Syntax and Semantics: Exercise Session 8

Exercise 1.
Consider the following three languages over the alphabet $\Sigma = \{a, b, c\}$

$$L_1 = \{a^i b^j c^k \mid i, j, k \geq 0,\ i = j = k\}$$
$$L_2 = \{a^i b^j c^k \mid i, j, k \geq 0,\ i = j\}$$
$$L_3 = \{a^i b^j c^k \mid i, j, k \geq 0,\ j = k\}$$

1. Find out which are the two context-free languages and prove that they are actually so.

2. Prove that one language is non context-free.

3. Conclude, using the results of the exercise, that the class of context-free languages is *not* closed under $\cap$.

Solution 1.

1. L_2 and L_3 are context free languages, and are generated by the following grammars

$$L_2\colon S \to TC \qquad\qquad L_3\colon S \to AT$$
$$C \to \varepsilon \mid cC \qquad\qquad A \to \varepsilon \mid aA$$
$$T \to \varepsilon \mid aTb \qquad\qquad T \to \varepsilon \mid bTc$$

2. Assume that L_1 is context free and let $p \geq 0$ be its pumping length. Consider the string $s = a^p b^p c^p$. Clearly $|s| \geq p$ and $s \in L_1$. Consider an arbitrary split of s of the form $uvxyz$. By $|vy| > 0$ we have that v and y cannot be both empty. By $|vxy| \leq p$ we have that the substring vxy cannot contain both a's and c's since this would imply $|ab^p c| \leq p$. Therefore we have the following cases:

- vxy is of the form σ^k for some $\sigma \in \{a, b, c\}$ and $k \le p$ (i.e., vxy only contains only one symbol). Then $uv^0xy^0z \notin L_1$ because it has less σ's than the other symbols.
- vxy has both $a's$ and $b's$. Then $uv^0xy^0z \notin L_1$ because it has less a's or b's than c's.
- vxy has both $b's$ and $c's$. Then $uv^0xy^0z \notin L_1$ because it has less b's or c's than a's.

This proves that L_1 is not context-free.

3. Note that $L_1 = L_2 \cap L_3$. This proves that context-free languages are not closed under intersection.

   Note that this also proves that context-free languages are not closed by complement. Indeed since context-free languages are closed by (finite) union, if we assume that they are also closed by complement we obtain that $\big((L_2)^c \cup (L_3)^c\big)^c$ is context-free. By DeMorgan Laws, this would correspond to say that $L_2 \cap L_3$ is context-free, leading to a contradiction.

Exercise 2.
Consider the language

$$L_4 = \{u\#w \mid u, w \in \{0, 1\}^*,\ u \text{ is a prefix of } w\}$$

1. Find two examples of strings which belong to L_4 and other two strings that are *not* elements of L_4.

2. Is L_4 context-free? Justify your answer by means of a proof.

Solution 2.
1. $\{\#0,\ 1\#1,\ 0\#011,\ \ldots\} \subseteq L_4$ and $\{0,\ 01,\ 0\#100,\ \ldots\} \cap L_4 = \emptyset$

2. Assume that L_4 is context-free and that $p \ge 0$ is its pumping length. Let $s = 0^p1^p\#0^p1^p$. Clearly, $|s| \ge p$ and $s \in L_4$. Consider an arbitrary split of s of the form $uvxyz$. By $|vy| > 0$, v and y cannot be both empty. Let us consider the following cases:

   - v or y contains $\#$. Then $uv^0xy^0z \notin L_4$ because the string has no symbol $\#$.

- v and y are both to the left of #. Then $uv^2xy^2z \notin L_4$ because the substring to the left of # is longer than that to the right, therefore the former cannot be a prefix of the latter.
- v and y are both to the right of #. Then $uv^0xy^0z \notin L_4$ because the substring to the right of # is shorter than that to the left, therefore the latter cannot be a prefix of the former.
- v and y are respectively to the left and to the right of #. By $|vxy| \leq p$ we have that if v is not empty it has only 1's and if y is not empty it has only 0's. Then $uv^2xy^2z \notin L_4$ because the substring to the left of # has more 1's than that to the right or the substring to the right has more 0's than that to the left.

This proves that L_4 is not context-free.

Exercise 3.

Here is a wrong attempt to show that a language is not context-free.

Consider the language

$$L_5 = \{w \mid \text{there exists } w_1 \in \{a,b\}^* \text{ such that } w = w_1w_1\}\,.$$

We use the pumping lemma to show that L_5 is not context-free. Choose $s = aabbaabb$. So we can choose $u = aa$, $v = b$, $x = baa$, $y = bb$ and $z = \varepsilon$. But then we have that $uv^2xy^2z \notin L_5$.

1. Explain the main reasons why the above attempt of proof is wrong.
2. Prove that L_5 is not context-free. (It's a good idea not to try to fix the wrong proof!).

Solution 3.

1. Some reasons are the following
   - no mentioning of the fact that L_5 is assumed to be context-free
   - no pumping length p is taken into consideration.
   - the length of s may not be greater than or equal to p.
   - only a single split is considered, rather than studying all possible splits of the form $uvxyz$.

