k$; then $a^mba^m \in L$. In any "pumping" decomposition such that $a^mba^m = uvx$, we have: $|uv| \leq k < m$. Hence, the "pumping" substring v consists entirely of a's, say $v = a^\ell$. Recall that $\ell > 0$, since the "pumping" substring cannot be empty. By the pumping, every word of the form uv^ix, $i \geq 0$, belongs to L. However, such a word is of the form:

$$w_1 = a^{m+(i-1)\ell}ba^m$$

while its reversal is:

$$w_1^R = a^mba^{m+(i-1)\ell}$$

Since $m + (i-1)\ell > m$ whenever $i > 1$, word w_1 has more a's before the single b than w_1^R, and it must be that:

$$w_1 \neq w_1^R$$

Since w_1 is not equal to its reversal w_1^R, we conclude that w_1 is not a palindrome. Hence, $w_1 \notin L$, which is a contradiction.

(b) Write a complete formal definition of a context-free grammar that generates L. If such grammar does not exist, prove it.

Answer: $G = \{V, \Sigma, P, S\}$, where
$\Sigma = \{a, b, c, d\}$, $V = \{S\}$, and P is:

$$S \to aSa \mid bSb \mid cSc \mid dSd \mid a \mid b \mid c \mid d \mid \lambda$$

Problem 4 Let L be a language over alphabet $\Sigma = \{a, b\}$, defined as follows:

$$L = \{x \mid (\exists w \in \Sigma^*)(x = ww)\}$$

(a) Write a regular expression that defines L. If such regular expression does not exist, prove it.

Answer: Such regular expression does not exist, since L is not regular.

To prove this, assume the opposite, that L is regular. Let k be the constant as in the Pumping Lemma for L. Let $m > k$; then $a^m b a^m b \in L$. In any "pumping" decomposition such that $a^m b a^m b = uvx$, we have: $|uv| \leq k < m$. Hence, the "pumping" substring v consists entirely of a's, say $v = a^\ell$. Recall that $\ell > 0$, since the "pumping" substring cannot be empty. By the pumping, every word of the form uv^ix, $i \geq 0$, belongs to L. Consider the word $uvvx$, obtained by setting $i = 2$:

$$w' = a^{m+\ell} b a^m b$$

The length of w' is:

$$|w'| = m + \ell + 1 + m + 1 = 2m + 2 + \ell$$

Note that ℓ has to be even, since the entire length $|w'|$ has to be even—w' has to be a concatenation of two identical words. However, if we write w' as a concatenation of two words of equal length:

$$w' = w_1 w_2, \text{ where } |w_1| = |w_2|$$

we see that this length is:

$$|w_1| = |w_2| = \frac{1}{2} \cdot |w'| = \frac{1}{2} \cdot (2m + 2 + \ell) = m + 1 + \frac{\ell}{2}$$

However:

$$m + 1 + \frac{\ell}{2} \leq m + \ell, \text{ since } \ell \geq 2$$

and we see that the the left-hand word w_1 consists entirely of a's:

$$\begin{aligned} w_1 &= a^{m+1+(\ell/2)} \\ w_2 &= a^{(\ell/2)-1} b a^m b \end{aligned}$$

In contrast, the right-hand word w_2 contains two b's. Hence:

$$w_1 \neq w_2$$

meaning that w' is not a concatenation of two identical words, and cannot belong to L.

(b) Draw a state-transition graph of a finite automaton that accepts L. If such automaton does not exist, prove it.

Answer: Such automaton does not exist, since L is not regular. This is proved in part (a).

Problem 5 Let L be the set of all strings over alphabet $\{a, b, c\}$ whose length is even and two middle symbols are equal.

(a) Write a complete formal definition of a context-free grammar that generates L. If such a grammar does not exist, prove it.

Answer: $G = (V, \Sigma, P, S)$, where $\Sigma = \{a, b, c\}$,
$V = \{S, Z\}$, and P is:

$$\begin{aligned} S &\to ZSZ \mid aa \mid bb \mid cc \\ Z &\to a \mid b \mid c \end{aligned}$$

(b) Draw a state-transition graph of a finite automaton M that accepts L. If such an automaton does not exist, prove it.

