CS 341 Homework 16
Languages that Are and Are Not Context-Free

1. Show that the following languages are context-free. You can do this by writing a context free grammar or a PDA, or you can use the closure theorems for context-free languages. For example, you could show that L is the union of two simpler context-free languages.

   (a) $L = a^n c b^n$

   (b) $L = \{a, b\}^* - \{a^n b^n : n \geq 0\}$

   (c) $L = \{a^m b^n c^p d^q : n = q,$ or $m \leq p$ or $m + n = p + q\}$

   (d) $L = \{a, b\}^* - L_1$, where L_1 is the language $\{babaabaaab\ldots ba^{n-1}ba^n b : n \ n \geq 1\}$.

2. Show that the following languages are not context-free.

   (a) $L = \{ a^{n^2} : n \geq 0\}$

   (b) $L = \{www : w \in \{a, b\}^*\}$

   (c) $L = \{w \in \{a, b, c\}^* : w$ has equal numbers of a's, b's, and c's$\}$

   (d) $L = \{a^n b^m a^n : n \geq m\}$

   (e) $L = \{a^n b^m c^n d^{(n+m)} : m, n \geq 0\}$

3. Give an example of a context free language ($\neq \Sigma^*$) that contains a subset that is not context free. Describe the subset.

4. What is wrong with the following "proof" that $a^n b^{2n} a^n$ is context free?

      (1) Both $\{a^n b^n : n \geq 0\}$ and $\{b^n a^n : n \geq 0\}$ are context free.

      (2) $a^n b^{2n} a^n = \{a^n b^n\}\{b^n a^n\}$

      (3) Since the context free languages are closed under concatenation, $a^n b^{2n} a^n$ is context free.

5. Consider the following context free grammar: $G = (\{S, A, a, b\}, \{a, b\}, R, S)$, where $R = \{$

      $S \rightarrow aAS$

      $S \rightarrow a$

      $A \rightarrow SbA$

      $A \rightarrow SS$

      $A \rightarrow ba \quad \}$

  (a) Answer each of the following questions True or False:

     (i) From the fact that G is context free, it follows that there is no regular expression for L(G).

     (ii) L(G) contains no strings of length 3.

     (iii) For any string $w \in L(G)$, there exists u, v, x, y, z such that $w = uvxyz$, $|vy| \geq 1$, and $uv^n xy^n z \in L(G)$
         for all $n \geq 0$.

     (iv) If there exist languages L1 and L2 such that $L(G) = L1 \cup L2$, then L1 and L2 must both be context
         free.

     (v) The language $(L(G))^R$ is context free.

  (b) Give a leftmost derivation according to G of aaaabaa.

  (c) Give the parse tree corresponding to the derivation in (b).

  (d) Give a nondeterministic PDA that accepts L(G).

6. Show that the following language is context free:

      $L = \{xx^R yy^R zz^R : x, y, z \in \{a, b\}^*\}$.

7. Suppose that L is context free and R is regular.
   (a) Is L - R necessarily context free?
   (b) Is R - L necessarily context free?

8. Let $L_1 = \{a^n b^m : n \geq m\}$. Let $R_1 = \{(a \cup b)^* :$ there is an odd number of a's and an even number of b's$\}$. Show a pda that accepts $L_1 \cap R_1$.

Solutions

1.   **(a)** $L = a^n c b^n$. We can easily do this one by building a CFG for L. Our CFG is almost the same as the one we did in class for $a^n b^n$:

   $S \rightarrow aSB$
   $S \rightarrow c$

   (b) $L = \{a, b\}^* - \{a^n b^n : n \geq 0\}$. In other words, we've got the complement of $a^n b^n$. So we look at how a string could fail to be in $a^n b^n$. There are two ways: either the a's and b's are out of order or there are not equal numbers of them. So our language L is the union of two other languages:

- $L_1 = (a \cup b)^* - a^* b^*$ (strings where the a's and b's are out of order)
- $L_2 = a^n b^m \ n \neq m$ (strings where the a's and b's are in order but there aren't matching numbers of them)

L_1 is context free. We gave a context-free grammar for it in class (Lecture 12). L_2 is the complement of the regular language $a^* b^*$, so it is also regular and thus context free. Since the union of two context-free languages is context free, L is context free.

