CS 181: Formal Languages and Automata Theory **Spring 2021**

Homework 6 Solutions Corrected

Assigned: Tuesday 4 May *Due: Monday 10 May 9:00pm PDT*

Note Submission deadline is now 9:00pm.

Problem 1

Let $\Sigma = \{a, b\}$. Consider the following language over Σ:

$$\{a^i b^j \mid 0 \leq i \leq j\}$$

Show a CFG for this language and briefly explain how your grammar is designed to correctly represent the language. You will receive extra credit if your grammar is unambiguous and you briefly explain why you believe that is true.

Solutions:

Ambiguous Solution:

$$S \to aSb|Sb|\epsilon$$

Brief explanation: The grammar generates the a's and b's in the right order and the two non-ϵ-moves guarantee that we generate at least one b for every a.

Note: It is ambiguous because we can generate extra b's (if any) at any point in the derivation. This is true even in a left-most derivation; so there can be more than one parse tree for any string that has symbol a's and extra symbol b's.

Unambiguous Solution:

$$\begin{aligned} S &\to TV \\ T &\to aTb|\epsilon \\ V &\to Vb|\epsilon \end{aligned}$$

Brief explanation: Variable T generates equal numbers of a and b while variable V generates a suffix with any number of b's. Variable T generates exactly one pair of a and b each time the first rule is applied. Variable V generates the extra b's (if any) similarly. So any combination of "i" and "j" in the definition of the language can only be generated by one parse tree using i applications of the rule for T and $j - i$ applications of the rule for V.

Problem 2

Let $\Sigma = \{0, 1\}$, and let:

$$L = \{xx^Rx \mid x \in \Sigma^*\}$$

Show using the Pumping Lemma for CFLs (and possibly other results), that this language is not a CFL.

Solution:
To make the answer easier to understand, we will divide the proof into two parts. First, we will prove a lemma which will help us with the CFL Pumping Lemma proof. Specifically, it will allow us to easily identify which strings of a certain form (0's followed by 1's followed by 0's followed by 1's) are part of L. Then we will continue with the main proof using that lemma.

Lemma 1 *Suppose $s = 0^h1^i0^j1^k$, with $i, j, k, h \geq 0$. Then there exists a w such that $s = ww^Rw$ if and only if $j = 2h$ and $i = 2k$.*

Proof: To prove the first direction, suppose that $j = 2h$ and $i = 2k$ and we need to show there is some w such that $s = ww^Rw$. We let $w = 0^h1^k$, so $s = ww^Rw$.
Going the other direction, suppose that there is some w such that $s = ww^Rw$, and we would like to show that $j = 2h$ and $i = 2k$. We begin by observing that w must have the form 0^a1^b for some $a, b \geq 0$. To see this, notice that w must start with a 0 (because s starts with a 0), and it must end with a 1 (because s ends with a 1), and can not have any 0's mixed in with 1's (or 1's mixed with the 0's) as that would not generate an s of our desired form.
Similarly, we can see that w must begin with h 0's immediately followed by a 1 (since s does), and must end with k 1's immediately preceded by a 0 (since s does). Putting these facts together means that the only possible matching choice is $w = 0^h1^k$. So if $s = ww^Rw$, expanding it out means that $j = 2h$ and $i = 2k$, which completes our proof. ∎

We now proceed with the proof that L is not a CFL.
Proof: Suppose for contradiction that L is a CFL. Let $p \geq$ be the pumping length, and choose $s = 0^p1^{2p}0^{2p}1^p$. Since $|s| > p$, by pumping lemma we can write $s = uvxyz$, with $|vy| \geq 1$, and $|vxy| \leq p$. Now, consider the cases of where vxy could be. There are 4 total "blocks" of the same symbol in s, each of length at least p, so vxy can only span either 1 or 2 adjacent blocks in s, giving us 7 total cases. We number the cases in order from left to right, according to where vxy begins. In every case, pumping down (that is, consider the string $s' = uv^0wx^0y \in L$ by pumping lemma) gives the following contradictions:

1. $s' = 0^i1^{2p}0^{2p}1^p$ where $i < p$ since $|vy| \geq 1$. By Lemma 1 $s' \notin L$ since $2i < 2p$.

2. $s' = 0^i1^j0^{2p}1^p$ where $i < p$ or $j < 2p$ since $|vy| \geq 1$. In both cases by Lemma 1 $s' \notin L$ since $2i < 2p$ or $j < 2p$.

3. $s' = 0^p1^i0^{2p}1^p$ where $i < 2p$ since $|vy| \geq 1$. By Lemma 1 $s' \notin L$ since $i < 2p$.

4. $s' = 0^p 1^i 0^j 1^p$ where $i < 2p$ or $j < 2p$ since $|vy| \geq 1$. In both cases by Lemma 1 $s' \notin L$ since $i < 2p$ or $j < 2p$.