Answer: The required regular expression does not exist, since this language is not regular. To prove this, assume the opposite, that L is regular. Let η be the constant as in the Pumping Lemma for L. Let $m > \eta$; then the string:

$$(ac)^m bb (ac)^m$$

belongs to L, as its length is equal to $2(m + 1)$ and the two middle symbols are b.

In any "pumping" decomposition such that $(ac)^m bb(ac)^m = uvx$, we have: $|uv| \leq \eta < m$. Hence, the "pumping" substring v is contained entirely in the segment containing letters a and c. Let ℓ be the length of the "pumping" substring v. Recall that $\ell > 0$, since the "pumping" substring cannot be empty. Moreover, it has to be the case that

$\ell \geq 2$, since ℓ has to be even, lest "pumping" once would produce a string of odd length, which is invalid. Hence we conclude that the "pumping" substring has one of the following two forms:

$$v = (ac)^k \text{ or } v = (ca)^k$$

for some $k = \ell/2 > 0$, whence the "pumped" part becomes:

$$v^i = (ac)^{ik} \text{ or } v^i = (ca)^{ik}$$

In both cases, letters a and c remain alternating, so that no two adjacent letters are equal in the entire segment to the left of the substring bb.

Let us "pump" once, to produce a word of the form:

$$(ac)^{m+k}bb(ac)^m \text{ where } k \geq 1$$

If $k = 1$, the two-letter substring in the middle of the word is cb; if $k > 1$, the two-letter substring in the middle of the word is either ca (if k is odd) or ac (if k is even.) In all cases, the two middle symbols are different, whence the contradiction.

Problem 6 **(a)** Give an example of **two context-free languages L_1 and L_2 whose union is not context-free.** Give a precise definition of both languages and explain your answer. If such languages do not exist, explain why.

Answer: Such languages do not exist, since the class of context-free languages is closed under union.

(b) Give an example of **two context-free languages L_1 and L_2 whose intersection is not context-free.** Give a precise definition of both languages and explain your answer. If such languages do not exist, explain why.

Answer:

$$\begin{array}{l} L_1 = \{a^n b^n c^m \mid n, m \geq 0\} \\ L_2 = \{a^n b^m c^m \mid n, m \geq 0\} \\ L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\} \end{array}$$

Whereas L_1 and L_2 are context-free, $L_1 \cap L_2$ is a canonical non-context-free language, easily proved to be so by the Pumping Lemma.

(c) Give an example of **two regular languages L_1 and L_2 whose intersection is not regular.** Give a precise definition of both languages and explain your answer. If such languages do not exist, explain why.

Answer: Such languages do not exist, since the class of regular languages is closed under intersection.

(d) Give an example of a **regular language L_1 that has a subset L_2 which is not context-free.** Give a precise definition of both languages and explain your answer. If such a language does not exist, explain why.

Answer: Let:

$$\begin{array}{l} L_1 = \boldsymbol{a^* b^* c^*} \\ L_2 = \{a^n b^n c^n \mid n \geq 0\} \end{array}$$

L_1 is certainly regular, since it has a regular expression. L_2 is not context-free, as stated in part (b). Evidently, $L_2 \subset L_1$.

(e) Give an example of an **infinite context-free language L_1 that is not regular but has a subset L_2 which is regular.** Give a precise definition of both languages and explain your answer. If such a language does not exist, explain why.

Answer: Let:

$$\begin{array}{l} L_1 = \{a^n b^n \mid n \geq 0\} \\ L_2 = \{ab\} \end{array}$$

L_1 is a canonical context-free language, easily shown not to be regular by the Pumping Lemma. L_2 is finite, and thereby certainly regular. Evidently, $L_2 \subset L_1$.

(f) Give an example of a **regular language L_1 that has a subset L_2 which is not regular.** Give a precise definition of both languages and explain your answer. If such a language does not exist, explain why.

