INFORMATICA, 2000, Vol. 11, No. 4, 353–370

Commutation in Global Supermonoid of Free Monoids

Anna BROSALINA, Boris MELNIKOV
Department of Applied Mathematics, Ulyanovsk State University
L. Tolstoy 42, 432700 Ulyanovsk, Russia
e-mail: ufire@mv.ru, bormel@mail.ru

Received: April 2000

Abstract. This work is an attempt of generalization of the simple statement about the requirements of commutation of words for the case of languages. In the paper, the necessary condition for commutation of languages are obtained, and in the prefix case the necessary and sufficient conditions are obtained. It is important to note that the considered alphabets and languages can be infinite.

The possibilities of application of the obtained results are shown in the other problems of the theory of formal languages. The boundary problems for the further solution are formulated.

Key words: formal languages, commutation of languages, infinite and prefix languages.

1. Introduction

In many textbooks on the theory of formal languages (see, e.g., Salomaa, 1981), the following simple statement is considered.

PROPOSITION 1.1. $uv = vu$ if and only if

$$(\exists w \in \Sigma^*,\ k, l \in \mathbb{N}_0) \quad (u = w^k,\ v = w^l).$$

This work is an attempt of generalization of the sectional statement about the requirements of a commutation of words for the case of languages. Up to the extremity the considered problem is not decided, i.e., *in general case* for the use in other problems necessary *and sufficient* requirements of this equality are not obtained.

The yet not solved problems are formulated as hypotheses. In the present paper, the obtained necessary conditions are reduced, and for the prefix case the necessary and sufficient are.

In the previous publications of the authors (see, e.g., Melnikov, 1995), the special cases of the sectional statements were applied to the solution of equivalent problems in special subclasses of context-free languages class. In the present paper more composite statements are obtained – in particular, when the considered alphabets (or the languages over them) are infinite. The authors hope that the same criteria and also criteria obtained in this paper for more common tasks, can be used in some other problems of the formal languages theory. (Some of such problems were considered in (Melnikov, 1995). In the same papers, the subjects were formulated for further examination, i.e., a complex of bound problems.)

2. Preliminaries

Let us consider free monoid $(\Sigma, \cdot, e)$ and its supermonoid $(\mathcal{P}(\Sigma), \cdot, \{e\})$.

The definitions, connected with the formal languages theory, are used by Salomaa (1981), Aho *et al.* (1985). Let us consider also some other definitions used below.

Let $v = a_1 a_2 \dots a_n$, and for some $m \leqslant n$, we have $u = a_1 a_2 \dots a_m$. Then u is called the prefix of the word v. We shall write this fact as $u \in \mathrm{pref}\,(v)$ (if $m < n$ we shall write $u \in \mathrm{opref}\,(v)$).

Let U and V satisfy of condition:

$$(\forall u \in U)(\exists v \in U)(u \in \mathrm{opref}\,(v)).$$

Then if $U \in \mathrm{pref}\,(V)$ we shall write $U \overset{\subseteq}{\infty} V$.

Let U and V is infinite languages. If $U \subseteq \mathrm{opref}\,(V)$ and $V \subseteq \mathrm{opref}\,(U)$ we shall write $U \overset{\sim}{\infty} V$ and the languages U and V is equivalent in infinity.

Let us remark that below we shall denote infinite languages by large bold letters ($\mathcal{A}, \dots, \mathcal{D}$), and finite ones by $A \dots, D$.

Also let us designate $\|A\|_{\min}$ as the length of the shortest word of the language A.

Let us designate for arbitrary language A

$$\mathrm{pv}\,(A) = \{u \in A \mid (\forall v \in A)(u \notin \mathrm{opref}\,(v))\}\,.$$

For example, let above alphabet $\Delta = \{0, 1\}$ the languages

$$C = \{0, 100, 101, 111\}, \quad D = \{01, 110\}$$

be given. Then we have

$$\mathrm{pv}\,(D) = C$$

for this languages.

If $(\forall u, v \in A)(u \notin \mathrm{opref}\,(v))$, then we shall call set (language) A prefix set (language), we shall write this fact as Pr (A).

The code above alphabet A is any subset from A^*, which is the basis of a free monoid from A^*. I.e., $C \subseteq A^*$ is a code, if any equality

$$x_1 \dots x_m = y_1 \dots y_n,$$

where from

$$x_1, \dots, x_m, y_1, \dots, y_n$$

follows, that

$$m = n \quad \text{and} \quad x_1 = y_1, \ldots, x_n = y_n.$$

The code $C \subseteq A^*$ is termed prefix if no word from C is not the left divider of other word from C, i.e., $CA^+ \cap C = \varnothing$.

