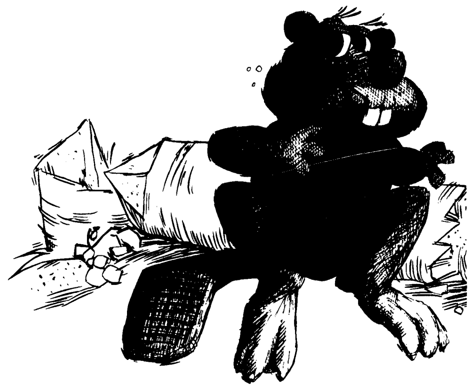
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT



Homomorphisms: Decidability, Equality and Test Sets

K. Culik II

CS-80-02

January 1980

Homomorphisms: Decidability, Equality and Test Sets

K. Culik II

Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

ABSTRACT

A number of recent results and open problems on homomorphisms on free monoids are discussed. Many of the results and conjectures state that various equivalence problems about homomorphisms are decidable. Also discussed are equality sets, test sets and new representation theorems for families of languages.

Presented at the International Symposium on Formal Language Theory, Santa Barbara, California, December 10-14, 1979.

Homomorphisms: Decidability, Equality and Test Sets

K. Culik II

Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

0. Introduction

We survey a number of recent results and open problems on homomorphisms on free monoids. Except for the last section, dealing with representation of language families, most of the results are decidability results. They were motivated or directly constitute problems in L-systems theory. However all of them are basic problems about free monoids and as such are not only of purely mathematical interest but also, since they are all simply formulated decidability problems, are of fundamental interest for theoretical computer science.

Whenever possible we give an algebraic formulation of each problem so that reading, not only the whole paper, but even a particular problem or theorem does not require any specialized knowledge. Open problems are specifically of interest, which makes us stress some topics. The only new results in this paper are some relations among the open problems (conjectures) mostly very easily shown.

In section 2 we deal with iterations of one or more homomorphisms (DOL, HDOL, DTOL systems) and some generalizations thereof. The next section is about "homomorphism equivalence on languages," i.e. the problem whether two given homomorphisms agree "string by string" on a given language, and its applications to transducers.

In Section 4 we consider elementary homomorphisms and questions about equality sets, in particular over a binary alphabet. In the next section we consider "homomorphism compatibility on languages," i.e. the problem whether there exists a string in given languages on which two given homomorphisms agree, in particular various restricted forms of the Post Correspondence Problem.

In Section 6 we discuss the Ehrenfeucht conjecture: Each language possesses a finite subset such that any two homomorphisms which agree (string by string) on the subset agree also on the whole language. Some partial solutions are discussed. Finally in the last section we list some new representation theorems for language families based on equality sets and related phenomena.

Research supported in part by the Natural Science and Engineering Research Council of Canada under Grant No. A-7403. Preparation of this paper was supported in part by the National Science Foundation under Grant No. MCS79-04012.

1. Preliminaries

We consider homomorphisms $\Sigma^* \to \Delta^*$, where Σ^*, Δ^* are free monoids generated by finite alphabets Σ, Δ. The monoid unit (empty word) is denoted by ϵ, the length of a word w in Σ^* by $|w|$. We also use $|n|$ to denote the absolute value of number n. The cardinality of set S is denoted by card S.

An alphabet Σ, homomorphism $h : \Sigma^* \to \Sigma^*$ and an (initial) word w in Σ^* form a *DOL system* $G = (\Sigma, h, w)$. *The sequence generated by* G, denoted $E(G)$, is defined by $E(G) = w, h(w), h^2(w), \ldots$; *the language generated by* G, denoted $L(G)$, *is defined by* $L(G) = \{h^n(w) \mid n \geqslant 0\}$.

A DOL system G and another homomorphism g form an HDOL system $K = \langle G, g \rangle$. It generates the sequence

$$E(K) = g(w), g(h(w)), g(h^2(w)), \ldots$$

and the language

$$L(K) = g(L(G)) = \{g(h^n(w)) \mid n \geqslant 0\}.$$

A DTOL system G is a tuple $(\Sigma, h_1, \ldots, h_n, w)$ where $h_i : \Sigma^* \to \Sigma^*$ for $i = 1, \ldots, n$. It generates a set of sequences

$$\{w, h_{i_1}(w), h_{i_1}(h_{i_2}(w)), \ldots \mid i_1, i_2, \ldots \in \{1, \ldots, n\}\}$$

and the language

$$L(G) = \{h_{i_1}(h_{i_2}(\ldots h_{i_k}(w)\ldots)) \mid i_1, \ldots, i_k \in \{1, \ldots, n\}\}.$$

For homomorphisms $g, h : \Sigma^* \to \Delta^*$ the *equality set for the pair* (g,h) is denoted by $E(g,h)$ and defined by $E(g,h) = \{x \in \Sigma^* \mid g(x) = h(x)\}$.

A *deterministic generalized sequential mapping* (*dgsm mapping*) is a mapping defined by deterministic generalized sequential machine with accepting states (dgsm) as in [32].

For other standard definitions and notations we refer the reader to [40, 41 or 47]

2. Iterated Homomorphisms

We will discuss a number of decision problems about iterative homomorphisms. The following problem and techniques used in its proof stimulated most of the research reported in this paper.

