Acta Univ. M. Belii
Math. no.1(1993), pp.33-38

A NOTE ON REGULAR LANGUAGES

Bohuslav Sivák

Abstract. In this paper there are constructed two non-regular languages satisfying the well-known necessary condition for regular languages and there is modified this necessary condition. It is not known if the new modification of the necessary condition is sufficient.

1. Introduction.

We are going to show that the wel-known necessary condition for regular languages (here: Theorem 2) can be replaced (without any change in the method of the proof) by certain more strong necessary condition (here: Theorem 3).

We shall use the following notations:

N ...the set of all non-negative integers

T^* ...the set of all strings of the form $x_1x_2\dots x_k$ with $k \in N,\ x_i \in T$, including the empty string **e**.
We shall assume that the set T is finite. The subsets of T^* are called **languages**.

w_1w_2 ...(the concatenation) if $w_1 = x_1\dots x_k$ and $w_2 = y_1\dots y_m$, then $w_1w_2 = x_1\dots x_ky_1\dots y_m$. (Special cases: $\mathbf{ee} = \mathbf{e}$, $\mathbf{e}w = w\mathbf{e} = w$.)

$|w|$ ...the length of the string w ($|\mathbf{e}| = 0$.)

a^n ...the string $aa\dots a$ (n-times, $a^o = \mathbf{e}$).

1991 *Mathematics Subject Classification.* 68Q45.
Key words and phrases. regular languages.

Remark. The notion of regular language can be defined by many ways. More precisely, there are many possible (pairwise non-equivalent) definitions of regular grammars. However, all these definitions yield the same class of languages.

2. Two conditions for regular languages.

Theorem 1. (An easy consequence of Theorem 3.1 in [4], resp. Theorem 1.3.3 in [5]. See also [1].) *Let T be a finite set and let L be a subset of the set T^*.*

a) *Put*
$R_L = \{[x, y] \mid L_x = L_y\}$, *where*
$L_z = \{t \mid zt \in L\}$.

(Here $x, y, z, t \in T^*$.) If the language L is regular then R_L is a right congruence on T^* such that L can be written in the form of the union of some classes of the equivalence relation R_L and this equivalence relation has a finite index.

b) *If R is a right congruence on T^* of a finite index and if the language L can be written in the form of the union of some classes of R, then L is regular.*

Remark. For each finite set T, the set T^* with the operation of concatenation is a free monoid over the set T. The "right congruence" is an equivalence relation R on the set T^* such that

$$xRy \quad \text{implies} \quad xyRyz \quad (x, y, z \in T^*).$$

The index of any equivalence relation R is the number of R-blocks.

Example 1. (by [5], Example 1.3.6.) Put $T = \{a, b\}$, $L = \{a^n b^n | n > 0\}$. Applying Theorem 1, we shall prove that the language L is not regular.

If L is an union of certain classes of a right congruence R on $\{a, b\}^*$ of a finite index p, then at least two of the strings

$$a, a^2, a^3, \dots, a^{p+1}.$$

are in the same class of R, say

$$a^i R a^j, 1 \le i < j \le p + 1.$$

Then (R is a right congruence!) it holds:

$$a^i b^i R a^j b^i,$$

a contradiction ($a^i b^i \in L$ but not $a^j b^i \in L$).

Lemma 1. *Put* $T = \{a, b\}$ *and put*
$L_1 = \{a^n b^n | n \geq 1\}$,
$L_2 = \{w | w \in T^*, \#_a(w) = \#_b(w) > 0\}$,
where $\#_t(w)$ *denotes the number of occurences of the symbol* t *in the string* w. *Then it holds:*

$$L_1 \subseteq L \subseteq L_2 \Longrightarrow L \text{ is not regular.}$$

Proof. Similarly as in Example 1, at least two of the strings $a, a_2, a_3, \ldots, a_{p+1}$ are in the same block of the right congruence R, say

$$a^i R a^j, 1 \leq i < j \leq p + 1.$$

Then it holds

$$w_1 = a^i b^i R a^j b^i = w_2,$$

a contradiction. (Here $w_1 \in L_1$ but not $w_2 \in L_2$.)

Theorem 2. (See [6], Theorem 3.6 with a fault in the last row.) *Let* L *be a regular language. Then there exists a constant* $p > 0$ *such that for every* $w \in L$, $|w| \geq p$, *the string* w *can be written in the form*

$$w = w_1 w_2 w_3,$$

where $0 < |W_2| \leq p$ *and for all* $i \in N$, $w_1 {w_2}^i w_3 \in L$.

Remark. In Theorem 2, the case $i = 0$ is included. In the proof of this Theorem there are used finite automata. The fundamental properties of regular languages and finite automata can be found, for instance, in [1], [2], [3].