The code above the alphabet A is termed maximum, if it is not contained in any other code above A.

For any alphabet Δ, we shall consider the following recursion definition of a set of maximum prefix codes above Δ. Let us designate the defined set of languages mp (Δ);

– we consider that $\Delta \in \text{mp}\,(\Delta)$;

– for any $C \in \text{mp}\,(\Delta)$ and $B \subseteq C$, we suppose

$$(C \setminus B \cup B\Delta \in \text{mp}\,(\Delta))\,.$$

If the set D is such that for some of $C \in \text{mp}\,(\Delta)$ the condition $C \subseteq D$ is carried out, we shall write $D \in \text{mp}^{\,+}(\Delta)$.

Let us consider the examples for the two last definitions. Let the languages

$$C = \{\, 0, 100, 101, 11 \,\} \quad \text{and} \quad D = C \cup \{\, 01, 110 \,\}$$

be given above the alphabet $\Delta = \{\, 0, 1 \,\}$. Then

$$C \in \text{mp}\,(\Delta)$$

besides $C, D \in \text{mp}^{\,+}(\Delta)$.

For the alphabet $\Sigma = \{\, a, b \,\}$ and set $A = \{\, ab, bbc \,\}$ it is possible, for example to consider that $\Delta_A = \Delta$, and the morphism

$$h : \Delta_A^* \to \Sigma^*$$

is given thus:

$$h_A(0) = ab, h_A(1) = bba.$$

Then the following

$$\begin{aligned} h_A(C) &= \{\, ab, \quad bba\ ab\ ab, \quad bba\ ab\ bba, \quad bba\ bba \,\} \in \text{mp}\,(A), \\ h_A(D) &= \{ab, \quad bba\ ab\ ab, \quad bba\ ab\ bba, \quad bba\ bba, \\ &\qquad ab\,bba, \quad bba\,bba\,ab \,\} \in \text{mp}^{\,+}(A) \end{aligned}$$

is fulfilled.

Thus, we shall consider equality

$$A\,B = B\,A \tag{1}$$

on sets of languages (i.e., in global supermonoids of free monoids).

Let us mark that the considered sets A and B can contain e, because of it there arise problems, considered in Section 3. Moreover, if e is not contained in considered sets, the further proved statement 3.1 is trivial.

3. The Case of Arbitrary Finite Languages

Thus, at first we shall consider the case of finite languages. Without any proof, we shall mark the following obvious fact: If $AB = BA$,

$$(\forall k, l \in \mathbb{N}_0)\ (A^k B^l = B^l A^k).^1 \tag{2}$$

PROPOSITION 3.1. If $AB = BA$, $B^* \overset{\subseteq}{\infty} A^*$.

Proof. As we have said before, we shall consider in this section the case of finite sets A and B only.[2]

Let us assume at first that $A \not\ni e$. Let us consider any word $v \in B^*$, choose $l \in \mathbb{N}_0$, for which $v \in B^l$, designate $k = |v| + 1$. The requirement

$$v \in \text{opref}\,(B^l A^k)$$

is carried out ($e \notin A$, consiquently $e \notin A^k$, therefore in the latter case it is really possible to write "opref", and not just "pref"). According to the marked above, $B^l A^k = A^k B^l$, therefore

$$v \in \text{opref}\,(A^k B^l).$$

As $e \notin A$, we have

$$\|A^*\|_{\min} \geqslant k > |v| \qquad v \in \text{opref}\,(A^k).$$

We surveyed an arbitrary word $v \in B^*$, hense at $e \notin A$ the condition

$$B^* \subseteq \text{opref}\,(A^*)$$

(i.e., $B^* \overset{\subseteq}{\infty} A^*$) is carried out. Further in the proved statement it is considered that $A \ni e$.

Let us assume that $B^* \not\subseteq \text{opref}\,(A^*)$, i.e.,

$$(\exists v \in B^*)\ (v \notin \text{opref}\,(A^*)).$$

[1] Generally, all the concatanations from $k + l$ of languages, among which A is met k times and B is met l times, are equal.

[2] Hence, it is possible to consider also the considered alphabet Σ finite.