   (c) $L = \{a^m b^n c^p d^q : n = q$ or $m \leq p$ or $m + n = p + q\}$. This one looks scary, but it's just the union of three quite simple context-free languages:

$L_1 = a^m b^n c^p d^q : n = q$
$L_2 = a^m b^n c^p d^q : m \leq p$
$L_3 = a^m b^n c^p d^q : m + n = p + q$

You can easily write context-free grammars for each of these languages.

   (d) $L = \{a, b\}^* - L_1$, where L_1 is the language $\{babaabaaab\ldots ba^{n-1}ba^n b : n \ n \geq 1\}$. This one is interesting. L_1 is not context free. But its complement L is. There are two ways to show this:

1. We could build a PDA. We can't build a PDA for L_1: if we count the first group of a's then we'll need to pop them to match against the second. But then what do we do for the third? L is easier though. We don't need to check that all the a groups are right. We simply have to find one that is wrong. We can do that with a nondeterministic pda P. P goes through the string making sure that the basic $b(a^+ b)^+$ structure is followed. Also, it nondeterministically chooses one group of a's to count. It then checks that the following group does not contain one more a. If any branch succeeds (i.e., it finds a mismatched pair of adjacent a groups) then P accepts.

2. We could use the same technique we used in (b) and (c) and decompose L into the union of two simpler languages. Just as we did in (b), we ask how a string could fail to be in L_1 and thus be in L. The answer is that it could fail to have the correct $b(a^+ b)^+$ structure or it could have regions of a's that don't follow the rule that each region must contain one more a than did its predecessor. Thus L is the union of two languages:

- $L_2 = (a \cup b)^* - b(a^+ b)^+$
- $L_3 = \{xba^m ba^n by \in \{a, b\}^* : m+1 \neq n\}$.

It's easy to show that L_2 is context free: Since $b(a^+ b)^+$ is regular its complement is regular and thus context free. L_3 is also context free. You can build either a CFG or a PDA for it. It's very similar to the simpler language $a^n b^m \ n \neq m$ that we used in (b) above. So $L_1 = L_2 \cup L_3$ is context free.

2. (a) $L = \{a^{n^2} : n \geq 0\}$. Suppose $L = \{a^{n^2} : n \geq 0\}$ were context free. Then we could pump. Let $n = M^2$. So w is the string with M^{2^2}, or M^4, a's.) Clearly $|w| \geq K$, since $M > K$. So uvvxyyz must be in L (whatever v and y

are). But it can't be. Why not? Given our w, the next element of L is the string with $(M^2+1)^2$ a's. That's $M^4 + 2M^2 + 1$ (expanding it out). But we know that $|vxy| \leq M$, so we can't pump in more than M a's when we pump only once. Thus the string we just created can't have more than $M^4 + M$ a's. Clearly not enough.

(b) L = {www : w ∈ {a, b}*}. The easiest way to do this is not to prove directly that L = {www : w ∈ {a, b}*} is not context free. Instead, let's consider L1 = L ∩ a*ba*ba*b. If L is context free, L1 must also be. L1 = {$a^n b a^n b a^n b$: n ≥ 0}. To show that L1 is not context free, let's choose w = $a^M b a^M b a^M b$. First we observe that neither v nor y can contain b, because if either did, then, when we pump, we'd have more than three b's, which is not allowed. So both must be in one of the three a regions. We consider the cases:

      (1, 1) That group of a's will no longer match the other two, so the string is not in L1.
      (2, 2)   "
      (3, 3)   "
      (1, 2) At least one of these two groups will have something pumped into it and will no longer match the
              one that is left out.
      (2, 3)   "
      (1, 3) excluded since $|vxy| \leq M$, so vxy can't span the middle region of a's.

(c) L = {w ∈ {a, b, c}*. Again, the easiest thing to do is first to intersect L = {w ∈ {a, b, c}* : w has equal numbers of a's, b's, and c's} with a regular language. This time we construct L1 = L ∩ a*b*c*. L1 must be context free if L is. But L1 = $a^n b^n c^n$, which we've already proven is not context free. So L isn't either.

