A UNIQUE STRUCTURE OF TWO-GENERATED BINARY EQUALITY SETS.

ŠTĚPÁN HOLUB

ABSTRACT. Let L be the equality set of two distinct injective morphisms g and h, and let L be generated by at least two words. Recently it was proved ([2]) that such an L is generated by two words and g and h can be chosen marked from both sides. We use this result to show that L is of the form $\{a^ib, ba^i\}^*$, with $i \geq 1$.

1. INTRODUCTION

Binary equality sets are the most simple non-trivial equality languages. Nevertheless, their precise description is still not known. They were for the first time extensively studied by K. Čulík II and J. Karhumäki in [3]. There the authors indicate that the only existing binary equality sets of rank two have the form $\{a^ib, ba^i\}^*$, but avoid to state it as a conjecture. Instead, they made a conjecture that in non-periodic cases (periodic cases being easy to deal with) the equality set is generated by at most two words. This statement was partially proved by A. Ehrenfeucht, J. Karhumäki and G. Rozenberg ([4]) leaving open the possibility of an infinitely generated equality set of the form $(\alpha\gamma^*\beta)^*$. The result is a corollary of the proof that the binary Post Correspondence Problem is decidable, previously achieved by the same authors ([1]). The mentioned possibility, contradicting the original conjecture, was excluded recently in [2], where we prove a stronger statement: the two words generating the equality set start (end resp.) with different letters. This in particular means that the equality set belongs to a pair of morphisms marked from both sides. In the present paper we therefore investigate such morphisms and show that their equality set can be generated by two words only if it is of the form $\{a^ib, ba^i\}^*$. This yields the complete characterization of binary equality sets generated by more than one word.

The paper at hand is actually an exercise in combinatorial analysis. After preliminaries (Section 2) we present some auxiliary lemmas based mostly on the primitivity of a word (Section 3). In the fourth section some general results concerning our morphisms are obtained. In Section 5 special cases are treated.

2. Assumptions and definitions

We first fix our notation.

By A we denote the binary alphabet $\{a, b\}$. The empty word is denoted by ε.

The set of all prefixes of u is denoted by $\mathrm{pref}(u)$. A prefix v of u is *proper* if $v \neq \varepsilon$ and $v \neq u$. Similarly *proper suffix* is defined. The set of all suffixes of u is denoted by $\mathrm{suff}(u)$. The first (the last resp.) letter of a non-empty word u is denoted by $\mathrm{pref}_1(u)$ ($\mathrm{suff}_1(u)$ resp.). A word v is called a *factor* of u if there exist words $w, w' \in A^*$ such that $u = w\,v\,w'$. A factor is said to be proper if and only if both w and w' are non-empty. If $v \in \mathrm{pref}(u)$ or $u \in \mathrm{pref}(v)$, we say that u and v are *comparable*. If $uv = w$ we also write $u = wv^{-1}$ and $v = u^{-1}w$. A word w is called an *overlap* of u and v if $w \in \mathrm{suff}(u) \bigcap \mathrm{pref}(v)$, or $w \in \mathrm{suff}(v) \bigcap \mathrm{pref}(u)$.

The *ratio* of a word $u \in A^*$ is defined by, $\mathrm{rat}(u) = \frac{|u|_a}{|u|_b}$. It is either a non-negative rational number or, in case $u \in a^+$, infinity.

Let $g, h : A^* \to A^*$ be binary morphisms. Their *equality set* is defined by

$$\mathrm{Eq}(g, h) = \{u \in A^* |\ g(u) = h(u)\}\,.$$

The choice of A as the target alphabet does not harm generality, since any alphabet can be encoded by two letters.

A binary morphism g is said to be *marked* if and only if $\mathrm{pref}_1(g(a)) \neq \mathrm{pref}_1(g(b))$. If, moreover, $\mathrm{suff}_1(g(a)) \neq \mathrm{suff}_1(g(b))$, we say that g is *marked from both sides*. Similarly we say that two non-empty words x and y are marked from both sides, if and only if $\mathrm{pref}_1(x) \neq \mathrm{pref}_1(y)$ and $\mathrm{suff}_1(x) \neq \mathrm{suff}_1(y)$.

We say that g is periodic, if words $g(a)$ and $g(b)$ commute, i.e., they have the same primitive root.

