RAIRO. INFORMATIQUE THÉORIQUE

K. CULIK II
J. KARHUMÄKI

On the equality sets for homomorphisms on free monoids with two generators

RAIRO. Informatique théorique, tome 14, n° 4 (1980), p. 349-369
<http://www.numdam.org/item?id=ITA_1980__14_4_349_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 14, n° 4, 1980, p. 349 à 369)

ON THE EQUALITY SETS FOR HOMOMORPHISMS ON FREE MONOIDS WITH TWO GENERATORS (*) (3)

by K. CULIK II (1) and J. KARHUMÄKI (2)

Communicated by J. BERSTEL

ABSTRACT. – *A problem which can be considered dual to the Post's Correspondence Problem is shown to be decidable.*

The equality sets for homomorphisms on free monoids with two generators are studied. In particular, for some words all possible homomorphisms agreeing on them are shown. If all such distinct homomorphisms are periodic, then the word (set) is called periodicity forcing. A number of periodicity forcing words or sets is shown. A number of equality sets of the form F^* with cardinality of F at most two is shown. In particular all such equality sets where $F \subset a^+ b^+ \cup b^+ a^+$ are given.

RÉSUMÉ. – *On prouve la décidabilité d'un problème qui peut être considéré comme le dual du problème de correspondance de Post.*

On étudie les ensembles d'égalité pour les homomorphismes sur un monoïde libre à deux générateurs. On donne, pour des mots particuliers, tous les homomorphismes qui coïncident sur eux. Si tous ces homomorphismes sont périodiques, le mot, ou l'ensemble de mots, est appelé « periodicity forcing ». On donne un certain nombre d'ensembles de cette nature. On donne des ensembles d'égalités de la forme F^*, où F a au moins deux éléments. En particulier, on donne tous les ensembles d'égalité de ce type où $F \subset a^+ b^+ \cup b^+ a^+$.

1. INTRODUCTION

Decision problems of whether two homomorphisms on a free monoid agree on at least one or all words from a given set are of crucial importance in computability and formal language theory. The former is the classical Post's Correspondence Problem while the latter played a crucial role in proving the decidability of the DOL equivalence problem [3]. Recently there has been much research done in this direction. For example in [4] it was shown that given a context-free language L and two homomorphisms it is decidable whether they

(*) Received June 1979, revised November 1979.

(1) Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada.
(2) Department of Mathematics, University of Turku, 20500 Turku 50, Finland. The paper was written during the author's visit at the University of Waterloo.

(3) This research was supported by the National Sciences and Engineering Council of Canada, Grant No. A 7403.

agree on every word of L. Already several years ago it was conjectured by A. Ehrenfeucht that for every language L there exists a finite "test set" F so that any pair of homomorphisms agree on L iff they agree on F. Very recently this problem was answered positively for L over a binary alphabet [5] but remains open in the general case.

In a somewhat different direction the notion of an equality set for two homomorphisms, i.e. the set of all words on which the homomorphisms agree, has been introduced in [12]. Equality sets have turned out to be a powerful tool in the characterization of various language classes. The reader is referred to [1, 2, 6] for further details. Equality sets are also useful in some decidability proofs, in particular the result that for elementary homomorphisms (*see* section 2) the equality set is always regular [7].

This paper approaches the topic of equality sets in still another direction. We study equality sets for the specific case of a binary alphabet. For this case we attempt to find all the sets which can be expressed as equality sets for some homomorphisms. This goal was not fully accomplished but we hope that our results can be extended in such a way that they would lead to solutions of some important problems of the kind discussed above. In particular we have in mind the decidability of the emptiness problem for equality sets in some special cases, for example for homomorphisms over a binary alphabet, which in other words is the decidability of the Post's Correspondence Problem for lists of length two.

After some preliminaries we show that a problem which can be considered dual to the Post's Correspondence Problem is decidable. This is shown by reducing it to the recent deep result of Makanin [11], namely the decidability of the existence of a solution for a system of equations over a free monoid.

Then we discuss some basic, mostly already known, properties of equality sets for homomorphisms over a binary alphabet. The restriction to binary alphabets ensures that every homomorphism is either elementary or periodic and consequently every equality set is either regular (it is not known whether this is effective) or of a very special form, namely the set of all words with a fixed ratio of the two symbols.

Next we solve the following problem for some cases: Given a word, find all possible pairs of homomorphisms which agree on this word.

Sections 6 and 7 give some partial solutions to the problem of characterizing all the equality sets for homomorphisms over a binary alphabet. Such a characterization is difficult but very interesting since it would probably imply the decidability of the emptiness problem for equality sets (Post's Correspondence Problem) in some special cases, and have other applications as discussed in section 8.

We call a set of words over a binary alphabet periodicity forcing if in every pair of distinct homomorphisms agreeing on every word from the set both the homomorphisms must be periodic. We first investigate the singleton periodicity forcing sets, that is periodicity forcing words. Finally, in section 7 we exhibit many two-element periodicity forcing sets and also some two-element sets which are not periodicity forcing. We give some results to support our conjecture that every set containing at least three r-primitive words (words for which no prefix has the same ratio of symbol occurrences as the whole word) is periodicity forcing. Equivalently this would mean that every equality set for two elementary homomorphisms over a binary alphabet is the star of a two element set.

2. PRELIMINARIES

We give here the basic definitions and some known results, which are needed later.

The free monoid generated by a finite alphabet Σ is denoted by Σ^*. For u, $v \in \Sigma^*$, we write $u \quad v$ if u is a prefix of v (not necessarily proper). The length of w in Σ^* is denoted $|w|$, specifically $|\varepsilon| = 0$ for the empty word, $\Sigma^+ = \Sigma^* - \{\varepsilon\}$. For a set A, $|A|$ denotes the cardinality of A. For w in Σ^* and a in Σ, the number of occurrences of a in w is denoted by $\#_a(w)$. For w in $\{a, b\}^+$, $r(w) = \#_a(w)/\#_b(w)$ is the *ratio* of w. A word w in Σ^* is *ratio primitive* (r-primitive) if $r(u) \neq r(w)$ for every nonempty proper prefix u of w.

Consider two homomorphisms g and h mapping Σ^* into Δ^* (possibly $\Sigma = \Delta$). The equality set of g and h [12] is defined by

$$E(g, h) = \{ w \in \Sigma^* : g(w) = h(w) \}.$$

The minimal equality set of g and h [1] is defined by

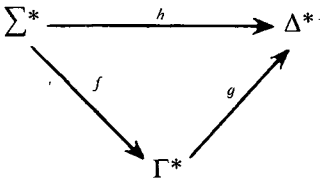
$$e(g, h) = \{ w \in \Sigma^+ : g(w) = h(w) \text{ and if } w = uv \text{ where } u, v \in \Sigma^+, \text{ then } g(u) \neq h(u) \}.$$

For a binary alphabet $\Sigma = \{a, b\}$, and distinct g, h, $e(g, h) = E(g, h) \cap P$ where $P = \{ w \in \Sigma^+ : w \text{ is } r\text{-primitive} \}$. Indeed, if $x \in E(g, h)$ and $x = yz$ where $r(y) = r(x)$, then also $y \in E(g, h)$. For the converse *see* lemma 4.1.

In this paper we study equality sets. We note that without loss of generality we can restrict ourselves to a binary target alphabet Δ since any alphabet Γ can be encoded over a binary alphabet Δ, and $E(g, h) = E(g', h')$ where g', h' are the compositions of g and h, respectively, with the encoding. On the other hand we

clearly cannot similarly encode the source alphabet Σ. Therefore results obtained for a binary alphabet need not be valid in the general case.

A homomorphism $h : \Sigma^* \to \Delta^*$ is *elementary* [7] if there does not exist a decomposition of h into f and g, $h = gf$,



so that $|\Gamma| < |\Sigma|$.