(d) L = {$a^n b^m a^n$: n ≥ m}. We'll use the pumping lemma to show that L = {$a^n b^m a^n$: n ≥ m} is not context free. Choose w = $a^M b^M a^M$. We know that neither v nor y can cross a and b regions, because if one of them did, then, when we pumped, we'd get a's and b's out of order. So we need only consider the cases where each is in one of the three regions of w (the first group of a's, the b's, and the second group of a's.)

      (1, 1) The first group of a's will no longer match the second group.
      (2, 2) If we pump in b's, then at some point there will be more b's than a's, and that's not allowed.
      (3, 3) Analogous to (1, 1)
      (1, 2) We must either (or both) pump a's into region 1, which means the two a regions won't match, or,
              if y is not empty, we'll pump in b's but then eventually there will be more b's than a's.
      (2, 3) Analogous to (1, 2)
      (1, 3) Ruled out since $|vxy| \leq M$, so vxy can't span the middle region of b's.

(e) L = {$a^n b^m c^n d^{(n+m)}$: m, n ≥ 0}. We can show that L = {$a^n b^m c^n d^{(n+m)}$: m, n ≥ 0} is not context free by pumping. We choose w = $a^M b^M c^M d^{2M}$. Clearly neither v nor y can cross regions and include more than one letter, since if that happened we'd get letters out of order when we pumped. So we only consider the cases where v and y fall within a single region. We'll consider four regions, corresponding to a, b, c, and d.
(1, 1) We'll change the number of a's and they won't match the c's any more.
(1, 2) If v is not empty, we'll change the a's and them won't match the c's. If y is nonempty, we'll change the number of b's and then we won't have the right number of d's any more.
(1, 3), (1, 4) are ruled out because $|vxy| \leq M$, so vxy can't span across any whole regions.
(2, 2) We'll change the number of b's but then we won't have the right number of d's.
(2, 3) If v is not empty, we'll change the b's without changing the d's. If y is not empty, we'll change the c's and they'll no longer match the a's.
(2, 4) is ruled out because $|vxy| \leq M$, so vxy can't span across any whole regions.
(3, 3) We'll change the number of c's and they won't match the a's.
(3, 4) If v is not empty, we'll change c's and they won't match a's. If y is not empty, we'll change d's without changing b's.
(4, 4) We'll change d's without changing a's or b's.

3. Let $L = \{a^n b^m c^p : n = m$ or $m = p\}$. L is clearly context free. We can build a nondeterministic PDA M to accept it. M has two forks, one of which compares n to m and the other of which compares m to p (skipping over the a's). $L1 = \{a^n b^m c^p : n = m$ and $m = p\}$ is a subset of L. But $L1 = a^n b^n c^n$, which we know is not context free.

4. (1) is fine. (2) is fine if we don't over interpret it. In particular, although both languages are defined in terms of the variable n, the scope of that variable is a single language. So within each individual language definition, the two occurrences of n are correctly interpreted to be occurrences of a single variable, and thus the values must be same both times. However, when we concatenate the two languages, we still have two separate language definitions with separate variables. So the two n's are different. This is the key. It means that we can't assume that, given $\{a^n b^n\}\{b^n a^n\}$, we choose the same value of n for the two strings we choose. For example, we could get $a^2 b^2 b^3 a^3$, which is $a^2 b^5 a^3$, which is clearly not in $\{a^n b^{2n} a^n\}$.

