Solution to Problem 13

Task. Let us denote the most frequent letter in a language by M, the second most frequent by S, and the third most frequent by T. (For example, in English, the most frequent letter is the letter E.) Prove that the language of all the sequences of letters M, S, and T in which there are twice more M's than S's and three times more M's than T's 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:

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

Let us take the word

$$w = M^{6p}S^{3p}T^{2p} = M \ldots MS \ldots ST \ldots T,$$

where M is repeated $6p$ times, S is repeated $3p$ times, and T is repeated $2p$ times. The length of this word – i.e., the number of symbols in this word – is equal to $6p + 3p + 2p = 11p$. Clearly, $11p \geq p$, 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 three different types of symbols: M's, S's, and T's – since then it would have to include all $3p$ symbols S plus additional M and T 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 M's;
2. it can be in M's and S's;
3. it can be in S's;
4. it can be in S's and T's;

5. it can be in T's.

Let us consider these cases one by one.

Case 1. If vxy is in the M'ss, this means that the parts v and y contain only M's. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add M's – but we do not add any S's or T's. In the original word $w = M^{6p}S^{3p}T^{2p}$, there was a balance between letters of the three types. When we add more M's, the balance is disrupted. Since the language L only contains the words for which the above proportions have to be satisfied, the word $uvvxyyz$ cannot belong to the language L.

Case 2. If vxy is in M's and in S's, this means that the parts v and y contain only M's and S's. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add M's and S's – but we do not add any T's. In the original word $w = M^{6p}S^{3p}T^{2p}$, there was the desired balance between numbers of letters of all three types. When we add more M's and/or S's, the balance is disrupted, so the word $uvvxyyz$ cannot belong to the language L.

Similarly, we can see that in the other 3 cases, we also 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.