A homomorphism $h : \Sigma^* \to \Delta^*$ is *periodic* if there is w in Δ^* such that, for each a in Σ, there is an integer q such that $h(a) = w^q$. It is clear that for $\Sigma = \{a, b\}$ every homomorphism on Σ^* is either elementary or periodic. A set $L \subseteq \Sigma^*$ is *periodicity forcing* if for any distinct homomorphisms g, h on Σ^*, the property $h(w) = g(w)$ for each w in L implies the periodicity of both g and h. If $L = \{x\}$, then we say simply that x is periodicity forcing.

For each w in Σ^* the *primitive root* of w is denoted by $\rho(w)$ and defined as the shortest word u in Σ^* such that $w = u^n$ for some $n \geq 1$. In particular, $\rho(u) = \varepsilon$ iff $u = \varepsilon$. It is well known that $\rho(w)$ is unique. The following lemmas turn out to be useful for this paper, *see* e.g. [10].

LEMMA 2.1: *For u, v in Σ^+, $uv = vu$ iff $\rho(u) = \rho(v)$.*

LEMMA 2.2: *For u, v, w in Σ^*, if $w \prec u^m$, $w \prec v^n$ for some $m, n \geq 1$, and $|w| \geq |u| + |v|$, then $\rho(u) = \rho(v)$.*

LEMMA 2.3: *For u, v in Σ^* and w in Σ^+, if $uv = vw$, then there exist p, s in Σ^* and $k \geq 0$ such that $u = sp$, $v = s(ps)^k$ and $w = ps$.*

LEMMA 2.4: *For u, v, w in Σ^+ and m, n, $p \geq 2$, if $u^m v^n = w^p$, then $\rho(u) = \rho(v) = \rho(w)$.*

3. DUAL POST'S CORRESPONDENCE PROBLEM

Here we show that a problem which can be considered dual to the Post's Correspondence Problem (PCP in short) is decidable.

The problem is: Given a string w in Σ^*, do there exist two distinct homomorphisms $h, g : \Sigma^* \to \Delta^*$ for some Δ, such that at least one of them is aperiodic and $h(w) = g(w)$?

Note that if the requirement that h and g are distinct or, for $w \notin a^+ \cup b^+$, that at least one of them is aperiodic is omitted then such homomorphisms always exist.

We show the decidability of the dual PCP by reducing it to Makanin's result concerning solvability of equations in free monoids.

THEOREM 3.1: *The dual PCP is decidable.*

Proof: Given w in Σ^*, we construct a finite number of systems of equations over a free monoid at least one of which has a solution iff there exist homomorphisms h and g satisfying the requirements of the given instance of the dual PCP.

In view of the discussion in section 2 we may assume that $\Delta = \{0, 1\}$. For $\Sigma = \{a_1, \ldots, a_n\}$ let $\overline{\Sigma} = \{\bar{a} : a \in \Sigma\}$ and $\overline{\overline{\Sigma}} = \{\overline{\bar{a}} : a \in \Sigma\}$. The word obtained from ξ in Σ^* by (double) barring of all symbols will be denoted $\overline{\xi}(\overline{\overline{\ }})$. In all systems of equations, the set of unknowns will be $\overline{\Sigma} \cup \overline{\overline{\Sigma}} \cup \{s, t, u, v, y, z\}$ and the constants 0 and 1. For each fixed i, j, k, $1 \leqq i, j, k \leqq n$, $k \neq i$ and α, β, γ, $\delta \in \{0, 1\}$, $\alpha \neq \beta$ and $\gamma \neq \delta$ we construct the following three systems of equations over Δ^*:

$$\bar{w} = \overline{\overline{w}}; \tag{1}$$

$$\bar{a}_k \bar{a}_i = u \alpha v, \qquad \bar{a}_i \bar{a}_k = u \beta z; \tag{2}$$

either

$$\left.\begin{array}{l} \bar{a}_j \gamma y = \overline{\overline{a}}_j; \\ \text{or} \\ \overline{\overline{a}}_j \gamma y = \bar{a}_j; \\ \text{or} \\ \bar{a}_j = y \gamma s, \quad \overline{\overline{a}}_j = y \delta t. \end{array}\right\} \tag{3}$$

By [11] we can test whether at least one of these systems has a solution. If so, then the given instance of the dual PCP has a solution. To see this consider homomorphisms g, h defined by $g(a) = \bar{a}$, $h(a) = \overline{\bar{a}}$ for each a in Σ, and verify that they satisfy the requirements of the dual PCP: (1) is equivalent to $g(w) = h(w)$. Equations (2) are equivalent to $g(a_k)\, g(a_i) \neq g(a_i)\, g(a_k)$ which, in turn, holds true iff $\rho(g(a_k)) \neq \rho(g(a_i))$, $g(a_k) \neq \varepsilon$ and $g(a_i) \neq \varepsilon$ (lemma 2.1). Hence, g is nonperiodic iff (2) is valid for some i and k (and α, β with $\alpha \neq \beta$). Finally, (3) holds for some j iff h and g are distinct.

If none of the systems of equations has a solution, then, clearly, also the dual PCP has no solution. □

Note, that it is easy to modify the above proof for the case when both g and h are required to be aperiodic.

COROLLARY 3.2: *Given $w \in \{a, b\}^*$, it is decidable whether there exist distinct elementary homomorphisms $g, h : \{a, b\}^* \to \Delta^*$ for some Δ such that $g(w) = h(w)$.*

Proof: Over a binary alphabet a homomorphism is elementary iff it is aperiodic. Hence, the result follows by the note above. □

Now we generalize theorem 3.1. We show that it is decidable whether there exist two homomorphisms as in theorem 3.1 which agree on every word from a given regular set.

THEOREM 3.3: *Given a regular set R, it is decidable whether there exist two distinct homomorphisms g and h such that g is aperiodic and $g(x) = h(x)$ for each x in R.*

Proof: By [4], for every regular set R there effectively exists a finite set F (called the test set) so that $g(x) = h(x)$ for each x in R iff $g(x) = h(x)$ for each x in F. Hence, we can restrict ourselves to the case when R is finite. For a finite R the proof is obtained by an easy modification of the proof of theorem 3.1, namely by replacing equation (1) by equations $\bar{x} = \bar{\bar{x}}$ for each x in R. □

In theorem 3.3, as in therorem 3.1, both g and h can be required to be aperiodic.

Furthermore, the theorem could clearly be extended to every language family for languages of which there effectively exists a finite test set. The results in [4] strongly suggest that the context-free languages are such a family.

Finally we note that it is also easy to see that it is decidable whether there exist two distinct periodic homomorphisms agreeing on every word of a given regular set.

4. PROPERTIES OF EQUALITY SETS OVER BINARY ALPHABETS

From now on we will be investigating the properties of equality sets and later on of their elements, that is of solutions of instances of PCP. In doing this we will restrict ourselves to homomorphisms over a binary alphabet, that is in terms of PCP to instances of PCP with lists of length two. Henceforth we assume $\Sigma = \{a, b\}$.

Example 4.1: Let the homomorphisms $h, g : \Sigma^* \to \Sigma^*$ be defined by

$$g : \begin{array}{l} a \to aab, \\ b \to a, \end{array} \qquad h : \begin{array}{l} a \to a, \\ b \to baa. \end{array}$$

Here every element of $E(g, h)$ must start with a and the "continuation" is uniquely determined. So we have a single minimal "solution" $aabb$

$$\begin{array}{c} g(a)\ \ g(a)\ \ g(b)\,g(b) \\ \ulcorner\!\!\urcorner\ \ulcorner\!\!\urcorner\!\!\urcorner\ \ulcorner\!\!\urcorner \\ a\ a\ b\ \ a\ a\ b\ \ a\ a \\ \llcorner\!\!\lrcorner\!\!\llcorner\!\!\lrcorner\ \ \llcorner\!\!\lrcorner\!\!\llcorner\!\!\lrcorner \\ h(a)\,h(a)\ \ h(b)\ \ \ h(b) \end{array},$$

therefore $E(g, h) = \{ aabb \}^*$.