Answer: Let:

$$\begin{array}{l} L_1 = \boldsymbol{a^* b^*} \\ L_2 = \{a^n b^n \mid n \geq 0\} \end{array}$$

L_1 is certainly regular, since it has a regular expression. L_2 is not regular, as stated in part (e). Evidently, $L_2 \subset L_1$.

Problem 7 Let:

$$L = \{a^i b^j c^k a^\ell \mid i = 2k,\ j = 2\ell,\ i, j, k, \ell \geq 0\}$$

Write a complete formal definition of a context-free grammar that generates L. If such grammar does not exist, prove it.

Answer: Such grammar does not exist, since L is not a context-free language. To prove this, assume the opposite, and let k be the constant as in the Pumping Lemma for L. Select a word $a^{2m}b^{2n}c^m a^n \in L$ such that $m > k$ and $n > k$. In any "pumping" decomposition: $a^{2m}b^{2n}c^m a^n = uvxyz$, the length of the "pumping window" vxy is not greater than k: $|vxy| \leq k < m$ and also $|vxy| \leq k < n$. Consider a non-empty, effectively "pumping" substring of the "pumping window", which is at least one of v, y. There are two cases: this "pumping" substring either falls within one of the four substrings containing a single letter: a^{2m}, b^{2n}, c^m, a^n, or it spans two such substrings—it is too short to extend through as many as three of them. In the first case, the "pumping" produces a surplus of occurrences of one letter, over against what should be the matching number of occurrences of another letter. In the second case, the "pumping" produces two kinds of letters that are out of sequence. In both cases, the word produced by "pumping" violates the given form of the strings in L, whence the contradiction.

Problem 8 Let:

$$L = \{a^i b^k c^j a^\ell \mid i = 2k + 3,\ j = 2\ell + 1,\ i, j, k, \ell \geq 0\}$$

Write a complete formal definition of a context-free grammar that generates L. If such grammar does not exist, prove it.

Answer: $G = \{V, \Sigma, P, S\}$, where $\Sigma = \{a, b, c\}$,
$V = \{S, A, B\}$, and P is:

$$\begin{aligned} &S \to AB \\ &A \to aaAb \mid aaa \\ &B \to ccBa \mid c \end{aligned}$$

Problem 9 **(a)** Give an example of a language that is not regular and does not have any regular subsets. Provide a precise definition of this language and explain your answer briefly. If such a language does not exist, explain why.

Answer: Impossible—$\emptyset$ is straightforwardly regular, but is a subset of any other language.

(b) Give an example of a language that is not context-free and does not have any context-free supersets. Provide a precise definition of this language and explain your answer briefly. If such a language does not exist, explain why.

Answer: Impossible—Σ^* is straightforwardly context-free, but is a superset of any other language.

(c) Give an example of a language that is not countable. Provide a precise definition of this language and explain your answer briefly. If such a language does not exist, explain why.

Answer: Impossible—every language is a subset of the set of all finite strings, and this set is countable.

Problem 10 Let:

$$L = \{a^{2i+1} b^j d^{3k+2} a^{\ell+1} \mid i = k, \text{ and } i, j, k, \ell \geq 0\}$$

(a) Write a regular expression that defines L. If such regular expression does not exist, prove it.

Answer: Such automaton does not exist, since:

$$L = \{a^{2i+1} b^j d^{3i+2} a^{\ell+1} \mid i, j, \ell \geq 0\}$$

and this language is not regular.

To prove this, assume the opposite, that L is regular. Let η be the constant as in the Pumping Lemma for L. Let $m > \eta$; then $a^{2m+1}d^{3m+2}a \in L$, since $a^{2m+1}d^{3m+2}a = a^{2m+1}b^0 d^{3m+2}a^{0+1}$. In any "pumping" decomposition such that: $a^{2m+1}d^{3m+2}a = uvx$, we have: $|uv| \leq \eta < m < 2m + 1$, hence uv is a prefix of a^{2m+1} and the "pumping" substring v consists entirely of a's, say $v = a^p$. Recall that $p > 0$, since the "pumping" substring cannot be empty. By the pumping, every word of the form $uv^i x$, $i \geq 0$, belongs to L. However, such a word has exactly $3m + 2$ occurrences of d and $2m + 1 + (i - 1)p$ occurrences of a. Whenever $i > 1$, it is true that $2m + 1 + (i - 1)p > 2m + 1$. Hence, in this case, such a word has more a's than is appropriate for its number of d's, and cannot belong to L, which is a contradiction.