It is easy to verify that the set $\mathrm{Eq}(g, h)$ is a free submonoid of A^* generated by the set of its minimal elements

$$\mathrm{eq}(g, h) = \mathrm{Eq}(g, h) \setminus (\mathrm{Eq}(g, h) \setminus \{\varepsilon\})^2 \setminus \{\varepsilon\}\,.$$

If $g \neq h$, and u and v are non-empty elements of $\mathrm{Eq}(g, h)$ then $\mathrm{rat}(u) = \mathrm{rat}(v)$. This follows easily from the length agreement of g and h on elements of their equality set.

The following is known about the structure of $\mathrm{Eq}(g, h)$ (see [2]).

Theorem 2.1. *Let g and h be non-periodic binary morphisms. Then*

$$E(h, g) = \{\alpha, \beta\}^*$$

for some (possibly empty) words $\alpha, \beta \in A^*$. If α and β are both non-empty then they are marked from both sides. Moreover, there are binary morphisms g' and h' marked from both sides, such that $\mathrm{Eq}(g, h) = \mathrm{Eq}(g', h')$.

In this paper we investigate binary morphisms $g, h : A^* \to A^*$, whose equality set is generated by two non-empty words α and β. The symmetry of letters a and b, and of morphisms g and h, and Theorem 2.1 allow to adopt following assumptions without loss of generality

Conditions 2.2.

- $|g(a)| > |h(a)|$
- $|g(b)| < |h(b)|$
- $|h(b)| \geq |g(a)|$
- $\mathrm{pref}_1(g(a)) = \mathrm{pref}_1(h(a)) = a$
- $\mathrm{pref}_1(g(b)) = \mathrm{pref}_1(h(b)) = b$
- $\mathrm{suff}_1(g(a)) = \mathrm{suff}_1(h(a)) \neq \mathrm{suff}_1(g(b)) = \mathrm{suff}_1(h(b))$
- $\mathrm{pref}_1(\alpha) = a$
- $\mathrm{pref}_1(\beta) = b$
- $\mathrm{suff}_1(\alpha) \neq \mathrm{suff}_1(\beta)$

We are going to prove the following

Theorem 2.3. *Let $g, h : A^* \to A^*$ be binary morphisms, such that $\mathrm{eq} = \{\alpha, \beta\}$, satisfying Conditions 2.2. Then there is a positive integer i such that $\alpha = a^i b$ and $\beta = ba^i$.*

Since $g \neq h$, both α and β contain both letters a and b. Note that the difference between letters a and b is given only by the condition $|h(b)| \geq |g(a)|$. Therefore if $|h(b)| = |g(a)|$ then $i = 1$.

Throughout the paper k, k', l and l' will be positive integers such that

- $a^k b$ is a prefix of α,
- $b^l a$ is a prefix of β,
- $ba^{k'}$ and $ab^{l'}$ are elements of $\mathrm{suff}\{\alpha, \beta\}$.

3. Auxiliary lemmas

In this section we present several auxiliary lemmas. The proofs are easy and we omit them. We also omit well known characterization of conjugate words and the Periodicity Lemma.

The following Lemma is a consequence of the fact that two words generate a free semigroup if and only if they do not commute.

Lemma 3.1. *Let all words $g(a)$, $g(b)$, $h(a)$ and $h(b)$ be generated by words x and y, which do not commute. Define morphism $\pi : A^* \to A^*$ by $\pi(a) = x$ and $\pi(b) = y$. Then π is injective and $\mathrm{Eq}(g, h) = \mathrm{Eq}(\pi^{-1} \circ g, \pi^{-1} \circ h)$.*

It is the well known fact that a primitive word p is not a proper factor of pp. This implies following list of claims.

Lemma 3.2. *Let swp be a factor of w^+. Then s is a suffix, and p a prefix of w^+.*

Lemma 3.3. *Let x and y be words marked from both sides. Let u be a factor of $(xy)^+$. Then any overlap of u and $xyyx$ is strictly shorter than $|xy|$.*

Lemma 3.4. *Let x and y be words marked from both sides. Let u be a word with a prefix (suffix resp.) xyx. Let w be a word such that*

$$w \in \mathrm{pref}(u) \bigcap \mathrm{suff}(xyyx) \qquad (w \in \mathrm{suff}(u) \bigcap \mathrm{pref}(xyyx) \text{ resp.})$$

Then w is strictly shorter than $|xy|$.