Example 4.2: Consider homomorphisms g, h defined by

$$\begin{array}{llll} g: & a \to aab, & h: & a \to a, \\ & b \to aa, & & b \to baa. \end{array}$$

Here it is easy to see that $e(g, h) = \emptyset$, i.e. $E(g, h) = \{ \varepsilon \}$, since clearly each potential solution would have to start with a and then to continue deterministically as indicated

$$\begin{array}{ll} g: & \ulcorner\!\!\urcorner\ \ulcorner\!\!\urcorner\!\!\urcorner\ \ulcorner\!\!\urcorner\!\!\urcorner\ \ulcorner\!\!\urcorner\!\!\urcorner\ \ulcorner\!\!\urcorner\!\!\urcorner\ \ulcorner\!\!\urcorner \\ & a\ a\ b\ \ a\ a\ b\ \ a\ a\ \ a\ a\ \ a\ a\ b\ \ a\ a\ b\ , \\ h: & \llcorner\!\!\lrcorner\!\!\llcorner\!\!\lrcorner\ \ \llcorner\!\!\lrcorner\!\!\llcorner\!\!\lrcorner\!\!\llcorner\!\!\lrcorner \end{array}$$

that is we are forced to generate the infinite word of the form

$$a\ a\ b\ b\ a^4\ b^4\ a^8\ b^8\ \ldots\ a^{2^n}\ b^{2^n}\ \ldots$$

The following fact concerning ratios is mentioned in [5].

LEMMA 4.1: *Let g, h be distinct homomorphisms over a binary alphabet. If u, $v \in E(g, h)$, then $r(u) = r(v)$.*

The equality sets for periodic homomorphisms over a binary alphabet are characterized by the following lemma which is easy to verify.

LEMMA 4.2: *Let g and h be distinct periodic homomorphisms over a binary alphabet with minimal periods p and q, respectively. Then $E(g, h)$ is of the form*

$$\{ \varepsilon \} \cup \{ w \in \Sigma^+ \mid r(w) = k \}, \tag{1}$$

where $k \geqq 0$ is a rational number or $k = \infty$ if $p = q$, and $E(g, h) = \{ \varepsilon \}$ if $p \neq q$. Every set of form (1) *is an equality set for some periodic homomorphisms.*

LEMMA 4.3: *A homomorphism over a binary alphabet is elementary iff it is injective.*

Proof: It is shown in [7], Thm. 3.7, that each elementary homomorphism is injective. Clearly, each injective homomorphism over an at least two-letter alphabet is aperiodic and finally, by definition, an aperiodic homomorphism over a binary alphabet is elementary. □

LEMMA 4.4: *Let g and h be homomorphisms over a binary alphabet. The equality set $E(g, h)$ is either regular or of the form* (1) *for some rational $k>0$.*

Proof: If $g=h$, then $E(g, h)=\Sigma^*$, a regular set. If g and h are distinct and at least one of them is elementary then the regularity of $E(g, h)$ was shown in [8]. Otherwise both g and h are periodic. By lemma 4.2, if $E(g, h)\neq\{\varepsilon\}$, then we have

$$E(g, h)=\{\varepsilon\}\cup\{w\in\Sigma^+ \mid r(w)=k\},$$

where $k\geqq 0$ is a rational number or $k=\infty$. This last set is regular in both cases $k=0$ or $k=\infty$, which completes the proof. □

From the above proof and lemma 4.3 we also have the following:

LEMMA 4.5: *Let g, h be distinct homomorphisms over a binary alphabet. If at least one of them is injective, then $E(g, h)$ is regular.*

The above result does not hold for an arbitrary alphabet even if both g and h are required to be injective, for a counterample *see* [9].

The following result is shown in [9].

LEMMA 4.6: *Let g and h be homomorphisms over a binary alphabet, g elementary and h periodic. Then there exists effectively a word w so that $E(g, h)=\{w\}^*$.*

5. HOMOMORPHISMS AGREEING ON A GIVEN WORD

In this section we consider the problem of finding all pairs of homomorphisms agreeing on a given word over $\{a, b\}$. Obviously, this problem is more difficult than the dual PCP and although we do not know any "practical" algorithm even for the dual PCP we will solve this more difficult problem in some special cases. Certainly, such solutions throw light on the theory of equality sets as a whole.

We start with:

LEMMA 5.1: *The word $ab\in E(h, g)$, with $|h(a)|>|g(a)|$, iff there exist words α, β and γ such that $\beta\neq\varepsilon$ and*

$$\left.\begin{array}{llll} h: & a\to\alpha\beta, & g: & a\to\alpha, \\ & b\to\gamma, & & b\to\beta\gamma. \end{array}\right\} \tag{1}$$

Proof: Indeed, $h(ab)=g(ab)$ with $|h(a)|>|g(a)|$ iff there exists a word $z\neq\varepsilon$ such that $h(a)=g(a)z$ and $zh(b)=g(b)$. Hence, the lemma follows when we choose $\alpha=g(a)$, $\beta=z$ and $\gamma=h(b)$. □

Formula (1) does not tell us very much about the equality sets, since it includes three variables. If we require that ba is also in $E(h, g)$ we can say much more:

THEOREM 5.2: *The set* $\{ab, ba\}\subseteq E(h, g)$, *with* $|h(a)|>|g(a)|$, *iff there exist nonnegative integers* t_1, t_2 *and* t_3 *and words* α *and* β *such that* $t_2>0$, $\alpha\beta\neq\varepsilon$ *and*

$$\left.\begin{array}{llll} h: & a\to\alpha(\beta\alpha)^{t_1+t_2}, & g: & a\to\alpha(\beta\alpha)^{t_1}, \\ & b\to\beta(\alpha\beta)^{t_3}, & & b\to\beta(\alpha\beta)^{t_2+t_3}, \end{array}\right\} \quad (2)$$

or

$$\left.\begin{array}{llll} h: & a\to\alpha, & g: & a\to\varepsilon, \\ & b\to\varepsilon, & & b\to\alpha. \end{array}\right\}$$

Moreover, if $\{ab, ba\}\subseteq E(h, g)$, $h\neq g$, *and* h *or* g *is elementary, then* $E(h, g)=\{ab, ba\}^*$.

Proof: Obviously for any pair (h, g) of the form (2) $ab, ba\in E(h, g)$. To prove the converse let $\{ab, ba\}\subseteq E(h, g)$. Then, by lemma 5.1

$$\begin{array}{llll} h: & a\to\alpha'\beta', & g: & a\to\alpha', \\ & b\to\gamma', & & b\to\beta'\gamma', \end{array}$$

for some words α', β' and γ' with $\beta'\neq\varepsilon$. Hence, $ba\in E(h, g)$ implies $\gamma'\alpha'\beta'=\beta'\gamma'\alpha'$. So, by lemma 2.1, either $\gamma'\alpha'=\varepsilon$ or else $\rho(\beta')=\rho(\gamma'\alpha')\neq\varepsilon$. In the first case we have $h(a)=g(b)=\beta'$ and $h(b)=g(a)=\varepsilon$. In the second case there exist integers t_1, $t_3\geqq 0$ and $t_2>0$ and words α and β, with $\alpha\beta\neq\varepsilon$, such that $\beta'=(\beta\alpha)^{t_2}$, $\gamma'=(\beta\alpha)^{t_3}\beta$ and $\alpha'=\alpha(\beta\alpha)^{t_1}$. Thus the first sentence of the theorem follows.