Example 2. Let L_2 be a language from Lemma 1. We know that L_2 is not regular. However, the non regularity of this language can not be proved by a direct application of Theorem 2. In fact, the necessary condition is satisfied for $p = 3$. (Each string $w \in L_2$, $|w| \geq 3$, contains a substring identical to "*ab*" or "*ba*" and this substring can be used in the role of w_2 in Theorem 2.)

Remark. (See [5], Example 2.2.13.) It is known that the class of all regular languages is closed to the operation of the intersection. Using this fact and Theorem 2 we can easily prove the non-regularity of L_2. In fact, if we put

$$L_3 = \{a^i b^j | i \geq 1\},$$

then $L_2 \cap L_3 = L_1$ and the non-regularity of L_1 can be proved by Theorem 2.

3. A more strong condition of regular languages.

Theorem 3. *Let L be a regular language. Then there exists a constant $p > 0$ such that for every $w \in L$, if $|w| \geq p$ and w is written in the form*

$$w = x_1 u x_2, |u| = p,$$

then the string u can be wrtitten in the form

$$u = y_1 v y_2, |v| > 0,$$

in such a way that for each $i \in N$, $x_1 y_1 v^i y_2 x_2 \in L$.

Remark. Theorem 3 can be proved by the same method as Theorem 2. We are going to show that the necessary condition in Theorem 3 is more strong than the necessary condition in Theorem 2.

Example 3. Let us continue the Example 2. The language L_2 is non-regular but this fact can not be proved by a direct application of Theorem 2. On the other hand, it is possible to apply Theorem 3: it suffices to put

$$w = a^p b^p, x_1 = a^p, u = b^p, x_2 = \mathbf{e}.$$

Remark. By the same argument it can be proved the non-regularity of the language L_1 (see Lemma 1).

Example 4. Put $T = \{a, b\}$ and put

$$L = \{a^m (ab)^j b^k : m \geq k \geq 0, j > 0\}.$$

First we shall try to apply Theorem 2. However, the necessary condition is satisfied for $p = 3$. In fact, assume that

$$w = a^a (ab)^j b^k, m \geq k \geq 0, j > 0, m + 2j + k \geq 3.$$

There are 3 possibilities:

1) $j > 1$. Then it suffices to put

$$w_1 = a^m, w_2 = ab, w_3 = (ab)^{j-1} b^k.$$

2) $j = 1, k > 0$. Then $w = a^m (ab) b^k, m \geq k \geq 1$ and it suffices to put

$$w_1 = a^m, w_2 = ab, w_3 = b^k.$$

3) $j = 1, k = 0, m > 0$. Then $w = a^{m+1} b, m \geq 1$ and it suffices to put

$$w_1 = \mathbf{e}, w_2 = a, w_3 = a^m b.$$

Therefore, it is impossible to prove the non-regularity of the language L by a direct application of Theorem 2, but it suffices to apply Theorem 3 to the strings

$$x_1 = a^p(ab), u = b^p, x_2 = \mathbf{e}, w = ap(ab)b^p.$$

Remark. The language L from Example 4 is not regular but it is context-free. In fact, it is generated by the following context-free grammar:

$$S \longrightarrow aSb \,|\, aS \,|\, R,$$
$$R \longrightarrow Rab \,|\, ab.$$

(Here S is the starting non-terminal symbol.) The author does not know if the non-regularity of this language can be proved by a similar method as in the remark after the Example 2.

4. Two open problems.

In [4] many unsolvable problems concerning context-free grammars and languages can be found. For instance, it is unsolvable to decide if the language generated by arbitrary context-free grammar is regular. Therefore, at least one of the following two problems has the negative answer.

Problem 1. *Is it possible (for an arbitary context-free grammar G) to decide if the language generated by G satisfies the necessary condition from the Theorem 3?*

Problem 2. *Is it true that each context-free language satisfying the necessary condition from Theorem 3 is regular?*

References

[1] Nerode, A., *Linear automaton transformations*, Proc. AMS (9/1958), 115-125.

[2] Rabin, M. O., Scott, D., *Finite automata and their decision problems*, IBM J. Res (3/1959), 115-125.

[3] Chomsky, N., Miller, G. A., *Finite state languages.*, Inf. and Control **2** (1/1958), 91-112.

[4] Hopcroft, J. E., Ullman, J. D., *Formal languages and their relation to automata*, Addision-Wesley Publ. Comp. Reading, Massachusets, 1969; Slovak transl. Alfa Bratislava 1978.

[5] Chytil M., *Automaty a gramatiky*, SNTL, Praha, 1984. (czek)

[6] Molnár, Ľ., Češka, M., Melichar, B., *Gramatiky a jazyky*, Sntl/Alfa, Bratislava, 1987. (slovak)

Department of Computer Science, Matej Bel University, Tajovského 40, 975 49 Banská Bystrica, Slovakia
E-mail address: sivak@fhpv.umb.sk

(Received October 18, 1991)