Lemma 3.5. *Let x and y be words marked from both sides. Then xyx is not a factor of $xyyx$, and $xyyx$ is not a factor of $\{xy, xyx\}^+$.*

Lemma 3.6. *Let x and y be words marked from both sides. Let u and v be non-empty words such that*

- $yx \in \mathrm{pref}(u)$, $xy \in \mathrm{suff}(u)$,
- $v \in \{xy, xyx\}^+$,
- $|v| > |u|$.

Then v is not a factor of u^+.

4. General considerations

Lemma 4.1. *The words $g(a)$ and $h(a)$ ($g(b)$ and $h(b)$ resp.) do not commute.*

Proof. Suppose $g(a) = t^i$ and $h(a) = t^j$, with $i > j \geq 1$. Then the maximal element of t^+, which is a prefix of $g(\alpha\beta)$, is $t^{i \cdot k}$. On the other hand, $t^{j \cdot k}$ is the maximal element of t^+, which is a prefix of $h(\alpha\beta)$. This is a contradiction with $g(\alpha\beta) = h(\alpha\beta)$. Similarly for $g(b)$ and $h(b)$. □

Lemma 4.2. *The word $g(b)^l$ ($h(a)^k$ resp.) is a prefix of $h(b)$ ($g(a)$ resp.). Similarly, $g(b)^{l'}$ ($h(a)^{k'}$ resp.) is a suffix of $h(b)$ ($g(a)$ resp.)*

Proof. Suppose, on the contrary, that $h(b)$ is a prefix of $g(b)^l$. Since $g(b)$ is a suffix of $h(b)$, the words $g(b)$ and $h(b)$ commute, a contradiction with Lemma 4.1. The rest is analogical. □

We list some characteristic situations, which are implied by a word in $\mathrm{Eq}(g, h)$.

Conditions 4.3.

(A) *There is a proper suffix s of $g(b)$, such that $h(b)$ is a prefix of $sg(a)^+$.*
(B) *There is a proper prefix p of $g(b)$, such that $h(b)$ is a suffix of $g(a)^+p$.*
(C) *The word $h(b)$ is a factor of $g(a)^+$.*
(D) *There is a non-empty suffix u of $g(a)^+$ and a non-empty prefix v of $g(a)^+$, such that $ug(b)v = h(b)$.*

Lemma 4.4. *Let ub be a prefix of $\beta\alpha$.*

(i) *If $|g(u)| < |h(u)|$ or $|g(u)| > |h(ub)|$ then condition* (C) *or* (A) *holds.*

(ii) *If $|g(ub)| > |h(ub)|$ or $|g(ub)| < |h(u)|$ then condition* (C) *or* (B) *holds.*

Proof. We first introduce some terminology. Let $m = |\alpha\beta|_b$ and $w = g(\alpha\beta) = h(\alpha\beta)$. Each letter b is mapped by g to a factor $g(b)$ of w, and by h to a factor $h(b)$ of w. The factor of w, which is an image of the i-th occurrence of letter b, with $1 \leq i \leq m$, will be called i-th $g(b)$-factor of w. Similarly we define i-th $h(b)$-factor of w. We shall consider the position of $g(b)$-factors with respect to corresponding $h(b)$-factors.

(i) Let $|g(u)| < |h(u)|$ and let u be the longest prefix of $\beta\alpha$ satisfying the assumption. Put $i = |ub|_b$. By assumption, the i-th $g(b)$-factor of w does not start within the i-th $h(b)$-factor. If the $(i+1)$-th $g(b)$-factor starts there, then $|g(u')| < |h(u')|$ for the prefix u' of $\alpha\beta$ such that $|u'b|_b = i+1$. But we supposed that u is the longest possible. Therefore no $g(b)$-factor starts within the i-th $h(b)$-factor and the claim follows. Similarly for the shortest possible u, if $|g(u)| > |h(ub)|$.

(ii) The proof is analogical.

□

Corollary 4.5.