5. (a)   **(i)** False, since all regular languages are also context free.
        (ii) True.
        (iii) False. For example $a \in L$, but is not long enough to contain pumpable substrings.
        (iv) False.
        (v) True, since the context-free languages are closed under reversal.
  (b) $S \Rightarrow a\underline{A}s \Rightarrow a\underline{S}SS \Rightarrow aa\underline{S}S \Rightarrow aaa\underline{S} \Rightarrow aaaa\underline{A}S \Rightarrow aaaaba\underline{S} \Rightarrow aaaabaa$.
  (c)
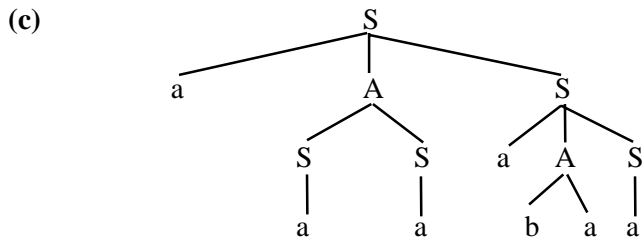
  (d) $M = (\{p, q\}, \{a, b\}, \{a, b\}, p, \{q\}, \Delta)$, where $\Delta =$
      $\{((p, \varepsilon, \varepsilon), (q, S)), ((q, a, a), (q, \varepsilon)), ((q, b, b), (q, \varepsilon)), ((q, \varepsilon, S), (q, aAS)), ((q, \varepsilon, S), (q, a)),$
        $((q, \varepsilon, A), (q, SbA)), ((q, \varepsilon, A), (q, SS)), ((q, \varepsilon, A), (q, ba))\ \}$

6. The easy way to show that $L = \{xx^R yy^R zz^R : x, y, z \in \{a, b\}^*\}$ is context free is to recall that we've already shown that $\{xx^R : x \in \{a, b\}^*\}$ is context free, and the context-free languages are closed under concatenation. But we can also do this directly by giving a grammar for L:
      $S \rightarrow AAA$
      $A \rightarrow aAa$
      $A \rightarrow bAb$
      $A \rightarrow \varepsilon$

7. (a) L - R is context free. $L - R = L \cap R'$ (the complement of R). R' is regular (since the regular languages are closed under complement) and the intersection of a context-free language and a regular language is context-free, so L - R is context free.
  (b) R - L need not be context free. $R - L = R \cap L'$. But L' may not be context free, since the context-free languages are not closed under complement. (The deterministic ones are, but L may not be deterministic.) If we let $R = \Sigma^*$, then R - L is exactly the complement of L.

8. M_1, which accepts $L_1 = (\{1, 2\}, \{a, b\}, \{a\}, \Delta, 1, \{2\})$, $\Delta =$
           $((1, a, \varepsilon), (1, a))$
           $((1, b, a), (2, \varepsilon))$
           $((1, \varepsilon, \varepsilon), (2, \varepsilon)$
           $((2, b, a), (2, \varepsilon)$

((2, ε, a), (2, ε)

M$_2$, which accepts R$_1$ = ({1, 2, 3, 4}, {a, b}, δ, 1, {2}), δ =

(1, a, 2)
(1, b, 3)
(2, a, 1)
(2, b, 4)
(3, a, 4)
(3, b, 1)
(4, a, 3)
(4, b, 2)

M$_3$, which accepts L$_1$ ∩ R$_1$ = ({(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2,), (2, 3), (2, 4)}, {a, b}, {a}, Δ, (1,1), {(2,2)}), Δ =

(((1, 1), a, ε), ((1, 2), a))
(((1, 1), b, a), ((2, 3), ε))
(((1, 2), a, ε), ((1, 1), a))
(((1, 2), b, a), ((2, 4), ε))
(((1, 3), a, ε), ((1, 4), a))
(((1, 3), b, a), ((2, 1), ε))
(((1, 4), a, ε), ((1, 3), a))
(((1, 4), b, a), ((2, 2), ε))

(((2, 1), b, a), ((2, 3), ε))
(((2, 2), b, a), ((2, 4), ε))
(((2, 3), b, a), ((2, 1), ε))
(((2, 4), b, a), ((2, 2), ε))

(((1, 1), ε, ε), ((2, 1), ε))
(((1, 2), ε, ε), ((2, 2), ε))
(((1, 3), ε, ε), ((2, 3), ε))
(((1, 4), ε, ε), ((2, 4), ε))
(((2, 1), ε, a), ((2, 1), ε))
(((2, 2), ε, a), ((2, 2), ε))
(((2, 3), ε, a), ((2, 3), ε))
(((2, 4), ε, a), ((2, 4), ε))