Let us choose $l \in \mathbb{N}$ so that $v \in B^l$, and consider the language

$$\mathcal{D} = \{\, w \in A^* \mid w \in \text{pref}\,(v) \,\} .$$

Let us remark that

$$v \notin \text{opref}\,(A^*),$$

hence $v \notin D$, therefore it is possible by the alternate mode to define $\mathcal{D}$ as

$$\mathcal{D} = \{\, w \in A^* \mid w \in \text{opref}\,(v) \,\} .$$

Let us choose any word $u \in \max(\mathcal{D})$, let

$$u \in A^{k_1} \quad \text{and} \quad |u| = m$$

(thus we shall choose the numbers k_1 and m). As $A \neq \{e\}$, that at any $i < j$ the inequality takes place

$$\left|A^i\right| < \left|A^j\right| .$$

Therefore it is possible to pick some $k_2 \in \mathbb{N}$, such that

$$\left|A^{k_2}\right| > (m+1)\cdot \left|B^l\right| .$$

Let us designate

$$k = \max(k_1, k_2).$$

Let us remark that according to requirements $e \in A$ and $u \in A^{k_1}$ the word u enters each of languages A^i at $i \geqslant k_1$, consiquently $A^k \ni u$.

Let us designate

$$v = ua_1a_2 \ldots a_n ,$$

where

$$a_1, a_2, \ldots, a_n \in \Sigma ,$$

and $n > 0$; let us define languages

$$C_1 = \{\, u, ua_1, ua_1a_2, \ldots, ua_1a_2 \ldots a_{n-1} \,\} ,$$

$$C_2 = v\Sigma^* , \qquad C = C_1 \cup C_2.$$

Let us consider language $D = A^k B^l$. According to the marked above, the equality $D = B^l A^k$ is also correct. Let us estimate the number of (various) elemets of set $D \cap C$ by two modes.

First, we shall consider an entry D as $A^k B^l$. All words of the type uw, where $w \in B^l$, can belong to $D \cap C$. As, according to the supposition $|u| = m$, the word u has exactly m various own prefixes (including e), and the condition

$$u'w \in D \cap C$$

for some $w \in B^l$ can be carried out for an arbitrary word u' of the last set ($u' \in \operatorname{opref}(u)$), the total number of elements of set $D \cap C$ is estimated *from above* as follows:

$$|D \cap C| \leqslant \left|B^l\right| + m\cdot \left|B^l\right| = (m+1)\cdot \left|B^l\right|. \tag{3}$$

(We really have counted all the words of the set $D \cap C$, since the requirement $u'' \in C$ at $u'' \in A^k$ and $u'' \neq u$ can not be carried out for the following reason. If u'' was included into language C_1, the mode of choice of u as an element of the set $\max(\mathcal{D})$ would be untrue, i.e.,

$$u'' \in A^* \quad \text{and} \quad |u''| > |u|.$$

And if condition $u'' \in C_2$ was satisfied, the condition $v \in \operatorname{opref}(A^*)$ would be also satisfied, but this contradicts the mode of choice v.)

Thus, there are no elements of language $D \cap C$, except for enumerated, i.e., the inequality (3) is carried out.

Second, we shall consider notation D as $B^l A^k$. Owing to $A \ni e$, the requirement $v \in B^l$ is carried out, therefore all the elements of set D of the type vw, where $w \in A^k$, belong also C_2, and hence C. Thus, the number of (various) elements of language $D \cap C$ is not less than the number of various elements A^k, and according to the requirement $k \geqslant k_2$ and the mode of the choice k_2, we have

$$|D \cap C| > (m+1)\cdot \left|B^l\right|. \tag{4}$$

Thus, the contradiction (3) and (4) is obtained, it proves the requirement

$$B^* \subseteq \operatorname{opref}(A^*)$$

i.e., $B^* \overset{\subseteq}{\infty} A^*$.

By double application of the statement 3.1 the following theorem is proved.

Theorem 3.1. *If* $AB = BA$, $A^* \overset{\sim}{\infty} B^*$.

Let us apply the theory from the paper (Melnikov, 1995) to any languages A and B, satisfying the requirement (1). We obtain that from the explored equality (1) and its corollary, i.e., the requirement $A^* \overset{\sim}{\infty} B^*$ (Theorem 3.1), it results that above the alphabet Σ^* there is some set C, such that

$$A, B \in \mathrm{mp}^+(C).$$

Or in an alternate statement:

$$\exists\, h_C^{-1}(A),\, h_C^{-1}(B) \in \mathrm{mp}^+(\Delta_C).$$

Let us remark that though for some $u \in A \cup B$ the set $h_C^{-1}(u)$ can contain more than one element (since the morphism h_C^{-1}, corresponding to language C, not without fall is injective), the languages

$$A' = h_C^{-1}(A) \quad \text{and} \quad B' = h_C^{-1}(B)$$

satisfy the relation

$$A'\,B' = B'\,A'. \tag{5}$$

The proof of the statement (5) will be carried out by contradiction by means of application of the morphism h_C to considered sets.