(i) *If $l > 1$ or $l' > 1$ then either conditions* (A) *and* (B) *hold, or condition* (C) *holds.*

(ii) *If none of conditions* (A), (B), (C) *and* (D) *holds, then* $\mathrm{eq}(g,h) = \{a^ib, ba^i\}$, $i \geq 1$.

Proof.

(i) Let $l > 1$ and put $u = b$. Then $|g(u)| < |h(u)|$ and, by Lemma 4.2, also $|g(ub)| < |h(u)|$. The statement now follows from Lemma 4.4. Similarly if $l' > 1$.

(ii) It is not difficult to deduce, by Lemma 4.4, that if none of the conditions holds, all letters b in $\alpha\beta$ must be starting or ending. Therefore there are only two letters b in $\{\alpha, \beta\}$. Since the case $\{ba^ib, a^j\}$ implies $g = h$, we are left with $\{a^ib, ba^j\}$. The length agreement yields $i = j$.

□

Lemma 4.6. *Let x and y be words such that xy is primitive, and*

$$g(a) \in (xy)^*x, \qquad h(a) \in (xy)^*x,$$
$$g(b) \in (yx)^*y, \qquad h(b) \in (yx)^*y.$$

Then $\mathrm{eq}(g,h) = \{ab, ba\}$.

Proof. Let $w = g(u) = h(u)$. By Lemma 3.1, we can suppose $x = a$ and $y = b$.

Let u be an element of $\mathrm{Eq}(g,h)$. Suppose that aa is a factor of u and $u = u_1aau_2$, where aa is not a factor of u_1a. The word $g(u_1a)a$ is the shortest prefix of $g(u)$ ending with aa. Similarly $h(u_1a)a$ is the shortest prefix of $h(u)$ of that form. Thus $g(u_1a) = h(u_1a)$.

This implies that aa is a factor of neither α nor β. In the same way we can show that neither α nor β contains bb as a factor. Thus either $\alpha \in (ab)^+a$ and $\beta \in (ba)^+b$, or $\alpha = ab$ and $\beta = ba$. The first possibility is excluded by the fact that α and β have the same ratio. □

The previous lemma has the following modification.

Lemma 4.7. *Let xy be a primitive word, with $x, y \in A^+$, such that*

$$g(a) \in (xy)^+x, \qquad g(b) \in (yx)^+y, \qquad h(b) \in (yx)^+y.$$

Then $\mathrm{eq}(g,h) = \{ab, ba\}$.

Proof. By Lemma 4.6 it is enough to show $h(a)$ is in $(xy)^*x$. The assumptions imply that x and y are marked from both sides.

1. Suppose ab is a prefix of α. Then $h(a)yx$ is a prefix of $(xy)^+$ and therefore $h(a) \in (xy)^*x$.
2. Suppose, on the other hand, that aa is a prefix of α. Then the word $yxxy$ is either a factor of $h(b)$, or $h(b)$ is a factor $yxxy$, or the two words have an overlap of length at least $|xy|$. This is a contradiction with Lemma 3.5 or Lemma 3.3.

□

5. Cases

The main *principium divisionis* is whether the word $g(ab)$ is longer or shorter than the word $h(b)$.

Case 5.1. $|g(ba)| \leq |h(b)|$.

The point of this case is to prove the following

Claim 5.1.

$$g(b^la) \in \mathrm{pref}(h(b)) \quad \textit{and} \quad g(ab^{l'}) \in \mathrm{suff}(h(b))\,.$$

Proof. It is enough to prove $|g(b^la)| \leq |h(b)|$ and $|g(ab^{l'})| \leq |h(b)|$.

Proceed by contradiction, and suppose, by symmetry, $|g(b^la)| > |h(b)|$. Since $|g(ba)| < |h(b)|$, $l \geq 2$. Therefore the word $g(b^la)$ is a prefix of $h(b)g(b)^{l-1}$ and there is a word u and a proper prefix q of $g(b)$, such that

$$h(b) = g(b)^lu, \qquad g(a) = ug(b)^iq,$$

with $0 \leq i \leq l-2$.

Suppose that $b^l ab$ is a prefix of β. Then $g(b)^i qg(b)$ is a factor of $g(b)^l$, and Lemma 3.2 yields that q is a suffix of $g(b)^+$, a contradiction. Therefore $b^l aa$ is a prefix of β.