(b) Write a complete formal definition of a context-free grammar that generates L. If such grammar does not exist, prove it.

Answer: $G = (V, \Sigma, P, S)$, where
$\Sigma = \{a, b, d\}$, $V = \{S, A, B, D\}$, and P is:

$$\begin{array}{l} S \to AB \\ A \to aaAddd \mid aDdd \\ D \to DD \mid \lambda \mid b \\ B \to aB \mid a \end{array}$$

(c) Is the grammar (if any) which you constructed in part (b) regular? Explain your answer.

Answer: No—the first production, for example, violates the prescribed form. More importantly, a regular grammar for L *cannot exist,* lest L would be regular, which is not the case, by the proof given in the answer to part (a).

Problem 11 Let:

$$L = \{a^i (bb)^{k+1} e^{\ell+2} (ddd)^{m+3} g^{n+4} h^j\}$$

where

$$k = 3s \text{ and } j = t \text{ and } m = s + t$$

for some non-negative integers $i, k, \ell, m, n, j, s, t$.

(a) Write a complete formal definition of a context-free grammar that generates L. If such a grammar does not exist, prove it.

Answer: The general template of the words in the language L is:

$$a^i (bb)^{3s+1} e^{\ell+2} (ddd)^{s+3} (ddd)^t g^{n+4} h^t$$

L is generated by the context-free grammar:
$G = (V, \Sigma, P, S)$, where $\Sigma = \{a, b, d, e, g, h\}$,
$V = \{S, A, B, E, D, H, J, K, L\}$, and the production set P is:

$$\begin{array}{l} S \to ABD \\ A \to aA \mid \lambda \\ B \to JJJBK \mid JEKKK \\ J \to bb \\ K \to ddd \\ E \to eE \mid ee \\ D \to LDh \mid H \\ L \to ddd \\ H \to gH \mid gggg \end{array}$$

(b) Write a regular expression that defines L. If such a regular expression does not exist, prove it.

Answer: This language is not regular, and there does not exist a regular expression that represents it. To prove this, we show that the Pumping Lemma does not hold for the language L.

First, note that if β is the number of b's in some word contained in L and if δ is the number of d's in the same word, then the general template requires:

$$\delta > \frac{\beta}{3}$$

since:

$$\delta = s + 3 + t \geq s + 3 > s + \frac{1}{3} = \frac{1}{3}(3s + 1) = \frac{\beta}{3}$$

Now, to prove that L is not regular, assume the opposite—that L is regular. Let η be the constant as in the Pumping Lemma for L. Let $s > \eta + 2$; then $(bbb)^{3s+1} ee (ddd)^{s+3} gggg \in L$, as it is obtained from the general template by setting $i = \ell = t = n = 0$.

In any "pumping" decomposition such that:

$$(bbb)^{3s+1} ee (ddd)^{s+3} gggg = uvx$$

we have: $|uv| \leq \eta < s$. Hence, the "pumping" substring v consists entirely of b's, say $v = b^p$. By the pumping, every word of the form $uv^q x$, $q \geq 0$, belongs to L. However, such a word is of the form:

$$w_1 = (bbb)^{3s+1+p(q-1)} ee (ddd)^{s+3} gggg$$

By selecting, say: $q = 6s + 1$, we obtain a word whose number of b's is equal to:

$$\beta = 3s + 1 + p(6s + 1 - 1) = 3s + 1 + 6sp > 3s(2p + 1) \geq 9s$$

where the last inequality holds because $p \geq 1$, as the "pumping" substring cannot be empty. However, the number of d's in this word is still:

$$\delta = s + 3 < s + 2s = 3s = \frac{\beta}{3}$$

(recall that $s > 2$, by the choice of s.) Since this word violates the required relationship between β and δ, we conclude that it does not belong to the language L, whence the contradiction.