- only one condition out of the 3 conditions of the Pumping Lemma has been checked. The condition $|vxy| \leq p$ may not hold.

2. Assume L_5 to be context-free and let $p \geq 0$ be its pumping length. Let $s = a^p b^p a^p b^p$. Clearly $|s| \geq p$ and $s \in L_5$. Consider an arbitrary split of s of the form $uvxyz$ such that $|vy| > 0$ and $|vxy| \leq p$. By $|vy| > 0$, v and y cannot be both empty. Let σ and σ' be respectively the first and last symbol of vxy. We consider the following cases:

$\sigma = \sigma'$) By $|vxy| \leq p$ we have that $vxy = \sigma^k$ for some $k \leq p$, meaning that vxy is entirely contained in one of the two substrings of the form a^p or one of the form b^p. Consider now the string $s' = uv^2xy^2z$ and assume (without loss of generality) that vxy was within the first sequence of a's. Then s' is of the form $a^k b^p a^p b^p$ for some $k > p$. The string $s' \notin L_5$, indeed if we assume that there exists $w \in \{a, b\}^*$ such that $s' = ww$, then w has to start with a and end in b because s' does so. The only way to do so is to cut s' exactly when the first sequence of b's ends. But this implies that $a^k b^p = a^p b^p$ which is impossible since $k > p$.

$\sigma \neq \sigma'$) Let $s' = uv^0xy^0z$. Consider the following 3 sub-cases according to where vxy is positioned in s.

  - If vxy is entirely contained in the first half of the string s. Then string s' is of the form $a^h b^k a^p b^p$ with $0 < h + k < 2p$ (note that if $h = 0$ or $k = 0$, then $\sigma = \sigma'$). If we assume that $s' \in L_5$ then there exists $w \in \{a, b\}^*$ such that $s' = ww$. Hence w has to start with a and end in b because s' does so. The only way to do so is to cut s' exactly when the first sequence of b's ends, but this implies that $a^h b^k = a^p b^p$ which is impossible since $h + k < 2p$.
  - If vxy is entirely contained in the second half of s. Analogously, to the previous case one can show that $s' = a^p b^p a^h b^k$ for some $h, k > 0$ such that $h + k < 2p$, proving that $s' \notin L_5$.
  - If vxy is in between the first and the second half of s. Analogously to the previous cases, one can show that $s' = a^p b^h a^k b^p$ for some $h, k > 0$ such that $h + k < 2p$, proving that $s' \notin L_5$.

Exercise 4.

Here is an attempt to write a strategy for using the Pumping Lemma for context-free languages to prove that a language in *not* context-free. Underline places where the strategy does something wrong or where it is used a wrong terminology or notation, and then type the correct proposed strategy.

1. Let L be a grammar.
2. Set the pump length $p = 2$ and select a string $s \in L$ so that $|s| = 2$.
3. Select a single split of s of the form $uvxyz$ and show that this split does not comply the conditions 1–3 of the Pumping Lemma, i.e., that we do not have
   (a) uv^ixy^i for $i \geq 0$
   (b) $vx > 0$
   (c) $|vxy| > p$
4. Therefore we can conclude that L is not context-free.

Solution 4.

In the following you can see the wrong parts highlighted in red.

1. Let L be a grammar.
2. Set the pump length $p = 2$ and select a string $s \in L$ so that $|s| = 2$.
3. Select a single split of s of the form $uvxyz$ and show that this split does not comply the conditions 1–3 of the Pumping Lemma, i.e., that we do not have
   (a) uv^ixy^i for $i \geq 0$
   (b) $vx > 0$
   (c) $|vxy| > p$
4. Therefore we can conclude that L is not context-free.

The correct procedure to use the Pumping Lemma for context-free languages is the following.

1. Assume L to be a context-free language.
2. Let $p \geq 0$ be a pump length for L and select a string $s \in L$ so that $|s| \geq p$.
3. Select an arbitrary split of s of the form $uvxyz$ and show that this split does not comply the conditions 1–3 of the Pumping Lemma, i.e., that we do not have
   (a) $uv^ixy^iz \in L$ for $i \geq 0$
   (b) $|vy| > 0$
   (c) $|vxy| \leq p$
4. Therefore we can conclude that L is not context-free.

Exercise 5.
Here is a version of the Pumping Lemma for context-free languages. Unfortunately, there are several errors in the formulation. Underline each error you may find. Then write the correct statement of the Pumping Lemma.

A language L is context-free if and only if there exists a $p > 0$ such that there exists $s \in L$ where $|s| > p$ such that for all divisions of s of the form $uvwxyz$ it holds that

1. uv^ixy^iz
2. $|vxy| \geq p$
3. $|vx| = 0$

Solution 5.
In the following you can see the wrong parts highlighted in red.

A language L is context-free if and only if there exists a $p > 0$ such that there exists $s \in L$ where $|s| > p$ such that for all divisions of s of the form $uvwxyz$ it holds that

1. uv^ixy^iz
2. $|vxy| \geq p$
3. $|vx| = 0$

The correct statement of the Pumping Lemma can be found in the book.