By Corollary 4.5 we have to consider two possibilities.

1. Suppose $h(b)$ is a factor of $g(a)^+$. Let t be the primitive root of $g(a)$ and let $v_1 \in \mathrm{suff}(t)$ and $v_2 \in \mathrm{pref}(t)$ be words such that $h(b) \in (v_1 t^* v_2)$. Since $g(b)^i qt \in \mathrm{suff}(t^+)$ is comparable with $h(b)$, it is also comparable with $v_1 t$, and primitivity of t yileds that $g(b)^i q \in v_1 t^*$. Therefore $h(b\,b)$ is a prefix of $g(b)^l t^+$. Similarly we deduce that $h(b\,b)$ is a suffix of $t^+ g(b)^{l'}$. Hence, by primitivity of t, $h(b\,b\,b) = g(b)^l t^m g(b)^{l'}$, for some positive integer m. From

$$|t| + |g(b)| \leq |g(a)| + |g(b)| < |h(b)|,$$
$$3 \cdot |h(b)| = (l + l') \cdot |g(b)| + m \cdot |t|$$

it is not difficult to deduce that either

$$(l + l') \cdot |g(b)| > |g(b)| + |h(b)|,$$

or

$$m \cdot |t| > |t| + |h(b)|.$$

This implies, by Periodicity Lemma, that either $g(b)$ or t commutes with $h(b)$. We thus get a contradiction with Lemma 4.1 or with $\mathrm{pref}_1(h(b)) \neq \mathrm{pref}_1(g(a))$.

2. Suppose now that $h(b)$ is a prefix of $sg(a)^+$ and a suffix of $g(a)^+ p$, with a proper suffix s and a proper prefix p of $g(b)$. Lemma 3.2 and g is marked from both sides imply that

$$|h(b)| < |s| + |p| + |g(a)|.$$

Therefore there are words x and y such that xy is primitive, $g(a) \in (xy)^+ x$, yx is a prefix and xy a suffix of $g(b)$. Therefore $xyyx$ occurs on the edge of $h(b)h(b)$ and it is easy to derive a contradiction with Lemma 3.5 or Lemma 3.4.

□

It is now straightforward to see that

Claim 5.2. *None of conditions* (A), (B) *and* (C) *holds.*

Proof.

1. If $h(b)$ is a factor of $g(a)^+$, then, by Claim 5.1, $g(b)^l g(a)$ is a factor of $g(a)^+$. This is a contradiction with Lemma 3.2 and g being marked from both sides.

2. Let (A) hold, and $h(b)$ be a prefix of $sg(a)^+$ for some proper suffix s of $g(b)$. Lemma 3.2 implies that $s^{-1}g(b)^l$ is a suffix of $g(a)^+$, a contradiction. Similarly for condition (B).

□

Lemma 4.4 now implies that $l = l' = 1$, and $h(b)$ is a prefix of $g(b)g(a)^+$ and a suffix of $g(a)^+g(b)$. It is slightly more complicated to see that

Claim 5.3. *The condition* (D) *does not hold.*

Proof. In this proof p_i (s_i resp.) will always denote a proper prefix (a proper suffix resp.) of $g(a)$.

Suppose that (D) holds. We have

$$h(b) = g(b)g(a)^{m_1}p_1 = s_2g(a)^{m_2}g(b) = s_3g(a)^{m_3}g(b)g(a)^{m_4}p_4,$$

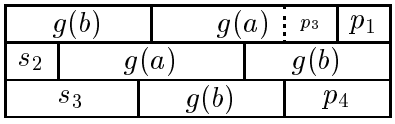
with $m_1, m_2, m_3, m_4 \geq 0$. Since $g(a)^{m_3}r$ is a factor of $g(a)^+$ for a nonempty prefix r of $g(b)$, Lemma 3.2 and g is marked imply that $m_3 = 0$. The mirrored consideration yields $m_4 = 0$.