Thus, we can apply the special inverse morphism to sets A and B, satisfying the equality (1), getting the similar correlation (5) for new "decoded" sets A' and B'. Moreover, it is known that each of languages A' and B' contains some maximum prefix code above the alphabet Δ_C as subset. Some possible algorithms of choice of an approaching set C were discussed in (Melnikov, 1995; Melnikov, 1993).

And, as it was said earlier, some of the problems of the theory of formal languages, possible for application of criteria obtained in this section, were surveyed in (Melnikov, 1995). However in those papers, there is essential the fact, whether the considered languages may be infinite. Therefore in the next section we shall consider generalizing of the statements, proved in this one.

4. The Case of Arbitrary Infinite Languages

The statement considered in this section is the generalizing of the theorem proved above for the sets having some finite numbers of atoms. Now we shall prove it for the case of infinite sets.

As we said before, sets denoted by $\mathcal{A}$, $\mathcal{B}$ (etc.) may contain infinitely many words.

Let us designate as A_i a sublanguage consisting from words of the language A of length i.

For sets of infinite languages over the alphabet Σ, we'll define some special partial linear order. It can be considered as generalizing of simple relation "contains more words".

DEFINITION 4.1. Let us define linear order $\succ$ for finite and infinite alphabets

- Let the alphabet Σ be finite.
  We shall write $\mathcal{A} \succ \mathcal{B}$, if there exists n such that the following condition holds:

$$|\mathcal{A}_1| = |\mathcal{B}_1|,\ |\mathcal{A}_2| = |\mathcal{B}_2|,\ \dots,\ |\mathcal{A}_{n-1}| = |\mathcal{B}_{n-1}|,\ |\mathcal{A}_n| > |\mathcal{B}_n|.$$

- Let the alphabet Σ be infinite.
  We shall write $\mathcal{A} \succ \mathcal{B}$, if there exists n such that the following conditions holds:

$$\mathcal{A}_1 = \mathcal{B}_1 \quad \text{or} \quad (\mathcal{A}_1 \not\subset \mathcal{B}_1 \quad \text{and} \quad \mathcal{B}_1 \not\subset \mathcal{A}_1),$$

$$\mathcal{A}_2 = \mathcal{B}_2 \quad \text{or} \quad (\mathcal{A}_2 \not\subset \mathcal{B}_2 \quad \text{and} \quad \mathcal{B}_2 \not\subset \mathcal{A}_2),$$

$$\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots$$

$$\mathcal{A}_{n-1} = \mathcal{B}_{n-1} \quad \text{or} \quad (\mathcal{A}_{n-1} \not\subset \mathcal{B}_{n-1} \quad \text{and} \quad \mathcal{B}_{n-1} \not\subset \mathcal{A}_{n-1}),$$

$$\mathcal{A}_n \supset \mathcal{B}_n.$$

For example, for any alphabet Σ the language Σ^* "contains more words" than the language $(\Sigma^2)^*$. It is obvious that if $\mathcal{A}_n \subset \mathcal{B}_n$, then $\mathcal{A}_n \prec \mathcal{B}_n$.

Now we shall prove the fact, which is the generalizing of Proposition 3.1 for an infinite case. Let us remark that the condition of proposition is, of course, the same as in that case, but the alphabets and languages may be infinite.

PROPOSITION 4.1. If $AB = BA$, then $B^* \overset{\subset}{\infty} A^*$.

Proof. Suppose that $\mathcal{A} \not\ni e$. Considering any word $v \in \mathcal{B}^*$, we shall choose $l \in N_0$, for which $v \in \mathcal{B}^l$; we shall designate $k = |v| + 1$. Thus, following condition holds:

$$v \in \text{opref}\,(\mathcal{B}^l \mathcal{A}^k).$$

According to mentioned above, $\mathcal{B}^l \mathcal{A}^k = \mathcal{A}^k \mathcal{B}^l$, therefore

$$v \in \text{opref}\,(\mathcal{A}^k \mathcal{B}^l).$$

Since $e \notin \mathcal{A}$, we have

$$\|\mathcal{A}^*\|_{\min} \geqslant qk > |v| \quad \text{and} \quad v \in \text{opref}\,(\mathcal{A}^k).$$

Let us consider any word $v \in \mathcal{B}^*$. Since $e \notin \mathcal{A}$, the following condition holds:

$$\mathcal{B}^* \subseteq \text{opref}\,(\mathcal{A}^*).$$

(i.e., $\mathcal{B}^* \overset{\subset}{\infty} \mathcal{A}^*$.)

Thus, further in this statement we shall suppose that $\mathcal{A} \ni e$.