To prove the second sentence of the theorem, let h and g be of the form (2) and let h (resp. g) be elementary. Define homomorphisms h_i, g_i and c by

$$\begin{array}{llll} h_i: & a\to a(ba)^{t_1+t_2}, & g_i: & a\to a(ba)^{t_1}, \\ & b\to b(ab)^{t_3}, & & b\to b(ab)^{t_2+t_3}, \end{array}$$

and

$$\begin{array}{ll} c: & a\to\alpha, \\ & b\to\beta. \end{array}$$

Then $h=ch_i$ and $g=cg_i$. Moreover, c is elementary since h (resp. g) is elementary. So c is injective implying $E(h, g)=E(h_i, g_i)$. Hence, the theorem follows since clearly $E(h_i, g_i)=\{ab, ba\}^*$. □

Corresponding to lemma 5.1 we also prove:

LEMMA 5.3: *The word* $aba \in E(h, g)$, *with* $|h(a)| > |g(a)|$, *iff there exist an integer* $t \geqq 0$ *and words* α, β *and* γ *such that* $\alpha\beta \neq \varepsilon$ *and*

$$\left.\begin{array}{lll} h: & a \to \alpha(\beta\alpha)^t \beta\alpha, & g: \quad a \to \alpha(\beta\alpha)^t, \\ & b \to \gamma, & \qquad b \to \beta\alpha\gamma\alpha\beta. \end{array}\right\} \tag{3}$$

Proof: Clearly, for any pair (h, g) satisfying (3) $aba \in E(h, g)$. To prove the converse let $aba \in E(h, g)$ with $|h(a)| > |g(a)|$. Then $h(a) = ug(a) = g(a)v$ for some nonempty words u and v. Hence, by lemma 2.3, there exist an integer $t \geqq 0$ and words α, β and γ such that $g(a) = \alpha(\beta\alpha)^t$, $u = \alpha\beta$ and $v = \beta\alpha$. Since $g(b) = vh(b)u$ the lemma follows if we choose $\gamma = h(b)$. □

LEMMA 5.4: *The word* $aab \in E(h, g)$, *with* $|h(a)| > |g(a)|$, *iff there exist an integer* $t \geqq 0$ *and words* α, β *and* γ *such that* $\alpha\beta \neq \varepsilon$ *and*

$$\left.\begin{array}{lll} h: & a \to (\alpha\beta)^t \alpha\alpha\beta, & g: \quad a \to (\alpha\beta)^t \alpha, \\ & b \to \gamma, & \qquad b \to \beta\alpha\alpha\beta\gamma. \end{array}\right\} \tag{4}$$

Proof: As above it is enough to show that if $aab \in E(h, g)$, with $|h(a)| > |g(a)|$, then h and g are of the form (4). Denote $h(a) = g(a)u$. Then $ug(a)uh(b) = g(a)g(b)$. Let v be the word satisfying $ug(a) = g(a)v$ which implies that $vuh(b) = g(b)$. By lemma 2.3, there exist an integer $t \geqq 0$ and words α and β, with $\alpha\beta \neq \varepsilon$, such that $u = \alpha\beta$, $g(a) = (\alpha\beta)^t\alpha$ and $v = \beta\alpha$. Denote $h(b) = \gamma$. Then $g(b) = vuh(b) = \beta\alpha\alpha\beta\gamma$ and $h(a) = g(a)u = (\alpha\beta)^t\alpha\alpha\beta$. Hence the lemma follows. □

As in lemma 5.1, formulae (3) and (4) contain three variables (and one parameter) and hence the equality set can not be immediately determined. Actually, as we shall see later, both $\{aab\}^*$ and $\{aab, baa\}^*$ are equality sets determined by (3). On the other hand, we shall show (theorem 7.2) that the only regular equality set obtained from (4) is $\{aba\}^*$.

Our next result gives another example of the case when the conditions $x \in E(h, g)$, $h \neq g$ and h is elementary imply that $E(h, g) = x^*$. Moreover, now all the homomorphisms agreeing on a given word are obtained using only two variables.

THEOREM 5.5: *The word* $aabb \in E(h, g)$, *with* $|h(a)| > |g(a)|$, *iff there exist nonnegative integers* t_1, t_2 *and* t_3 *and words* α *and* β *such that* $\alpha\beta \neq \varepsilon$ *and*

$$\left.\begin{array}{lll} h: & a \to \alpha(\beta\alpha)^{t_1}\alpha(\beta\alpha)^{t_2}\beta, & g: \quad a \to (\alpha\beta)^{t_1}\alpha, \\ & b \to \alpha(\beta\alpha)^{t_3}, & \qquad b \to \beta(\alpha\beta)^{t_2}\alpha(\alpha\beta)^{t_3}\alpha. \end{array}\right\} \tag{5}$$

Hence, $aabb \in E(h, g)$, $h \neq g$, and h or g elementary, imply $E(h, g) = \{aabb\}^*$.

Proof: It is easy to see that the second sentence is a consequence of the first one (*cf.* the proof of theorem 5.2, It is also clear that any pair (h, g) satisfying (5) also satisfies $aabb \in E(h, g)$.

So it remains to be shown that if $aabb \in E(h, g)$ and $|h(a)| > |g(a)|$, then h and g are of the form (5). Denote $g(a) = \bar{a}$, $g(b) = \bar{b}$, $h(a) = \bar{\bar{a}}$ and $h(b) = \bar{\bar{b}}$. Then the equality $h(a^2 b^2) = g(a^2 b^2)$ becomes

$$\bar{\bar{a}}\,\bar{\bar{a}}\,\bar{\bar{b}}\,\bar{\bar{b}} = \bar{a}\,\bar{a}\,\bar{b}\,\bar{b}. \qquad (\star)$$

Since $|\bar{\bar{a}}| > |\bar{a}|$, then $|\bar{\bar{b}}| < |\bar{b}|$ and so there exist words x and y such that $\bar{\bar{a}} = \bar{a}x$ and $\bar{b} = \bar{\bar{b}}y$. Thus, from $(\star)$ it follows that

$$x\bar{a}x\bar{b} = \bar{a}y\bar{\bar{b}}y,$$

which implies that $|x| = |y|$ and therefore

$$x\bar{a} = \bar{a}y,$$
$$x\bar{\bar{b}} = \bar{\bar{b}}y.$$

Now we apply lemma 2.3 to both of these equalities and conclude the existence of words u, v, z and r and integers k, $k' \geqq 0$ such that

$$x = uv = zr, \qquad y = vu = rz,$$
$$\bar{a} = (uv)^k u, \qquad \bar{\bar{b}} = (zr)^{k'} z.$$

We first assume that $|u| \geqq |z|$. Hence equalities $uv = zr$ and $vu = rz$ lead, by theorem 5.2, to the following three subcases: Either $|u| = |z|$ or $u = r = \alpha$ and $v = z = \varepsilon$ for some nonempty word α or

$$u = \alpha(\beta\alpha)^{t_1 + t_2}, \qquad z = \alpha(\beta\alpha)^{t_1},$$
$$v = \beta(\alpha\beta)^{t_3}, \qquad r = \beta(\alpha\beta)^{t_2 + t_3},$$

for some words α and β, with $\alpha\beta \neq \varepsilon$, and for integers t_1, $t_3 \geqq 0$ and $t_2 > 0$.