Hence $|h(b)| < |g(b)| + 2 \cdot |g(a)|$, and therefore $m_1 = m_2 = 1$. We can write

(1) $h(b) = g(b)g(a)p_1$

(2) $h(b) = s_2g(a)g(b)$

(3) $h(b) = s_3g(b)p_4,$



where $|s_2| < |s_3|$ and $|p_1| < |p_4|$. From (1) and (3) we deduce $p_4 = p_3p_1$ and

$$g(b)g(a) = s_3g(b)p_3\,,$$

with $s_3p_3 = g(a)$. Hence

$$g(b)p_3s_3 = s_3g(b)p_3\,,$$

and words $g(b)p_3$ and s_3 have a common primitive root, say t. Let $t = t_1t_2$ be a factorization of t such that

$$g(b) = (t_1t_2)^{i_1}t_1, \qquad p_3 = t_2(t_1t_2)^{i_2}, \qquad s_3 = (t_1t_2)^j.$$

with $i_1, i_2 \geq 0$, $j \geq 1$. Then also

$$\begin{aligned} g(a) = p_3s_3 &= (t_2t_1)^{i_2+j}t_2, \\ g(b)g(a) &= (t_1t_2)^{i_1+i_2+j+1}, \\ g(a)g(b) &= (t_2t_1)^{i_1+i_2+j+1}. \end{aligned}$$

From (2) and (1) it follows that $s_2(t_2t_1)$ is a prefix of $g(b)g(a)$ and thus

$$s_2 = (t_1t_2)^{i_3}t_1, \qquad h(b) = s_2g(a)g(b) = (t_1t_2)^{i_1+i_2+i_3+j+1}t_1,$$

with $i_3 \geq 0$. The equality (3) gives

$$p_4 = (t_1 t_2)^{i_2+i_3+1}$$

and, since p_4 is a prefix of $g(a)$, the words t_1 and t_2 commute. Therefore also $g(a)$ and $g(b)$ commute, a contradiction. □

Corollary 4.5 together with the above claims now yields $\mathrm{Eq}(g,h) = \{a^i b, ba^i\}^*$.

Case 5.2. $|g(ab)| > |h(b)|$

We first consider a special situation:

Lemma 5.4. *If $k = k' = l = l' = 1$, then $\mathrm{eq}(g,h) = \{ab, ba\}$.*

Proof. If $|g(ab)| = |h(ab)|$, we are through. Suppose that $|g(ab)| > |h(ab)|$. The case $|g(ab)| < |h(ab)|$ is analogical. Assumptions now imply that

$$g(ab) = h(ab)v \tag{4}$$

for some non-empty word v. Since $h(b)$ is a suffix of $g(ab)$, there is a word u such that

$$uh(b) = h(b)v.$$

Let xy be a primitive word such that x is non-empty and

$$u = (yx)^i, \qquad v = (xy)^i, \qquad h(b) = (yx)^j y,$$

with $i \geq 1$ and $j \geq 0$. From $|h(ab)| > |g(a)|$ and from (4) we deduce $|g(b)| > |v|$. Since $g(b)$ is both prefix and suffix of $h(b)$, primitivity of xy yields $g(b) = (yx)^{j_1} y$, $j_1 \geq 1$. We have

$$g(a) = h(a)(yx)^{i+j-j_1} = (xy)^{i+j-j_1} h(a).$$

Therefore, by characterization of conjugates, $h(a), g(a) \in (xy)^* x$ and we are through by Lemma 4.6. □

Subcase 5.2.1. $(l + l')|g(b)| \geq |h(b)|$

If $(l + l')|g(b)| = |h(b)|$ then $g(b)$ and $h(b)$ commute, a contradiction with Lemma 4.1.

If $(l + l' - 1)|g(b)| \geq |h(b)|$, then $g(b)$ and $h(b)$ again commute, by Periodicity Lemma.

Therefore $|h(b)| + |g(b)| > (l + l')|g(b)| > |h(b)|$. This implies that there exists a primitive word xy, with $x, y \in A^+$, such that

$$g(b) = (yx)^i y, \qquad h(b) = ((yx)^i y)^{l-1} (yx)^m y ((yx)^i y)^{l'-1},$$

with $i \geq 1$, and $i < m \leq 2i$. The factor $(yx)^m y$ in the expression of $h(b)$ represents the overlapping occurrences of $g(b)$.

Then also

(5) $(xy)^{m-i}((yx)^i y)^{l'-1} \in \mathrm{pref}(g(a)), \quad ((yx)^i y)^{l-1}(yx)^{m-i} \in \mathrm{suff}(g(a)).$

Note that x and y are marked from both sides.