Let us suppose that $\mathcal{B}^* \not\subset \text{opref}\,(\mathcal{A}^*)$, i.e.,

$$(\exists v \in \mathcal{B}^*)\ (v \notin \text{opref}\,(\mathcal{A}^*)).$$

Let us choose $l \in N$ such that $v \in \mathcal{B}^l$. Let us consider also language

$$D = \{w \in A^* | w \in \text{pref}\,(v)\}.$$

By the method of construction D is finite.

Let us remark that

$$v \notin \text{opref}\,(\mathcal{A}^*),$$

then $v \notin D$, therefore it is possible to determine D by another way, i.e.,

$$D = \{w \in A^* | w \in \text{opref}\,(v)\}.$$

Since D is finite, it is possible to choose some word $u \in \max(D)$. Let the following condition holds:

$$u \in \mathcal{A}^{k_1} \quad \text{and} \quad |u| = m.$$

(We choose here the numbers k_1 and m.)

Since $\mathcal{A} \neq \{e\}$, for each $i < j$ the inequality holds:

$$\mathcal{A}^i \prec \mathcal{A}^j.$$

Therefore it is possible to choose some $k_2 \in N$, such that

$$|\mathcal{A}^{k_2}| \ > (m+1)\cdot |B^l|.$$

As we did in the proof of Proposiotion 3.1, denote

$$k = \max(k_1, k_2).$$

Let us remark that according to condidtions $e \in \mathcal{A}$ and $u \in \mathcal{A}^{k_1}$, the word u belongs to the languages $\mathcal{A}^i$ for each $i \geqslant k_1$, therefore $\mathcal{A}^k \ni u$.

Let us denote

$$v = ua_1a_2 \dots a_n\,,$$

where

$$a_1, a_2, \dots, a_n \in \Sigma\,,$$

and, obviously, $n > 0$. We'll define finite languages

$$\mathcal{C}_1 = \{u, ua_1, ua_1a_2, \ldots, ua_1a_2 \ldots a_{n-1}\},$$

$$\mathcal{C}_2 = v\Sigma^*, \qquad \mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2.$$

Let us consider the language $\mathcal{D} = A^k B^l$. According to Section 3, the equality $\mathcal{D} = \mathcal{B}^l \mathcal{A}^k$ also holds. Let us count the number of various elements of the set $\mathcal{D} \cap \mathcal{C}$ by two ways.

First, consider $\mathcal{D}$ as $\mathcal{D} = \mathcal{A}^k \mathcal{B}^l$. All the words of uw, where $w \in \mathcal{B}^l$, can belong to $\mathcal{D} \cap \mathcal{C}$. Therefore, according to the supposition $|u| = m$, the word u has equally M various own prefixes (including e), then the condition

$$u'w \in \mathcal{D} \cap \mathcal{C}$$

holds for any word u' of the last set (i.e., $u' \in$ opref (u)) and some $w \in \mathcal{B}^l$ is choosed by u'.

By the definition of the relation $\succ$, for the language $\mathcal{D} \cap \mathcal{C}$, we obtain the following conditions:

$$|\mathcal{D} \cap \mathcal{C}| > |\mathcal{B}^l| + m \cdot |\mathcal{B}^l| = (m+1) \cdot |\mathcal{B}^l|. \tag{6}$$

(We have really counted all the words of the set $\mathcal{D} \cap \mathcal{C}$, since the condition $u'' \in \mathcal{C}$, $u'' \in \mathcal{A}^k$ and $u'' \neq u$ cannot be true because of the following facts. If u'' belongs to the language $\mathcal{C}_1$, then we have a contradiction with the way of choosing u as the element of the set $\max(D)$, since

$$u'' \in \mathcal{A}^* \quad \text{and} \quad |u''| \succ |u|.$$

And if condition $u'' \in \mathcal{C}_2$ holds, we would have $v \in$ opref $(\mathcal{A}^*)$, and the last fact contradicts to the way of choosing v.)

Thus, there are no elements of the language $\mathcal{D} \cap \mathcal{C}$, except enumerated ones, i.e., the inequality (6) holds.

