The language of left brackets as a morphism of a maximum prefix code

B F Melnikov\textsuperscript{1}, S Yu Korabelshchikova\textsuperscript{2} and V V Dang\textsuperscript{3}

\textsuperscript{1} Shenzhen MSU – BIT University, 1 International University Park Road, Dayun New Town, Longgang District, Shenzhen, Guangdong Province, 518172, China
\textsuperscript{2} Northern Arctic Federal University named after M.V. Lomonosov, 17 Northern Dvina, Arkhangelsk, 163001, Russia
\textsuperscript{3} Vietnam National University, Ho Chi Minh city University of Technology, District 10, Ho Chi Minh city, Vietnam

E-mail: bf-melnikov@yandex.ru

Abstract. We investigate the solvability of the equivalence problem in various subclasses of the class of context-free languages. Another fundamental problem necessary for the material in this paper is also the special binary relation for a set of finite languages. Combining these two topics, we obtain in this paper the necessary and sufficient conditions for the equality of two bracketed languages for the prefix-and-suffix case.

1. Introduction

This article is a continuation of the publications that we completed during 1991–2019, see [1 – 6] etc. Some of these papers are not available on the Internet, therefore, we must repeat notation and some propositions of these papers below.

In most of these publications, we primarily investigate the solvability of the equivalence problem in various subclasses of the class of context-free languages. To do this, we consider languages representing morphisms of the so-called bracketed languages or simple bracketed languages. We especially note that bracketed languages may seem simpler than variants of the Dick language, [7] etc.; however, according to the authors, they have been widely used since 1960s in automation systems for building compilers. Among the works on this subject, we especially note the publication of one of the authors of this paper, i.e. [4].

The second fundamental problem necessary for the material in this paper is also the special binary relation for a set of finite languages that we have studied; such a relation is satisfied if and only if the infinite iterations of these languages are equal. In some of the mentioned papers, we have obtained, in particular cases, the conditions for such equality; moreover, these conditions are sometimes necessary, sometimes sufficient, and sometimes necessary and sufficient.

Combining these two topics, we obtain in this paper the necessary and sufficient conditions for the equality of two bracketed languages for the so-called prefix-and-suffix case.

2. Notation and preliminaries

Like previous publications, we denote the set of pairs of brackets $\mu$, usually with subscripts:
\[ \mu \in \Sigma^* \times \Sigma^* \]

for the given finite alphabet \( \Sigma \).

For the sets of left brackets, we shall write
\[ \alpha(\mu) = \{ u | (\exists v)(u, v) \in \mu \}. \]
\[ \alpha(\mu, v) = \{ u | (u, v) \in \mu \}; \]

similarity,
\[ \beta(\mu) = \{ v | (\exists u)(u, v) \in \mu \}, \beta(\mu, u) = \{ v | (u, v) \in \mu \}. \]

Language \( \text{pref}(u) \) is the whole set of prefixes of the word \( u \), \( \text{opref}(u) \) is the set of its proper prefixes. Similarly, for languages, instead of the word \( u \).

We shall write \( \text{Pr}(A) \) if the language \( A \) has the prefix property; similarly, \( \text{Su}(A) \) for the suffix property. If \( \text{Pr}(\alpha(\mu)) \) and \( \text{Su}(\beta(\mu)) \), then we shall write \( \text{PrSu}(\mu) \).

For the considered \( L \) and \( \mu \), we shall write:
\[ \Psi^0(\mu, L) = L \]
\[ \Psi^j+1(\mu, L) = \{ u\Psi^j(\mu, L)v | (u, v) \in \mu \} \]
for each \( j \geq 0 \),
\[ \Psi^*(\mu, L) = \bigcup_{j \geq 0} \Psi^j(\mu, L), \]
\[ \Psi^+(\mu, L) = \bigcup_{j \geq 1} \Psi^j(\mu, L). \]

Thus,
\[ \Psi^*(\mu, L) = \{ u\Psi^j(\mu, L)v | (u, v) \in \mu \} \]
\[ \cup \{ w | w \in L \}. \]

Also,
\[ s(\mu) = \Psi^1(\mu, \{e\}) = \{ uv | (u, v) \in \mu \} \]

For any language \( L \), we shall write
\[ \Psi^*_i = \Psi^*(\mu_i, L); \]

remark that \( L \) could be infinite. Similarly, for languages \( \Psi^+_i \) and \( \Psi^+_j \) for any \( j \in \mathbb{N} \). We shall denote
\[ \alpha_i(\mu) = \alpha(\mu, v); \]

similarly, \( \alpha_i, \beta_i \) and \( \beta_i(\mu) \) are defined. Notation \( A^i_u, A^i_u \), and \( A^i_u \) (for the given language \( A \) and word \( u \)) is defined by [1].