In the first case $u = z$ and $v = r$ and hence

$$h: \; a \to (uv)^k uuv, \qquad g: \; a \to (uv)^k u,$$
$$b \to (uv)^{k'} u, \qquad\qquad b \to vuu(vu)^{k'}.$$

In the second case

$$h: \; a \to \alpha^{k+2}, \qquad g: \; a \to \alpha^{k+1},$$
$$b \to \alpha^{k'}, \qquad\qquad b \to \alpha^{k'+1}.$$

Finally, in the third case after setting $t = t_1 + t_2 + t_3$:

$$h: \begin{array}{l} a \to \alpha(\beta\alpha)^{(t+1)k+t_1+t_2}\alpha(\beta\alpha)^t\beta, \\ b \to \alpha(\beta\alpha)^{(t+1)k'+t_1}, \end{array} \qquad g: \begin{array}{l} a \to (\alpha\beta)^{(t+1)k+t_1+t_2}\alpha, \\ b \to \beta(\alpha\beta)^t\alpha(\alpha\beta)^{(t+1)k'+t_1}\alpha. \end{array}$$

So in all the cases the homomorphisms are of the form (5).

In the main case $|u| \leqq |z|$ the homomorphisms h and g are obtained in all three subcases from above by interchanging u with z and v with r. It is straightforward to see that these homomorphisms are still of the form (5). Hence the proof is complete. □

As an application of theorem 5.5 one can show the following. The word $a^4 b^4$ belongs to $E(h, g)$, with $|h(a)| > |g(a)|$ and h or g elementary, iff there exist nonempty words α and β such that $\rho(\alpha) \neq \rho(\beta)$ and

$$\left.\begin{array}{ll} h: & a \to \alpha^4\beta, \\ & b \to \alpha, \end{array} \qquad \begin{array}{ll} g: & a \to \alpha, \\ & b \to \beta\alpha^4. \end{array}\right\} \tag{6}$$

So the pair (h, g) is now, in a sense, unique. The proof of this fact is straightforward but long, therefore we omit it here.

It is interesting to note that the uniqueness of the pair (h, g) above is not a consequence of the fact that the exponents in $a^4 b^4$ are "large". This is demonstrated in the following example with words $a^4 b^3$ and $a^4 b^5$.

Example 5.1: Define, for all $n, m \geqq 1$, homomorphisms h_1, g_1, h_2 and g_2 by setting

$$h_1: \begin{array}{l} a \to (a^{mn}b)^n, \\ b \to a^m, \end{array} \qquad g_1: \begin{array}{l} a \to a^n, \\ b \to (ba^{mn})^m. \end{array}$$

and

$$h_2: \begin{array}{l} a \to (a^n ba^n)^{mn}a^n b, \\ b \to a, \end{array} \qquad g_2: \begin{array}{l} a \to a^n ba^n, \\ b \to ((ba^{2n})^{mn}ba^n)^m. \end{array}$$

Then

$$E(h_1, g_1) = (a^m b^n)^*,$$

and

$$E(h_2, g_2) = (a^{mn+1}b^n)^*.$$

6. PERIODICITY FORCING WORDS

In this section we are looking for periodicity forcing words over $\{a, b\}$, i.e. words w having the property: $w \in E(h, g)$ implies h and g are periodic. By theorem 3.1, it is decidable whether a given word is periodicity forcing.

However, our proof of theorem 3.1 does not give any example of a periodicity forcing word. Here we will show that such words really exist.

It is a simple task to show that some words which are not r-primitive are periodicity forcing. For example the word *abaabb* is such since $h(abaabb)=g(abaabb)$ implies $h(ab)=g(ab)$ and $h(aabb)=g(aabb)$ which is (by theorem 5.5 or a simple direct argument) possible only if both h and g are periodic. It is a little more complicated to show that there also exist r-primitive periodicity forcing words. Before proving this we show that there is no periodicity forcing word shorter than five.

LEMMA 6.1: *For any word w with $|w| \leqq 4$ there exist elementary homomorphisms h and g such that $h(w)=g(w)$.*

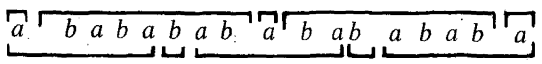
Proof: By example 5.1, $\{a^i b^j\}^*$ is an equality set for all $i, j \geqq 1$. Further $\{a^i b a^j\}^*$, for all $i, j \geqq 1$, is the equality set of the elementary homomorphisms defined by $h(a)=a$, $h(b)=a^i b a^j$, $g(a)=a^2$ and $g(b)=b$. The word *abba* is not periodicity forcing by theorem 5.2. Hence the lemma follows. □

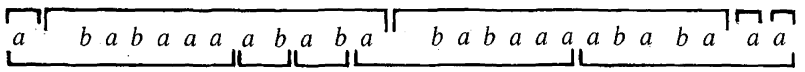
In the next result we characterize periodicity forcing words of lenght five.

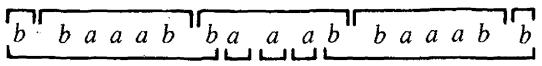
THEOREM 6.2: *A word w in $\{a, b\}^*$ of length 5 is periodicity forcing iff $w=abaab$ or $w=babaa$ or w is obtained from these by interchanging letters or by taking mirror images or by making both the operations.*

Proof: By symmetry, and by the fact that all words containing at most one b are in equality sets of some elementary homomorphisms we may assume that w contains two occurrences of b. Further, by symmetry, it is enough to show that the words *abaab* and *babaa* are periodicity forcing while the words *aaabb*, *ababa*, *abbaa* and *baaab* are not.

The fact that *aaabb* is not periodicity forcing follows from example 5.1. The same holds true for *ababa*, *abbaa* and *baaab* since:







So it remains to be shown that the words *abaab* and *babaa* are periodicity forcing. We consider each of them separately.

I. *abaab*: Assume that $abaab \in E(h, g)$ with $|h(a)| > |g(a)|$. Define $g(x) = \bar{x}$ and $h(x) = \bar{\bar{x}}$ for all $x \in \{a, b\}^*$. Let $c = ab$. Then we have $\bar{c}\,\bar{a}\,\bar{c} = \bar{\bar{c}}\,\bar{\bar{a}}\,\bar{\bar{c}}$. Moreover

$$|\bar{c}| - |\bar{\bar{c}}| = |g(ab)| - |h(ab)| > 0.$$

Hence, by lemma 5.3,

$$\begin{cases} \bar{c} = \bar{a}\,\bar{b} = \alpha(\beta\alpha)^t\,\beta\alpha, \\ \bar{a} = \gamma, \end{cases} \qquad \begin{cases} \bar{\bar{c}} = \bar{\bar{a}}\,\bar{\bar{b}} = \alpha(\beta\alpha)^t, \\ \bar{\bar{a}} = \beta\alpha\gamma\beta\alpha, \end{cases}$$

for some words α, β and γ and some integer $t \geqq 0$. Since $\beta\alpha\gamma\alpha\beta \prec \alpha(\beta\alpha)^t$ it follows that $\rho(\alpha) = \rho(\beta)$ and hence also $\rho(\gamma) = \rho(\alpha)$. So $\rho(\bar{a}) = \rho(\bar{b}) = \rho(\bar{\bar{a}}) = \rho(\bar{\bar{b}})$ proving that *abaab* is periodicity forcing.