1. Suppose first that either $l > 1$ or $l' > 1$, and apply Corollary 4.5. By Lemma 3.6, the word $h(b)$ is not a factor of $g(a)^+$. Therefore there is a proper suffix s of $g(b)$, such that $h(b)$ is a prefix of $sg(a)^+$, and

(6) $$s^{-1}g(b)^l x \quad \text{is a prefix of} \quad g(a).$$

The prefix yx of $h(b)$ is also a prefix of $sxy \in \mathrm{pref}(sg(a))$, which is a suffix of $(yx)^{i+1}y$. This implies that $s = (yx)^{i_1}y$, with $i_1 \geq 0$. Therefore

(7) $$(yx)^{i_1+m-i}y((yx)^i y)^{l'-1} \in \mathrm{pref}(h(b)).$$

We shall show that $l' \leq l$ and $i_1 = 2i - m$.

1.1. Suppose that $l = 1$ and $l' > 1$. By (5), the word $(yx)^i y(xy)^{m-i}y$ is the shortest prefix of $h(b)$ ending with xyy. From (7) we get another expression of this word, namely $(yx)^{i_1+m-i}yy$. This implies $m = i_1 + m - i$. Therefore $i_1 = i$, a contradiction with s being proper prefix of $g(b)$.

1.2. If $l > 1$, the shortest prefix of $h(b)$ ending with xyy is $(yx)^i yy$, and, as above, we deduce $i = i_1 + m - i$, in accordance with the claim. Thus both $((yx)^i y)^{l'}$ and $((yx)^i y)^l x$ are prefixes of $h(b)$, and l' is at most l.

Mirror considerations yield $l \leq l'$ and thus $l = l'$. From (6) we now conclude that

$$(xy)^{m-i}y((yx)^i y)^{l-1}x \quad \text{is a prefix of} \quad g(a).$$

It follows that the word

$$g(b)^l(xy)^{m-i}((yx)^i y)^{l-1}x$$

is a prefix of $h(b)^l$, and x is a prefix of $h(b)$, a contradiction.

2. Suppose then that either $k > 1$ or $k' > 1$. By symmetry, let $k > 1$. We shall use the fact that $g(aa)$ contains a factor $yxxy$. By Lemma 3.6, the word $h(b)$ is not a factor of $g(a)^+$. Therefore $g(a)^k$ is a factor of $h(b)$. Since $|h(a^k)| < |g(a)|$, we get a contradiction with Lemma 3.4.

We have shown that any possibility, except $k = k' = l = l'$, is contradictory, and can use Lemma 5.4.

Subcase 5.2.2. $(l + l')|g(b)| < |h(b)|$

We have

$$g(a)g(b)^{l'}u = vg(b)^l g(a) = vh(b)u = vg(b)^l wg(b)^l u,$$

with $u, v, w \in A^+$. The word w is both a prefix and a suffix of $g(a)$, and

$$g(a) = vg(b)^l w = wg(b)^{l'} u.$$

Thus there is a primitive word xy, such that x and y are marked from both sides, and

$$w = (xy)^j x, \qquad g(b)^{l'} u = (yx)^i, \qquad vg(b)^l = (xy)^i,$$

with $i \geq 1$, $j \geq 0$. We have

$$g(a) = (xy)^{i+j} x$$

and $h(b)$ is a factor of $(yx)^+$. We first prove the following

Claim 5.5. *If $k > 1$, or $k' > 1$, or* (C) *holds, then $h(b)$ is a factor of $yxxy$.*

Proof. This is a direct consequence of Lemma 3.3. □

1. Let first $|g(b)| \geq |y|$. Then yxy is a factor of $h(b)$. Claim 5.5 and Lemma 3.5 imply that (C) does not hold, and $k = k' = 1$.

   Suppose $l > 1$. Then $u = g(b)^{l-1} q$, with a prefix q of $g(b)$, and $g(b)^l q$ is a factor of $(xy)^+$. By Corollary 4.5, the condition (A) holds. Consequently, the word $g(b)^l xy$ is a prefix of $s(xy)^+$ for some suffix s of $g(b)$. Lemma 3.2 implies that $s^{-1} g(b)^l$ commutes with xy, and xy is a suffix of $g(b)^l$. Again by Lemma 3.2, we conclude that q is a prefix of $(xy)^+$, a contradiction. Similarly if $l' > 1$.