If some set (language or a lot of word pairs) is the same for both subscripts \( i \in \{1, 2\} \), then we shall sometimes denote such set without these subscripts. For example, when we have already proved that \( \Psi^1 = \Psi^2_2 \), then we can denote this set \( \Psi^2 \).

Thus, for some infinite language \( L \) and sets of word pairs \( \mu_1 \) and \( \mu_2 \) consider equality
\[ \Psi^*_1 = \Psi^*_2 \]

(1)

with the following restrictions:
\[ \text{PrSu}(L), \text{PrSu}(\mu_1), \]
\[ L \cap \Psi^*_i = \emptyset, \]
\[ \Psi^1 = \Psi^2 \text{ and } \Psi^1 = \Psi^2. \]

(2)

Note that due to conditions (1) and (2), the equality \( \Psi^+_1 = \Psi^+_2 \). We will consider all these restrictions to be fulfilled throughout this article; we will formulate and prove all statements, taking them into account.

Without loss of generality, we assume that \( \mu_i \) do not contain the pair \( \{e, e\} \). Note that, despite this, for both \( i \), one of the sets \( \alpha_i \) or \( \beta_i \) can be equal to \( \{e\} \). However, all these possibilities will be considered
Let us consider language $\mathcal{L}$, and $w = u_i z_i v_i$ for some $(u_i, v_i) \in \mu_i$ and $z_i \in L$. Then we shall write $\Psi_{\alpha_1}^{-1}(w) = (u_i, v_i)$.

Note that according to the condition $\text{PrSu}(\mu_1)$, for any word $w \in \Psi_1$ there exists an unique pair $(u_i, v_i) \in \mu_i$, such that $w = u_i z_i v_i$ for some $z_i \in L$. Thus, the last notation is correct: it really defines exactly one pair of words from $\mu_i$, not a set of such pairs.

3. The inverse morphism

**Proposition 1.** $\alpha_1^* \odot \alpha_2^*, (\beta_2^*)^* \odot (\beta_1^*)^*$.

*Proof.* At first, let us prove that $\alpha_1^* \subseteq \text{opref}(\alpha_2^*)$.

Let $u \in \alpha_1^*$, $|u| = m$, then consider $n = ||\Psi_2^1||_{\min} + 1$. After that, we shall select a word $w \in \Psi_2^1$, such that $u \in \text{opref}(w)$ (we can select such $w$, because condition $n > m$ holds, then $\varepsilon \notin \alpha_1$).

By (1), $w \in \Psi_2^1$, then by the method of selecting $n$, condition $w \in \Psi_2^1$ for some $p > m$ holds. Therefore, $w = u_1 \ldots u_p z v \ldots v_1$, where $(u_j, v_j) \in \mu_2$ and $z \in L$.

The following condition is evident:

$$|u_1 \ldots u_p| \geq p > m,$$

then $u \in \text{opref}(u_1 \ldots u_p)$, i.e. $u \in \text{opref}(\alpha_2^*)$. Therefore, $\alpha_1^* \subseteq \text{opref}(\alpha_2^*)$.

In the same way, we can prove $\alpha_1^* \subseteq \text{opref}(\alpha_2^*)$. Then, using definition of [1], we have $\alpha_1^* \odot \alpha_2^*$.

The second condition $(\beta_2^*)^* \odot (\beta_1^*)^*$ can be proved similarly.