II. *babaa*: Assuming that $babaa \in E(h, g)$ with $|h(a)| > |g(a)|$ and using the above notations we now have $\bar{c}\,\bar{c}\,\bar{a} = \bar{\bar{c}}\,\bar{\bar{c}}\,\bar{a}$ where $c = ba$. So, by lemma 5.4, we get

$$\begin{cases} \bar{c} = \bar{b}\,\bar{a} = (\alpha\beta)^t\alpha\alpha\beta, \\ \bar{a} = \gamma, \end{cases} \qquad \begin{cases} \bar{\bar{c}} = \bar{\bar{b}}\,\bar{\bar{a}} = (\alpha\beta)^t\alpha, \\ \bar{\bar{a}} = \beta\alpha\alpha\beta\gamma, \end{cases}$$

for some words α, β and γ and some integer $t \geqq 0$. Since, $\bar{\bar{b}}\,\bar{\bar{a}}$ is a suffix of a word in $\{\beta\alpha\}^*$ and $\beta\alpha\alpha\beta\gamma$ is a suffix of $\bar{\bar{b}}\,\bar{\bar{a}}\,\beta\alpha = \alpha\beta$, and hence $\rho(\alpha) = \rho(\beta)$. From this point onwards the proof continues as in case I. □

At this point we want to summarize what kinds of not periodicity forcing words are known to us. Let us call such a word as a solution referring to PCP. By example 5.1, any nonempty word in $a^* b^*$ is a solution. In $a^+ b^+ a^+$ all solutions known to us are as follows: any word of the form $a^i b a^j$ (*cf.* the proof of lemma 6.1), the words *abbaa* and *aabba* (*cf.* the proof of theorem 6.2) and the words $ba^{2i+1}b$, for $i \geqq 1$ (*cf.* the proof of theorem 6.2). The only other solutions known to us are those pointed out by the referee of this paper, namely the solutions of the form $(ab)^i a$ for $i \geqq 2$ (*cf.* again the proof of theorem 6.2).

As regards to periodicity forcing words we want to mention the following. In theorem 6.2 there are examples of r-primitive periodicity forcing words. Besides these we know that any r-primitive word in $(a^3 a^* b^3 b^*)^2$ is periodicity forcing. However the proof of this is tedious because of the many cases needed to be considered, and hence we omit it.

7. PERIODICITY FORCING SETS

In this section we consider periodicity forcing sets. We start with:

THEOREM 7.1: *Each of the following sets is periodicity forcing*

$$\{ ab, ax \}, \quad \textit{where } x \neq b \textit{ and } ax \textit{ is } r\textit{-primitive},$$
$$\{ aab, ax \}, \quad \textit{where } x \neq ab \textit{ and } ax \textit{ is } r\textit{-primitive}.$$

Proof: By lemma 4.1 each subset containing two words u and v such that $r(u) \neq r(v)$ is periodicity forcing. We shall therefore suppose that in the first case $r(ab) = r(ax)$ and in the second case $r(aab) = r(ax)$. In particular $x \neq \varepsilon$ in both cases. We consider these separately.

I. $\{ ab, ax \}$: Let $\{ ab, ax \} \subseteq E(h, g)$ with $|h(a)| > |g(a)|$. Then, by lemma 5.1, there exist words α and β such that

$$\left.\begin{array}{lll} h: & a \to \alpha\beta, & g: \quad a \to \alpha, \\ & b \to \gamma, & \quad\quad b \to \beta\gamma. \end{array}\right\} \tag{1}$$

If $\alpha = \varepsilon$ then g is periodic and hence, by lemma 4.6, also h must be periodic. Consequently $\alpha \neq \varepsilon$. Let $ax = a^i bz$ for some word z and $i \geqq 0$. Since ax is r-primitive $i > 1$. For $i \geqq 2$ we write

$$\alpha\beta(\alpha\beta)^{i-1}\gamma u = \alpha\alpha^{i-1}\beta\gamma v,$$

for some words u and v. Now we have two cases.

First, if $|\alpha| + |\beta| \leqq |\alpha^{i-1}|$, i. e. $|\beta| \leqq |\alpha^{i-2}|$, then by lemma 2.2 we have $\rho(\alpha\beta) = \rho(\alpha) \neq \varepsilon$ and therefore $\rho(\alpha) = \rho(\beta)$.

Secondly, if $|\alpha^{i-2}| < |\beta|$ we write $\beta = \alpha^{i-2} z$, $z \neq \varepsilon$, and we conclude that $\alpha\beta \prec z\alpha\beta$. Hence, $\alpha\beta = \alpha^{i-1} z = z\alpha^{i-1}$ showing that $\rho(z) = \rho(\alpha)$. So also in this case $\rho(\alpha) = \rho(\beta)$ and therefore $E(h, g) = \{ ab \}^*$, a contradiction. Hence the proof of case I is complete.

II. $\{ aab, ax \}$: Let again $\{ aab\ ax \} \subseteq E(h, g)$ with $|h(a)| > |g(a)|$. Then, by lemma 5.4, there exist words α, β and γ and an integer $t \geqq 0$ such that

$$\begin{array}{lll} h: & a \to (\alpha\beta)^t \alpha\alpha\beta, & g: \quad a \to (\alpha\beta)^t \alpha, \\ & b \to \gamma, & \quad\quad b \to \beta\alpha\alpha\beta\gamma. \end{array}$$

Since ax is r-primitive it is either of the form $ax = abu$ for some word u or of the form $ax = aaa^i bv$ for some word v and some integer $i \geqq 1$. If $ax = abu$, then for some w and z:

$$(\alpha\beta)^t \alpha\alpha\beta\gamma w = (\alpha\beta)^t \alpha\beta\alpha\alpha\beta z.$$

Hence, $\rho(\alpha)=\rho(\beta)$ and $E(h, g)=\{aab\}^*$, a contradiction. If $ax=aaa^i bz$, then for some w' and z' :

$$(\alpha\beta)^t\alpha\alpha\beta(\alpha 3)^t\alpha\alpha\beta(\alpha\beta)^t\alpha\alpha\beta w'=(\alpha\beta)^t\alpha(\alpha\beta)^t\alpha(\alpha\beta)^t\alpha\beta z'.$$

Hence, also in this case $\rho(\alpha)=\rho(\beta)$, which completes the proof of case II. □

We also prove:

THEOREM 7.2: *Each of the following sets is periodicity forcing*

$$\{aabb, x\}, \qquad \text{where } \rho(x)\neq aabb,$$
$$\{aba\ , x\}, \qquad \text{where } \rho(x)\neq aba\ .$$

Proof: The first assertion is an immediate consequence of theorem 5.5. To prove the second claim let us assume that $\{aba, x\}\subseteq E(h, g)$, with $|h(a)|>|g(a)|$. Now lemma 5.3 gives general expressions for h and g and using these and the arguments of the proof of theorem 7.1 it is straightforward to see that $\{aba, aay\}$, for any y, is periodicity forcing.

So it remains two cases: either $x=abbu$ for some u or $x=bv$ for some v. Let us assume first that $x=abbu$. For each word z we define so-called balance $B(z)$ of z by setting $B(z)=|h(z)|-|g(z)|$. Let $B(a)=n$. Then clearly $B(b)=-2n$ and $B(x)=0$. Obviously for any word z $B(z)$ is the multiple of n. So the r-primitiveness of x and the fact $B(abb)<0$ guarantees that $B(w)<0$ for all proper prefixes of x different from a. Hence x has a suffix aa, i. e. $x=u'aa$ for some u'. But, by the beginning of this proof and by symmetry $\{aba, u'aa\}$ is periodicity forcing.

Finally, $\{aba, bv\}$ is periodicity forcing for all v because of exactly the same reasons as $\{aba, abbu\}$. □

Our next aim is to characterize periodicity forcing sets included in $a^*b^*\cup b^*a^*$. We start with:

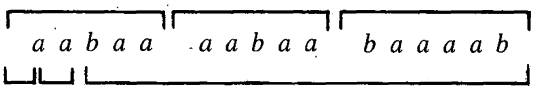
Example 7.1: Define, for $i\geq 1$, homomorphisms h and g by

$$h:\ a\to a^i ba^i, \qquad g:\ a\to a,$$
$$b\to (ba^{2i})^{i-1}b, \qquad b\to (ba^{2i})^{i-1}ba^i(ba^{2i})^{i-1}b.$$

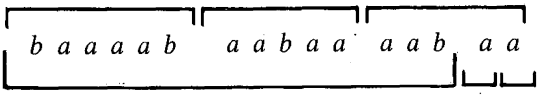
It is straightforward to see that

$$E(g, h)=\{a^i b, ba^i\}^*.$$

For instance, if $i=2$ we have



and



The above example provides the only examples (besides $\{a, b\}^*$) known to the authors when a regular equality set over a binary alphabet has two or more generators. These equality sets are also maximal in the sense that any set $\{a^i b, ba^i, x\}^*$ where $x \notin \{a^i b, ba^i\}^+$ is not an equality set, or even included in a regular equality set different from $\{a, b\}^*$, as we will next show.