2. Let now $|g(b)| < |y|$.

2.1. Suppose $l > 1$. Then the word u is a prefix of $h(b)$ and consequently yx is a prefix of $g(b)^{l'+l}$. Since $g(b)$ is a suffix of y, Lemma 3.2 yields that x is a prefix of $g(b)^+$, a contradiction. Similarly if $l' > 1$.

2.2. Suppose now $l = l' = 1$ and $k > 1$. Claim 5.5 implies $|h(b)| \leq |yxxy|$ and $|h(b)| \geq |g(a)|$ yields $i + j = 1$. Thus $i = 1$ and $j = 0$, and from

$$2|g(b)| + |x| = |h(b)| \geq |g(a)| = 2|x| + |y|$$

we deduce

$$|x| + |y| \leq 2|g(b)|.$$

The word $g(b)$ is a prefix and a suffix of y. Therefore there exist a primitive word $x_1 y_1$, with $y_1 \in A^+$, $x_1 \in A^*$, and integers $1 \leq i_1 \leq j_1$ such that

$$y = (y_1 x_1)^{i_1 + j_1} y_1, \qquad g(b) = (y_1 x_1)^{j_1} y_1, \tag{8}$$

and

$$|(y_1 x_1)^{j_1 - i_1} y_1| \geq |y_1| \geq |x|.$$

Therefore

$$h(b) = (y_1 x_1)^{j_1} y_1 \, x \, (y_1 x_1)^{j_1} y_1$$

and Claim 5.5 now yields

(9) $(y_1x_1)^{j_1}y_1\,x\,(y_1x_1)^{j_1}y_1$ is a factor of $(y_1x_1)^{i_1+j_1}y_1\,xx\,(y_1x_1)^{i_1+j_1}y_1$.

Let u_1 and v_1 be words such that

$$u_1(y_1x_1)^{j_1}y_1\,x\,(y_1x_1)^{j_1}y_1v_1 = (y_1x_1)^{i_1+j_1}y_1\,xx\,(y_1x_1)^{i_1+j_1}y_1.$$

Note that x and y_1 are marked from both sides.
This implies that y_1xy_1 is not a factor of y_1xxy_1, by Lemma 3.5, and either

(10) $u_1(y_1x_1)^{j_1}y_1$ is a proper prefix of $((y_1x_1)^{i_1+j_1}y_1)$,

or

(11) $(y_1x_1)^{j_1}y_1v_1$ is a proper suffix of $((y_1x_1)^{i_1+j_1}y_1)$.

By symmetry, suppose (10). Consider the factor $x_1y_1xy_1x_1y_1$ in claim (9). Primitivity of its prefix x_1y_1 yields that $xy_1x_1y_1$ is comparable with $(x_1y_1)^mxx$, $m \geq 1$. If $m = 1$ then y_1 and x are comparable, a contradiction. On the other hand, $m > 1$ implies that xy_1x_1 is a prefix of $x_1y_1x_1y_1$, and primitivity of y_1x_1 yields $x = x_1$. From (8) we have $yx = (y_1x_1)^{i_1+j_1+1}$, a contradiction with primitivity of xy.

We are left with $l = l' = k = k' = 1$, and Lemma 5.4 concludes the proof.

References

1. A. Ehrenfeucht, J. Karhumäki, and G. Rozenberg, *The (generalized) Post correspondence problem with lists consisting of two words is decidable*, Theoret. Comput. Sci. **21** (1982), no. 2, 119–144. MR **84k:**68035
2. Š. Holub, *Binary equality sets are generated by two words*, to appear.
3. K. Culik II and J. Karhumäki, *On the equality sets for homomorphisms on free monoids with two generators*, RAIRO Theor. Informatics **14** (1980), 349–369.
4. A. Ehrenfeucht, J. Karhumäki and G. Rozenberg, *On binary equality sets and a solution to the test set conjecture in the binary case*, J.Algebra **85** (1983), 76–85.

Turku Center for Computer Science, Turku, Finland and Charles University, Prague, Czech Republic

E-mail address: holub@karlin.mff.cuni.cz