By the condition $\text{PrSu}(\mu_1)$, Proposition 1 and [1; Theorem 1], there exist a prefix language $A$ and a suffix language $B$, such that the following condition holds:

$$\alpha_1 \in mp(A), \quad \beta_1 \in mp(B).$$

For some $u \in A^*$ and $U \subseteq A^*$, we shall write

$$u' = h_1^-(u), \quad U' = \{h_{\alpha_1}^-(u) \mid u \in U\}.$$

For the alphabet $\Delta_A$ defined by [1], we have $\alpha_1^* \in mp(\Delta_A)$. Below, we shall write $\Delta$ instead of $\Delta_A$.

4. The language of left brackets as a morphism of the maximum prefix code

Let us consider language $D = \{w \in \Psi_1^1 \mid \Psi_{\alpha_1}^{-1}(w) \neq \Psi_{\alpha_1}^{-1}(w)\}$.

If $D = \emptyset$, then $\mu_1 = \mu_2$ and (1) is evident; therefore, we shall suppose below that $D \neq \emptyset$.

Consider some word $y \in D$, let

$$\Psi_{\alpha_1}^{-1}(y) = (u_i, v_i).$$

Without loss of generality, we can suppose that

$$|u_1^*| - |u_2^*| = m > 0.$$

Suppose that at the same time the inequality

$$|v_1^*| \geq |v_2^*|$$

holds. In this case, let us select some word $z \in \min(L)$ and consider $x = u z v$. By the way of selecting $x$, we have $x \in \Psi_2^1$, then by the considered limitation, $x \in \Psi_1^1$, then there exist the words $\alpha_1^{-1}(x)$ and $\beta_1^{-1}(x)$. By the condition $\text{PrSu}(\mu_1)$, we have

$$|\alpha_1^{-1}(x)| > |u_2|, |\beta_1^{-1}(x)| \geq |v_1|,$$

Then for some $z' \in L$, such that

$$x = \alpha_1^{-1}(x) z' \beta_1^{-1}(x),$$
the inequality $|z'| < |z|$ holds. For the last condition, we obtain $z \notin \text{min}(L)$.

The obtained contradiction ($z \in \text{min}(L)$ and $z \notin \text{min}(L)$ shows, that (3) cannot be performed; i.e. for some $n > 0$, we have $|v'_n| - |v'_1| = n$.

Now, let us consider the languages

$$E = \alpha_1(v_1)|_{u_2} = \{ \omega | (u_2, \omega, v_1) \in \mu_1\}$$

and

$$F = \alpha_1|_{u_2} = \{ \omega | (\exists \nu | (u_2, \omega, \nu) \in \mu_1\}$$

(below, we shall not give such detailed definitions of the considered languages and sets of pairs). Let us also remark, that:

- $E \neq \emptyset$, because $(u_2, v_1) \in \mu_1$ and $u_2 \in \text{orpref}(u_1)$;
- $E \subseteq F$ by definitions of these sets;
- $E, F \subseteq A^+$; and, therefore, there exist the languages $E'$ and $F'$.

By proposition 1, $\alpha_1 \in \text{mp}(A)$, then by definition of maximum prefix codes, we have $F \in \text{mp}(A)$, i.e. $F' \in \text{mp}(A)$.

**Proposition 2.** $E = F$.

**Proof.** Consider the language

$$G = \Psi^+|_{u_2} |_{v_1}.\$$

Let us write it in the two following ways.

Considering sets $\mu_2$ and $\Psi^+_2$, we can write $G$ in the following way:

$$G = \Psi^+ \beta_2(u_2).$$

(4)

By $\alpha_2 \in \text{mp}(A)$, the following condition holds:

$$(\forall \omega \in A^+)(\omega \in \text{pref}(G)).\$$

(5)

Vice versa, considering sets $\mu_1$ and $\Psi^+_1$, we can write $G$ by $E \Psi^* v_1$. Then (4) is true if and only if

$$E \in \text{mp}^+(A).$$

(6)

By $E \subseteq F \in \text{mp}(A)$ and (6), we obtain the following: $E \in \text{mp}(A)$, then $E = F$.