Second, consider the language $\mathcal{D}$ as $\mathcal{D} = \mathcal{B}^l \mathcal{A}^k$. Owing to $\mathcal{A} \ni e$, the condition $v \in \mathcal{B}^l$ is carried out, therefore all the elements of the set $\mathcal{D}$ of the type vw (where $w \in \mathcal{A}^k$) belong also to $\mathcal{C}_2$, and therefore to $\mathcal{C}$. Thus, the number of various elements of the language $\mathcal{D} \cap \mathcal{C}$ is not less than the number of various elements of $\mathcal{A}^k$. And according to the condition $k \geqslant k_2$ and the way of choosing k_2, we have

$$|\mathcal{D} \cap \mathcal{C}| > (m+1) \cdot |\mathcal{B}^l|. \tag{7}$$

Thus we have obtained the contradction of (6) to (7), therefore

$$\mathcal{B}^* \subseteq \text{opref}\,(\mathcal{A}^*),$$

i.e., $\mathcal{B}^* \overset{\subseteq}{\infty} \mathcal{A}^*$.

For an infinite case the following theorem also holds.

Theorem 4.1. *If* $AB = BA$*, then* $A^* \overset{\sim}{\infty} B^*$.

5. The Case of Prefix Languages

Below we shall consider other individual cases of equality (1), i.e., when one of considered languages contains no more than two words.

Let us consider equality (1) in the prefix case, i.e., let the requirements Pr (A) and Pr (B) be carried out.

Theorem 5.1. *If* $AB = BA$*,* Pr (A) *and* Pr (B)*,*

$$(\exists C \in \Sigma^*,\ k, l \in \mathbb{N}_0)\ (A = C^k, B = C^l).$$

Proof. If $A = B$, the proved statement is correct at

$$C = A = B \qquad k = L = 1.$$

And if $A = \{e\}$ (in the prefix case it is equivalent to that $A \ni e$), the statement is correct at

$$C = B\,, \quad k = 0\,, \quad l = 1\,;$$

similarly for $B = \{e\}$.

Considering these cases, in the below proved theorem we can expect that $A \neq B$, and none of these sets contains e (i.e., not equal to $\{e\}$). At least one of two following statements is carried out with this restriction:

$$(\exists u \in A,\ v \in B)\ (u(v)) \tag{8}$$

$$(\exists u \in A,\ v \in B)\ (v \in \text{opref}\,(u)) \tag{9}$$

(so far we do not confirm that exactly one of them) is true.

Let (8) be carried out (if (9), the reasonings are similar). Let us consider $w \in \Sigma^*$ such, that $v = uw$ for any words u and v chosen according to the requirement (8). The requirement

$$uwA \subseteq BA = AB\,,$$

is carried out, therefore $wA \subseteq B$.

We shall designate D the association of all words w, which can be chosen according to fixed $u \in A$ and some satisfying (8) word $v \in B$. According to the mentioned the above, $DA \subseteq B$; therefore in case $DA \neq B$ the requirement (1) is not fulfilled owing to Pr (B). Thus,

$$D \cdot A = B, \tag{10}$$

and the requirement (9) is not fulfilled still.

According to (1) and (10), the equality

$$ADA = DAA$$

is carried owt, therefore owing to Pr (A) the following is correct:

$$AD = DA \quad \text{and} \quad \text{Pr}\,(D).$$

Thus, the requirements of the proved theorem are fulfilled and for languages A and D (instead of A and B), and, as $e \notin A \cup B$, the following is also correct:

$$\|A \cup D\|_{\min} < \|A \cup B\|_{\min} \tag{11}$$

(we have used elementary inequality, linking numbers $\|A\|_{\min}$ in two languages).

Let us designate now

$$A_0 = A_1 = A\,, \quad B_0 = B\,, \quad B_1 = D.$$

If $A_1 \neq B_1$, so by applying all the explained above to languages A_1 and B_1, we receive some A_2 and B_2 etc. According to the introduced labels of languages A_i and B_i and to generalization of the statement (10) for all possible i the following inequalities are carried out

$$\|A_{i+1} \cup B_{i+1}\|_{\min} < \|A_i \cup B_i\|_{\min}\ ,$$

similar to (11). It follows from these inequalities that the described process of building-up of pairs sets A_i and B_i is finite, i.e., for some $m \in \mathbb{N}$ the following is correct:

$$A_0 \neq B_0,\ \ A_1 \neq B_1,\ \ldots,\ A_{m-1} \neq B_{m-1},\ \ A_m = B_m.$$

Let us designate

$$C = A_m = B_m.$$

According to the method of building-up C,

$$\{\, A_{m-1}, B_{m-1} \,\} = \{\, C, C^2 \,\}\ ,$$

hence

$$\{ A_{m-2}, B_{m-2} \} \subseteq \{ C, C^2, C^3 \} ,$$

and so on. Thus for some natural k and l the equalities

$$A = C^k , \quad B = C^l$$

are fulfilled. The case when one of the sets is equal to $\{e\}$ was surveyed earlier. Uniting these two cases, we obtain that for the language C chosen by us

$$(\exists k, l \in \mathbb{N}_0) \, (A = C^k, B = C^l) ,$$

as was to be proved.