THEOREM 7.3: *Let* $i \geqq 1$. *If the set* $\{a^i b, ba^i\} \subseteq E(h, g)$, $h \neq g$, *and* h *is elementary then* $E(h, g) = \{a^i b, ba^i\}^*$. *Consequently,* $\{a^i b, ba^i, x\}$ *is periodicity forcing for all* $i \geqq 1$ *and* $x \notin \{a^i b, ba^i\}^+$.

Proof: The case $i=1$ is covered by theorem 5.2. So consider $i \geqq 2$. By lemma 4.6, g must be elementary, hence, we may assume that $|h(a)| > |g(a)|$. Again let $h(x) = \bar{\bar{x}}$ and $g(x) = \bar{x}$ for $x \in \{a, b\}^*$. By theorem 5.2, or its obvious interpretation, there exist words p and s and integers n, $m \geqq 0$ and $t \geqq 1$ such that

$$\bar{\bar{a}}^i = s(ps)^{n+t}, \qquad \bar{a}^i = s(ps)^n,$$
$$\bar{\bar{b}} = (ps)^m p, \qquad \bar{b} = (ps)^{m+t} p.$$

If $n+t \geqq 2$, then by lemma 3.2, $\rho(\bar{\bar{a}}) = \rho(ps)$. Hence also $\rho(p) = \rho(s)$ contradicting the nonperiodicity of h.

So we conclude that $n=0$ and $t=1$. If $|\bar{a}^{i-1}| \geqq 1/2 |\bar{\bar{a}}|$, then clearly $\rho(\bar{a}) = \rho(\bar{\bar{a}})$ and hence $\rho(s) = \rho(p)$, a contradiction. Thus there exist words u, v and w such that

$$\bar{\bar{a}} = \bar{a}^{i-1} u \bar{a}^{i-1},$$

and

$$u = \bar{a}v = w\bar{a}.$$

From this last equality we conclude, by lemma 3.3, that

$$\bar{a} = (\alpha\beta)^k \alpha,$$
$$v = \beta\alpha,$$
$$w = \alpha\beta,$$

for some words α and β and some integer $k \geqq 0$. It is easily seen that

$$\left.\begin{aligned} s &= ((\alpha\beta)^k \alpha)^i, \\ p &= (\beta\alpha((\alpha\beta)^k \alpha)^{2i-1})^{i-2} \beta\alpha((\alpha\beta)^k \alpha)^{2i-2} \alpha\beta. \end{aligned}\right\} \tag{2}$$

So the homomorphisms h and g are of the form

$$\left.\begin{aligned} h: \quad & a \to g, & \qquad g: \quad & a \to r, \\ & b \to (ps)^m p, & & b \to (ps)^{m+1} p, \end{aligned}\right\} \tag{3}$$

where p and s are arbitrary words of the form (2), $g^i = sps$, $r^i = s$ and $m \geqq 0$.

From now on it is straightforward to see (*cf.* the proof of theorem 5.2) that $E(h, g) = \{ a^i b, ba^i \}^*$. □

It is interesting to note that the proof of the above theorem gives a characterization of all pairs of homomorphisms having the equality set $\{ a^i b, ba^i \}^*$, $i \geqq 2$. All such pairs are obtained from (2) and (3) by fixing the constants k and m and choosing words α and β such that $\rho(\alpha) \neq \rho(\beta)$. In particular, example 7.1 is obtained with $k = 0$, $m = 0$, $\alpha = a$ and $\beta = b$, that is

$$[h(a)]^i = sps = a^i (ba^{2i})^{i-1} ba^i = (a^i ba^i)^i \qquad \text{and} \qquad [g(a)]^i = a^i,$$

thus

$$h(a) = a^i ba^i, \qquad g(a) = a,\ h(b) = p = (ba^{2i})^{i-1} b$$

and

$$g(b) = psp = (ba^{2i})^{i-1} ba^i (ba^{2i})^{i-1} b.$$

Hence the homomorphisms in example 7.1 are in a sense minimal. More precisely, if we define the size of the pair (h, g) of homomorphisms to be $\max \{ |h(a)|, |g(a)| \mid a \in \Sigma \}$, then the smallest pair needed to define the language $\{ a^5 b, ba^5 \}^*$, for example, as an equality language is of size 95.

In contrast to example 7.1 we now show:

LEMMA 7.4: *Let i, j, m and n be natural numbers satisfying $in \neq 1$ and $jm \neq 1$. Then the set $\{ a^i b^j, b^m a^n \}$ is periodicity forcing.*

Proof: Assume that $\{ a^i b^j, b^m a^n \} \subseteq E(h, g)$ with $|h(a)| > |g(a)|$. Let $i \geqq n$, then by lemma 4.1 $j \geqq m$. We first consider the case when $i, j \geqq 3$. Again let $h(x) = \bar{\bar{x}}$ and $g(x) = \bar{x}$ for all $x \in \{ a, b \}^*$. We have two subcases.

(i) $|\bar{a}^i| \geqq |\bar{\bar{a}}|$ (or symmetrically $|\bar{\bar{b}}^j| \geqq |\bar{b}|$). Then there exist words p and s and integer $t \geqq 1$ such that $\bar{\bar{a}} = (ps)^t p$ and $\bar{a} = ps$. But $\bar{a}$ is also a suffix of $\bar{\bar{a}}$ showing that $\rho(p) = \rho(s)$ and hence $\rho(\bar{a}) = \rho(\bar{\bar{a}})$.

This gives us the equality

$$\overline{b}^j = \rho(\overline{a})^k \overline{\overline{b}}^j,$$

for some $k \geqq 2$. Also since $j \geqq 2$, $\rho(\overline{b}) = \rho(\overline{a}) = \rho(\overline{\overline{b}})$ by lemma 3.4, so both g and h are periodic.

(ii) $|\overline{a}^i| \leqq |\overline{\overline{a}}|$ and $|\overline{\overline{b}}^j| \leqq |\overline{b}|$. Let $\overline{\overline{a}} = \overline{a}x^i$ and $\overline{b} = y\overline{\overline{b}}^j$.

Then we have

$$\overline{b}^{j-1} y = x\overline{\overline{a}}^{i-1}.$$

Since $j-1, i-1 \geqq 2$ we conclude, by lemma 3.2, that there exist words q and r and natural numbers k_1 and k_2 such that $\rho(qr) = qr$ and

$$\overline{b} = (qr)^{k_1}, \qquad \overline{\overline{a}} = (rq)^{k_2}.$$

Hence $\overline{a}^i qrqr$ is a prefix of a word in $(rq)^*$ which implies that $\overline{a}^i \in (rq)^* r$ [recall that $\rho(rq) = rq$]. Symmetrically, $\overline{a}^n \in q(rq)^*$. These together guarantee that $\rho(r) = \rho(q)$. Thus, also $\rho(\overline{b}) = \rho(\overline{\overline{a}}) = \rho(\overline{a}) = \rho(\overline{\overline{b}})$.