**Proposition 3.** $(\forall v \in \beta_1)(\alpha_1(v) \in \text{mp}(A))$.

**Proof.** For the selected words $u_2$ and $v_1$ and set $G$ (see Proposition 2), let us consider the languages

$$H = G|_{v_1} \text{ and } K = H|_{v_1}.\$$

By the representing set $G$ in the form $E \Psi^* v_1$, which can be obtained considering $\mu_1$ and $\Psi^+_1$, we have $H = E \Psi^*$. By the last equation,

$$K = E(\Psi^*)|_{v_1},$$

then

$$E \subseteq \text{pref}(K).\$$

(7)

Let us consider the second representation of the language $G$, i.e. (4). By this representation, if

$$\alpha_i(v) \notin \text{mp}(A)$$

for some $i$, then (7) proved before cannot hold.

Now, let us consider the language

$$M = \Psi^1|_{u_2-v_1}.\$$

This language is not empty, because it contains the word $y$ selected before, besides

$$u_2 \in \text{pref}(y) \text{ and } v_1 \in \text{suffix}(y).$$
Let us consider some word \( x \in M \). By \( \Pr(\alpha_2) \), the following conditions hold:
\[
\Psi_1^{-1}(x) \neq \Psi_2^{-1}(x), \quad \alpha_2^{-1}(x) = u_2;
\]
similarly, \( \beta_1^{-1}(x) = v_1 \). Then the following condition holds:
\[
(\forall x \in \Psi^1)_{u_2-v_2})(\alpha_2^{-1}(x) = u_2, \ \beta_1^{-1}(x) = v_1).
\]

**Proposition 4.** \( E' = \Delta^m \) for \( m \) fixed before.

**Proof.** Supposing \( E' \neq \Delta^m \), we obtain \( \overline{\Sigma u}(E') \); therefore \( \overline{\Sigma u}(E) \).

Let us write the last fact in the following way:
\[
(\exists x_1, x_2 x_1 \in E)(|x_2| > 0).
\]

By (8), the following conditions hold:
\[
(u_2 x_1, v_1) \in \mu_1, \quad (u_2 x_2 x_1, v_1) \in \mu_2;
\]
then we have for any \( z \in L \):
\[
(u_2 x_1 z v_1) \in \Psi^1, \quad (u_2 x_2 x_1 z v_1) \in \Psi^1.
\]

Also, we obtain by (8), that
\[
\alpha_2^{-1}(u_2 x_1 z v_1) = \alpha_2^{-1}(u_2 x_2 x_1 z v_1) = u_2.
\]

Then
\[
x_1 z v_1 \in L \beta_2 \text{ and } x_2 x_1 z v_1 \in L \beta_2,
\]
i.e. \( \overline{\Sigma u}(L \beta_2) \) holds, and the last fact contradicts to Proposition 3.

**5. Conclusion**

Thus, we show that in some cases, the language of left brackets can be represented as a special morphism of a maximum prefix code. This fact can be used to prove the solvability of the equivalence problem in some special subclasses of the class of context-free languages (like (2) etc.), which, in turn, can be used in various automation systems for building compilers.

**References**

[1] Melnikov B 1993 The equality condition for infinite catenations of two sets of finite words International Journal of Foundation of Computer Sciences 4(3) 267–74

[2] Melnikov B 1995 Some equivalence problems for free monoids and for subclasses of the CF-grammars class Number theoretic and algebraic methods in computer science. World Sci. Publ. 125–37

[3] Melnikov B F and Vakhitova A A 1998 Some more on the finite automata The Korean Journal of Computational and Applied Mathematics (Journal of Applied Mathematics and Computing) 5(3) 495–505

[4] Melnikov B and Kashlakova E 2000 Some grammatical structures of programming languages as simple bracketed languages Informatica (Lithuanian Acad. of Sciences) 11(4) 441–6

[5] Dang V V, Korabelshchikova S Yu, and Melnikov B F 2015 On the problem of finding the minimum approximation semigroup News of higher educational institutions. Volga region. Physics and Mathematics 3(35) pp 88–99

[6] Melnikov B F, Korabelshchikova S Yu, and Dolgov V N 2019 On the task of extracting the root from the language International Journal of Open Information Technologies 7(3) 1–6

[7] Gross M and Lentin A 1967 Notions sur les gramm aires formelles Gauthier-Villars Ed