Note. In case of the infinite alphabet Σ the proved theorem has the following numerical interpretation.

First, according to the Theorem 3.1 it is possible to consider that the sets A and B contain some maximum prefix codes (i.e.,

$$A, B \in \text{mp}^+ (\Sigma);$$

see the following subsection on this theme). And as we consider the case of prefix sets, A and B are maximum prefix codes, i.e.,

$$A, B \in \text{mp} (\Sigma).$$

Second, we shall rank the alphabet Σ, i.e., we shall consider any ration of the strict order "$\prec$": if

$$\Sigma = \{ a_1, a_2, \dots, a_n \} ,$$

then let us assume, for example,

$$a_1 \prec a_2 \prec \cdots \prec a_n .$$

By means of the entered relation "$\prec$" let us rank also words from Σ^*, i.e., we shall consider them in the lexicographic (alphabetic) order. We shall consider any languages here only in specified order, i.e., let our sets A and B be

$$A = (u_1, u_2, \dots, u_k) , \text{ where } u_1 \prec u_2 \prec \cdots \prec u_k ,$$

$$B = (v_1, v_2, \dots, v_l) , \text{ where } v_1 \prec v_2 \prec \cdots \prec v_l$$

(we use parentheses, as we consider A and B as serially ordered sets, i.e., the finite sequences). We shall designate the coresponding sequences of lengths of words

$$\mathcal{D}(A) = \Big(|u_1|, |u_2|, \ldots, |u_k|\Big) \quad \mathcal{D}(B) = \Big(|v_1|, |v_2|, \ldots, |v_l|\Big).$$

It is simple to be convinced that as the order of the letters of the alphabet is fixed beforehand, so the arbitrary set $A \in \text{mp}\,(\Sigma)$ is recovered in the unique fashion sequence of lengths $\mathcal{D}(A)$ (i.e., such map from the set of maximum prefix codes into the set of numerical sequences is an injection).

On the definition of maximum prefix codes it is simple to show that

$$AB \in \text{mp}\,(\Sigma)$$

(i.e., the set of maximum prefix codes is the undermonoid of considered global supermonoid of a free monoid). Therefore we can use label $\mathcal{D}(AB)$ in just particular sense. It is easily proved that

$$\begin{aligned}\mathcal{D}(AB) \;=\; \Big(&|u_1|\cdot|v_1|, \ldots, |u_1|\cdot|v_l|,\\ &|u_2|\cdot|v_1|, \ldots, |u_2|\cdot|v_l|,\ \ldots,\ |u_k|\cdot|v_1|, \ldots, |u_k|\cdot|v_l|\Big).\end{aligned}$$

Hence, condition (1) is equivalent to equality

$$\mathcal{D}(AB) = \mathcal{D}(BA).$$

6. Conclusion

In this section some interesting examples and problems for further solution are considered.

We shall consider some examples. In the prefix case all pairs of languages A and B, for which the requirement (1) is carried out, can be constructed on the basis of the Theorem 5.1. It is obvious that (1) is correct also for any languages A and B, each of which can be noted as

$$\bigcup_{i \in I} \Sigma^i$$

For some set of indexes $I \subseteq \mathbb{N}_0$.

Let us present a less obvious example nonprefix A and B:

$$\Sigma = \{0, 1\},$$

$$A = \Sigma \cup \{\, 00, 01, 11 \,\} \cup \Sigma^3, \quad B = \Sigma \cup \Sigma^2; \tag{12}$$

then

$$AB = BA = \bigcup_{2 \leqslant i \leqslant 5} \Sigma^i.$$

All examples (these and any others) can be changed by means of application to languages A and B of the same morphism, but it is not necessary injective.

Two hypotheses, for which the authors now have neither proofs, nor counterexamples, are formulated further.

The essence of the first hypothesis is: if (1) is fulfilled, that equality

$$\text{pv}\,(A)\cdot \text{pv}\,(B) = \text{pv}\,(B)\cdot \text{pv}\,(A) \tag{13}$$

is carried out.

Let us remark that for an arbitrary language

$$A \in \text{mp}^+\,(\Sigma)$$

according to the definition of the set mp the requirement

$$\text{pv}\,(A) \in \text{mp}\,(\Sigma)$$

is carried out, therefore if the first hypothesis is correct (i.e., if equality (13) for arbitrary commuting languages A and B is really fulfilled), according to the Theorem 3.1,

$$\text{pv}\,(A) = C^k \quad \text{and} \quad \text{pv}\,(B) = C^l$$

for some $C \subseteq \Sigma^*$ (of the prefix set) and numbers $k, l \in \mathbb{N}_0$.