So, there remain the cases where either i or $j \leqq 2$. Since $a^i b^j$ and $b^m a^n$ has a fixed ratio and $i \geqq n, j \geqq m$, it is sufficient to consider the sets $X_1 = \{a^t b^2, b^2 a^t\}$, $t \geqq 2$, and $X_2 = \{a^{2t} b^2, ba^t\}$, $t \geqq 1$. To see that X_1 is periodicity forcing let $X_1 \subseteq E(h, g)$, with $|h(a)| > |g(a)|$. Using the earlier notation of this proof let $\overline{\overline{a}}^t = \overline{a}^t z$ and $\overline{b}^2 = z\overline{\overline{b}}^2$. Then $z\overline{\overline{b}}^2 \overline{a}^t = \overline{\overline{b}}^2 \overline{a}^t z$ implying that $\rho(z) = \rho(\overline{\overline{b}}^2 \overline{a}^t)$. So $\overline{\overline{b}}^2 \overline{\overline{a}}^t \in \rho(z)^2 \rho(z)^*$, which means that

$$\overline{\overline{b}}^2 \overline{\overline{a}} t = \rho(z)^l,$$

for some $l \geqq 2$. Hence, by lemma 2.4, $\rho(\overline{b}) = \rho(\overline{\overline{a}}) = \rho(z)$. So both h and g are periodic.

To see that also $X_2 = \{a^{2t}, b^2, ba^{2t}\}$ is periodicity forcing, assume that $X_2 \subseteq E(h, g)$ with h elementary. We define a homomorphism $h' : a \to [h(a)]^t$, $b \to h(b)$. Clearly h' is also elementary. Moreover, since $X_2 \subseteq E(h, g)$, $\{a^2 b^2, b^2 a^2\} \subseteq E(h', g)$ contradicting theorem 7.2. So the proof of lemma 7.4 is complete. □

LEMMA 7.5: *Any two element set in $a^+ b^+$ is periodicity forcing.*

Proof: The proof can be carried out, for example, by employing the ideas of the proof of lemma 7.4. Moreover, theorem 7.1 is also useful. We omit the details. □

Now we are ready for:

THEOREM 7.6: *The subset X of $a^+ b^+ \cup b^+ a^+$ is periodicity forcing iff $|X| \geqq 2$ and X is not of the form $\{a^i b, ba^i\}$, $i \geqq 1$ (or symmetrically of the form $\{ab^i, b^i a\}$, $i \geqq 1$).*

Proof: By examples 6.1 and 7.1 if $|X| = 1$ or X is of the form $\{a^i b, ba^i\}$, for some $i \geqq 1$, X is not periodicity forcing. The converse follows from lemmas 7.4, and 7.5. □

We want to conclude this section by mentioning the following generalization of theorem 7.1. One can show that any two element set in $a^+ b^* a^*$ is periodicity forcing. In other words, for any pair of homomorphisms (h, g), $h \neq g$, h elementary,

$$|e(h, g) \cap a^+ b^* a^*| \leqq 1, \tag{4}$$

and hence

$$|e(h, g) \cap a^+ b^* a^* \cup b^+ a^* b^*| \leqq 2.$$

The proof we know for the above fact is tedious, and since the result itself is not so important, although it supports our conjecture (*cf.* section 8) we omit it.

8. OPEN PROBLEMS

Throughout this section only homomorphisms over a binary alphabet are considered. In section 7 it has been shown that $\{a^i b, ba^i\}^*$ is an equality set for each $i \geqq 1$. No other regular equality set ($\neq \{a, b\}^*$) freely generated by at least two words is known to us. Without claiming that no such set exists we make a somewhat weaker conjecture in this direction: Every regular equality set for homomorphisms over a binary alphabet is of the form F^* where F is of the cardinality at most two. Lemma 4.6, theorems 7.1, 7.2, 7.3 and 7.6 give considerable support to this conjecture. Further evidence is given in the discussion following theorem 7.6.

Actually, (4) in section 7 indicates that an even somewhat stronger statement than our conjecture would hold, namely that any two r-primitive elements of a regular equality set must start with distinct symbols. Hence, over a binary alphabet there could be at most two. That also would mean that any pair (g, h) where g is elementary, would have the "unique continuation property", i. e. there would not exist two minimal solutions with common proper prefixes.

In conclusion, we want to mention two results which would follow from the affirmative answer to our conjecture. A. Ehrenfeucht has conjectured that for every language L (over any alphabet) there exists a finite "test set" $F \subset L$, that is a set F such that every pair of homomorphisms agree word by word on L iff they agree on F. This has been shown to hold in the case of a binary alphabet in [5]. It is easy to see that our conjecture would imply a considerably simpler proof and also sharpen the result given in [5], namely it would imply that the test set could be always chosen (noneffectivelty) to be of the cardinality at most three.

The DOL sequence equivalence problem is decidable [3]. In [13] it is conjectured that the equivalence can be determined after considering no more than first $2n$ words in both sequences, where n is the cardinality of the alphabet. It is not hard to show that our conjecture would imply this for $n=2$.

Finally, although our conjecture does not imply the decidability of the Post's Correspondence Problem for lists of length two, it together with results of this paper supports the belief that this problem is decidable.

ACKNOWLEDGEMENT

The authors are grateful to the anonymous referee and C. Choffrut for very useful comments and improvements and especially for correcting a mistake in our original theorem 6.2.

REFERENCES

1. K. Culik II, *A Purely Homomorphic Characterization of Recurively Enumerable Sets*, J. Assoc. Comput. Mach., Vol. 26, 1979, pp. 345-350.
2. K. Culik II, *On Homomorphic Characterization of Families of Languages*. Proceedings of the Six Inter. Colloquium an Automata, Languages and Programming, Graz, Austria, pp. 161-170, July 1979.
3. K. Culik II and I. Fris, *The Decidability of the Equivalence Problem for DOL Systems*, Inf. and Control, Vol. 35, 1977, pp. 20-39.
4. K. Culik II and A. Salomaa, *On the Decidability of Homomorphism Equivalence for Languages*, J. Comp. System Sc., Vol. 17, 1978, pp. 163-175.
5. K. Culik II and A. Salomaa, *Test Sets and Checking words for Homomorphism Equivalence*, J. Comp. System Sc., Vol. 20, 1980, pp. 379-395.
6. K. Culik II and H. A. Maurer, *On Simple Representations of Language Families*, R.A.I.R.O., Informatique théorique/Theoretical Informatics, Vol. 13, No. 3, 1979, pp. 241-250.
7. A. Ehrenfeucht and G. Rozenberg, *Elementary Homomorphisms and a Solution to the DOL Sequence Equivalence Problem*, Theoretical Computer Science, Vol. 7, 1978, pp. 169-183.
8. J. Engelfriet and G. Rozenberg, *Equality Languages and Fixed Point Languages*, Inf. and Control, Vol. 43, 1979, pp. 20-49.
9. J. Karhumaki and I. Simon, *A Note on Elementary Homomorphisms and the Regularity of Equality Sets*, Bulletin E.A.T.C.S., No. 9, October 1979, pp. 16-24.
10. R. C. Lyndon and M. P. Schutzenberger , *The Equation $a^M = b^N c^P$ in a Free Group*, Michigan Math. J., Vol. 9, 1962, pp. 289-298.
11. G. S. Makanin, *The Problem of Solvability of Equations in a Free Semigroup* (in Russian), Matematiceskij Sbornik, Vol. 103, No. 145, 1977, pp. 148-236.
12. A. Salomaa, *Equality Sets for Homomorphisms of Free Monoids*, Acta Cybernetica, Vol. 4, 1978, pp. 127-139.
13. A. Salomaa, DOL *Equivalence: The Problem of Iterated Morphisms*, Bulletin E.A.T.C.S., No. 4, January 1978, pp. 5-12.