Solution to Problem 15

Task. Prove that the language consisting of all the words that have equal number of +'s and −'s and exactly two times as many opening parentheses is not context-free.

Solution. Let us prove that this language is not context-free.

The proof will be by contradiction. Let us assume that this language is context-free. Then, by pumping lemma, there exists an integer p such that every word w from the language L whose length is at least p can be represented as $w = uvxyz$, where:

- $\text{len}(vxy) \le p$;
- $\text{len}(vy) > 0$, and
- for all natural numbers i, the word uv^ixy^iz also belongs to the language L.

Let us take the word

$$w = +^p \, -^p \, (^{2p} = + \ldots + - \ldots - (\ldots (,$$

where + and − are repeated p times each, and (is repeated $2p$ times. The length of this word – i.e., the number of symbols in this word – is equal to $p + p + 2p = 4p$. Clearly, $4p \ge p$, so, according to the Pumping Lemma, this word can be described as $uvxyz$ with the above properties.

Where can the central part vxy of this word be? We know that the length $\text{len}(vxy)$ of this part cannot exceed p. Thus, it cannot contain all three types of symbols: +'s, −'s, and ('s – since then it would have to include all p letters − plus additional − and (symbols, so its length would have been larger than p. So, there are only 5 cases remaining for the location of the part vxy:

1. it can be in the +'s;
2. it can be in +'s and −'s;
3. it can be in −'s;
4. it can be in −'s and ('s; and
5. it can be in ('s.

Let us consider these cases one by one.

Case 1. If vxy is in $+$'s, this means that the parts v and y contain only $+$'s. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add $+$'s – but we do not add any $-$'s or ('s. In the original word $w = +^p\,-^p\,(^{2p}$, there were exactly as many $+$'s as $-$'s. When we add more $+$'s, the balance is disrupted. Since the language L only contains the words which have the exact same number of $+$'s and $-$'s, the word $uvvxyyz$ cannot belong to the language L.

Case 2. If vxy is in $+$'s and in $-$'s, this means that the parts v and y contain only $+$'s and $-$'s. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add $+$'s and $-$'s – but we do not add any ('s. In the original word $w = +^p\,-^p\,(^{2p}$, there were exactly two times as many ('s as $+$'s and as $-$'s. When we add more $+$'s and $-$'s, the balance is disrupted. Since the language L only contains the words which have exactly two times as many ('s than $+$'s, the word $uvvxyyz$ cannot belong to the language L.

Case 3. If vxy is in $-$'s, this means that the parts v and y contain only $-$'s. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add $-$'s – but we do not add any $+$'s or ('s. In the original word $w = +^p\,-^p\,(^{2p}$, there were exactly as many $-$'s as $+$'s. When we add more $-$'s, the balance is disrupted. Since the language L only contains the words which have exactly as many $-$'s as $+$'s, the word $uvvxyyz$ cannot belong to the language L.

Case 4. If vxy is in $-$'s and ('s, this means that the parts v and y contain only $-$'s and ('s. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add $-$'s and ('s – but we do not add any $+$'s. In the original word $w = +^p\,-^p\,(^{2p}$, there were exactly as many $+$'s as $-$'s and exactly two times more ('s than $+$'s. When we add more $-$'s and ('s, the balance is disrupted. Since the language L only contains the words which have the exact same number of $+$'s and $-$'s and two times more ('s, the word $uvvxyyz$ cannot belong to the language L.

Case 5. If vxy is in ('s, this means that the parts v and y contain only ('s. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add ('s – but we do not add any $+$'s or $-$'s. In the original word $w = +^p\,-^p\,(^{2p}$, there were exactly two times as many ('s as $+$'s. When we add more ('s, the balance is disrupted. Since the language L only contains the words which have exactly two times more ('s than $+$'s, the word $uvvxyyz$ cannot belong to the language L.

So, in all five cases, we get a contradiction. This means that the original assumption – that the language L is context-free – is wrong. Thus, the language L is not context-free.