On the basis of these equalities the second hypothesis is stated: with the marked labels it is possible to pick the set C thus, that $A, B \subseteq C^*$.

(We shall remark that, for example, pairs of infinite languages

$$A = \Sigma^2\Sigma^* \quad \text{and} \quad B = \Sigma^4\Sigma^*$$

for the arbitrary alphabet Σ are not counterexamples to the second hypothesis, as we can pick not only $C = \Sigma^2$, but also $C = \Sigma$.)

For the analysis of both formulated hypotheses we use expression:

$$\text{pv}\,(A)\cdot \text{pv}\,(B) \neq \text{pv}\,(A\cdot B). \tag{14}$$

For confirmation (14) we shall consider such an example:

$$\Sigma = \{0, 1\},\ A = B = \{\,0, 01, 100, 101, 110, 111\,\}.$$

Let us remark that here

$$A,\, B,\, AB \in \text{mp}^+(\Sigma).$$

The inequality (14) thus can be tested immediately.[3]

In summary let us mark that with the problem of exposition of criteria of equality $AB = BA$, and also with many problems, considered in previous sections and in (Melnikov, 1995; Melnikov, 1993), the problem "of taking the root" from some given language immediately is bound: for the given language $A \subseteq \Sigma^*$ it is required to find the greatest possible $n \in \mathbb{N}$ and depending on n language $B \subseteq \Sigma^*$, such that $A = B^n$. Up to the extremity this problem is not explored by the authors yet, therefore we shall adduce only one interesting example, having some analogy to the example (12): for the arbitrary alphabet Σ (including at $|\Sigma| = 1$ and $|\Sigma| = \omega$) from language

$$\bigcup_{2\leqslant i\leqslant 10} \Sigma^i$$

not only the "obvious" square root is extracted,

$$\bigcup_{1\leqslant i\leqslant 5} \Sigma^i$$

but also

$$\bigcup_{i\in\{\,1,2,4,5\,\}} \Sigma^i.$$

Thus, in case of languages, the operation of taking the root of a given degree is not a (unique) function.

References

Salomaa, A. (1981). *Jewels of the Formal Languages Theory*. Rockville.

Melnikov, B. (1995). Some equivalence problems for free monoids and for subclasses of the CF-grammars class. In *Number Theoretic and Algebraic Methods in Computer Science*. World Sci. Publ. pp. 125–137.

Aho A.V., R. Sethi, J.D. Ullman (1985). *Compilers: Principles, Techniques and Tools*. N.J., Addison-Wesley.

Melnikov B. (1993). The equality condition for infinite catenations of two sets of finite words. *Int. J. of Found. of Comp. Sci.*, **43**, 267–274.

[3] And it is possible to show almost without "evaluations", using the following reasons:

$$\text{pv}\,(A) = \text{pv}\,(B) = \{\,0, 100, 101, 110, 111\,\}$$

Thus,

$$\text{pv}\,(A)\cdot \text{pv}\,(B) \not\ni 010\,,$$

whereas pv $(AB) \ni 010$.

A.G. Brosalina. Since 1997–2000: post-graduate student of Ulyanovsk State University (Russia). 1991–1996: student of Novocherkassk State Technical University (Russia). Research interests: heuristical algorithms; finite automata, regular languages, theory of semigroups.

B.F. Melnikov. Since 1999: Professor of Ulyanovsk State University (Russia). 1995–1999: Associated Professor of Ulyanovsk State University. 1991–1995: Assistant Professor of Ulyanovsk Branch of Moscow State University. 1988–1991: post-graduate student of Moscow State University. 1984–1988: programmer in the Research Institute (ex-Leningrad, ex-USSR). Research interests: heuristical algorithms (genetic, branch-and-bounds etc.); artificial intelligence; finite automata, regular languages, theory of semigroups; mathematical aspects of the object-oriented programming.

Laisvųjų monoidų globaliojo supermonoido perstatiniai

Anna Brosalina, Boris Melnikov

Straipsnyje siekiama apibendrinti formaliosios kalbos žodžių perstatymo reikalavimus. Jame suformuluotos pakankamos formaliosios kalbos komutatyvumo sąlygos. Prefiksinėms kalboms suformuluotos taip pat ir būtinosios sąlygos. Rezultatai galioja taip pat ir begaliniams alfabetams bei begalinėms kalboms. Parodyta, kaip naudoti gautus rezultatus kai kurioms kitoms svarbioms formaliųjų kalbų teorijos problemoms spręsti.