Solution to Problem 13

Task. Prove that the language consisting of all the digit sequences that have equal numbers of 0s, 1s, 2s, and 3s is not context-free. For example, the sequences 0123, 3210, and 01233210 are in this language.

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:

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

Let us take the word

$$w = 0^p1^p2^p3^p = 0\ldots01\ldots12\ldots23\ldots3,$$

where 0 is repeated p times, 1 is repeated p times, 2 is repeated p times each, and 3 is repeated p times. The length of this word – i.e., the number of symbols in this word – is equal to $p + p + p + p = 4p$. 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 $\mathrm{len}(vxy)$ of this part cannot exceed p. Thus, it cannot contain three different types of symbols: 0s, 1s, and 2s (or 1s, 2s, and 3s) – since then it would have to include all p symbols 1 plus additional 0 and 2 symbols, so its length would have been larger than p. So, there are only 7 cases remaining for the location of the part vxy:

1. it can be in the 0s;
2. it can be in 0s and 1s;
3. it can be in 1s;
4. it can be in 1s and 2s;
5. it can be in 2s;

6. it can be in 2s and 3s; and

7. it can be in 3s.

Let us consider these cases one by one.

Case 1. If vxy is in 0s, this means that the parts v and y contain only 0s. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add 0s – but we do not add any 1s, 2s, or 3s. In the original word $w = 0^p1^p2^p3^p$, there were exactly as many 0s as 1s. When we add more 0s, the balance is disrupted. Since the language L only contains the words which have equal number of 0s, 1s, 2s, and 3s, the word $uvvxyyz$ cannot belong to the language L.

Case 2. If vxy is in 0s and in 1s, this means that the parts v and y contain only 0s and 1s. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add 0s and 1s – but we do not add any 2s or 3s. In the original word $w = 0^p1^p2^p3^p$, there was an equal number of 0s, 1s, 2s, and 3s. When we add more 0s and/or 1s, the balance is disrupted. Since the language L only contains the words which have the same number of 0s, 1s, 2s, and 3s, the word $uvvxyyz$ cannot belong to the language L.

Similarly, we can see that in the other 5 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.