5. $s' = 0^p 1^{2p} 0^i 1^p$ where $i < 2p$ since $|vy| \geq 1$. By Lemma 1 $s' \notin L$ since $i < 2p$.

6. $s' = 0^p 1^{2p} 0^i 1^j$ where $i < 2p$ or $j < p$ since $|vy| \geq 1$. In both cases by Lemma 1 $s' \notin L$ since $i < 2p$ or $2j < 2p$.

7. $s' = 0^p 1^{2p} 0^{2p} 1^i$ where $i < p$ since $|vy| \geq 1$. By Lemma 1 $s' \notin L$ since $2i < 2p$.

In all cases of $s = uvxyz$ we have a contradiction, so L is not a CFL. ■

Problem 3

Let Σ be an alphabet with at least two symbols. Consider the following languages over Σ:

R and S are a FSLs
C is a language which can be represented by a PDA
G is a language which can be represented by an unambiguous CFG
A is a language which can be represented by an ambiguous CFG
I is an inherently ambiguous CFL
L is a language which *cannot* be represented by a CFG
X, given that $X \cup S = L$

Given only the information above, classify each of the following languages by stating exactly which of the following four families it could possibly be:

1 FSL | 2 CFL and Not FSL | 3 Inherently Ambiguous CFL | 4 Non-CFL

For this set of classification questions, provide a *very brief* explanation, but you do *not* need to provide any detailed justifications, proofs, examples, nor counter-examples.

Examples:

$R \cap S$: It must be 1. The Family of FSLs is closed under intersection.

$R \cup I$: It could be a CFL; so it could be 1, 2, or 3. I is a CFL. And since R is a FSL, it is also a CFL. The Family of CFLs is closed under union.

$C \cup L$: It could be a non-CFL; so it could be 1, 2, 3, or 4. C is a CFL, but since L is not a CFL we cannot apply any closure properties to narrow-down the possibilities.

a. $R \cap L$

b. A

c. $\bar{C}$ (the complement of C with respect to Σ^*)

d. G

e. $\bar{L}$ (the complement of L with respect to Σ^*)

f. X

Solution:

a. $R \cap L$: it can be 1, 2, 3, 4.

   Brief explanation: L is not a CFL, so it is also not a FSL. Therefore, the fact that the family of FSLs is closed under intersection does not apply here. So we cannot say anything about $R \cap L$ based upon that closure property nor based upon any properties of the family of CFLs. Therefore, we cannot narrow-down the possibilities for this language at all.

   Notes for further study:

   Can you find an example of a FSL, R, such that you can force $R \cap L$ to be a FSL for *any* L? Hint: There are some examples of such an R which are very simple and should be familiar to you.

   Can you find examples of a FSL, R, and a non-CFL, L, such that their intersection is a CFL?

b. A can be 1, 2, 3

   Brief explanation: If A can be represented by an ambiguous CFG, then A is a context-free language. Since the family of FSLs is a subset of the family of context-free languages, then A can also be a FSL; so it could be 1 or 2. The fact that A can be represented by an ambiguous grammar says nothing about whether there might also be an unambiguous grammar for it. Thus, we do not know that A is an inherently ambiguous language, but we also cannot rule out the possibility that it is 3 inherently ambiguous.

c. $\bar{C}$ can be 1, 2, 3, 4

   Brief explanation corrected: Since C can be represented by a PDA, C could be a CFL or a FSL. The family of FSLs is closed under complement, but the family of CFLs is not closed under complement. If C is a FSL, then $\bar{C}$ will be a FSL. But if C is a non-FSL CFL (in category 2 or 3), then we do not know whether $\bar{C}$ will be a CFL or a 4 non-CFL; so it could also be 2, 3, or 4.

d. G can be 1, 2

   Brief explanation: G can be represented by a CFG; so it is a CFL. Since G can be represented by an unambiguous CFG, it is not 3 inherently ambiguous language. The Venn diagram shown in lecture showed that the family of FSLs is a proper subset of the family of CFLs, and the FSLs are disjoint from the inherently ambiguous CFLs. Therefore, G could also be 1 FSL.

e. $\bar{L}$ can be 2, 3, 4

Brief explanation: The family of FSLs are closed under complement. If $\bar{L}$ were a FSL, so would be L. Since L is not a CFL, $\bar{L}$ cannot be a FSL either, and thus $\bar{L}$ cannot be a FSL. However, since the family of CFLs is not closed under complement, we cannot say anything further about $\bar{L}$.

f. X can be 4

Brief explanation: Since S is a FSL, it is both a FSL and a CFL. Since the family of FSLs and the family of CFLs are each closed under union, X cannot be a FSL or a CFL. Therefore, it can only be non-CFL.