Theorem 2.1 [15] (DOL sequence equivalence problem). Given two homomorphisms $g, h : \Sigma^* \to \Sigma^*$ and w in Σ^* it is decidable whether $g^n(w) = h^n(w)$ for all $n \geqslant 0$.

The strategy of the solution of this problem is to show that any two (normal) equivalent systems must behave in certain "similar" ways and then to show the decidability for similar systems only. Here a pair of DOL systems is similar if the pair (g,h) has "bounded balance" on the language $\{g^n(w) \mid n \geqslant 0\}$.

The balance of a string w in Σ^* with respect to a pair of homomorphisms g,h on Σ^* is defined as

$$B(w) = |g(w)| - |h(w)|$$

The pair (g,h) is said to have *bounded balance* on language L if there is a $C > 0$ so that $|B(w)| \leqslant C$ for each prefix of every word in L.

A property of a pair of DOL systems $G_1 = (\Sigma,g,w)$ and $G_2 = (\Sigma,h,w)$ equivalent to "bounded balance" is introduced in [15]. The pair (G_1, G_2) is said to have a *true envelope* R if $L(G_1) \cup L(G_2) \subseteq R \subseteq E(g,h)$. Obviously, if a pair (G_1,G_2) has a true envelope, then G_1 and G_2 are sequence equivalent. It is shown in [15] that a pair of equivalent DOL systems (G_1,G_2) has a regular true envelope iff the pair of homomorphisms (g,h) has bounded balance on $L(G_1)$, and consequently that each pair of equivalent normal DOL systems has a regular true envelope. The latter result is extended in [25] to all pairs of equivalent DOL systems.

The "bounded balance technique" is also useful when testing homomorphism equivalence discussed in Section 3. (See [20]). The same holds also for another technique introduced in [8], the "shifting argument". Roughly speaking, it is used to show that if homomorphisms g,h agree on two words of the form xwy and uwv, i.e. with a common subword w, where w is "sufficiently long" and $|B(x) - B(u)|$ "sufficiently small", then either $B(x) = B(u)$ or $g(w)$ and $h(w)$ are periodic.

The bounded balance technique is not helpful in proving the following generalization of the DOL sequence equivalence problem.

Conjecture 2.2: (HDOL equivalence problem). Given four homomorphisms $g_1:\Sigma_1^*\rightarrow\Sigma_1^*$, $g_2:\Sigma_1^*\rightarrow\Sigma_2^*$, $h_1:\Delta_1^*\rightarrow\Delta_1^*$, $h_2:\Delta_1^*\rightarrow\Delta_2^*$ and strings $u \in \Sigma_1^*$, $v \in \Delta_1^*$. It is decidable whether $g_2(g_1^n(u)) = h_2(h_1^n(v))$.

We show later a problem equivalent to the HDOL equivalence problem (Theorem 3.3). There are two other interesting extensions of DOL equivalence which have been shown decidable by reducing them to DOL equivalence. (Theorems 2.3 and 2.6). The proof of the following theorem also uses results about monoids generated by integer matrices obtained by [34] and by [37].

Theorem 2.3: [10] (Ultimate sequence equivalence). Given two homomorphisms $g,h:\Sigma^*\rightarrow\Sigma^*$ and u,v in Σ^* it is decidable whether there exists $n \geqslant 0$ such that $g^k(u) = h^k(v)$ for all $k \geqslant n$.

It is natural to ask whether sequence equivalence remains decidable for more complicated mappings than homomorphisms, in particular for mappings defined symbol by symbol but in a context dependent manner. This is also strongly biologically motivated since such mappings abstract developmental systems of higher level where individual cells interact, i.e. their behaviour is context dependent. The simplest case is dependence on one symbol at the left, the so called D1L system. The sequence equivalence has been shown undecidable even for propagating (nonerasing) version of these systems.

Theorem 2.4: [52] The PD1L sequence equivalence problem is undecidable.

In the view of the last theorem it is rather surprising that the equivalence problem becomes decidable when the rewriting of a letter might depend on one neighbour from each side but only when the letter is being rewritten by at least two new letters. That is any letter-to-letter rewriting must be context free (no erasing is allowed). A deterministic system based on this type of rewriting is introduced in [16] and called an e-GD2L system. Two main results of [16] are that e-GD2L systems have essentially context-free behaviour and that the sequence equivalence for them is decidable. The former result could be compared to "Baker's Theorem" (29, Theorem 10.2.1) giving a condition under which context-sensitive grammar generates a context-free language.

Theorem 2.5: If the sequence $s_0, s_1, \ldots$ is generated by an e-GD2L system, then there exist a nonerasing homomorphism h and a letter-to-letter homomorphism (coding) g so that $s_n = g(h^n(s_0))$ for all $n \geqslant 0$.

Theorem 2.6: [16] The sequence equivalence problem for e-GD2L systems is decidable.

We are not directing our attention here to the languages generated by various parallel rewriting systems, but for completeness of the decidability results we mention the following two theorems. The DOL language equivalence had already been reduced to DOL sequence equivalence in [38] before the latter was shown to be decidable. Recently even the inclusion problem has been shown decidable.

Theorem 2.7: [45] The inclusion problem for DOL languages is decidable.

In the nondeterministic case we have the following result which follows from the undecidability of the equality problem for sentential forms of context free languages.

Theorem 2.8: [3] The equivalence problem for OL (even POL) languages is undecidable.

Another biologically important generalization of DOL systems is obtained when several starting strings and several homomorphisms (tables) are considered. Given two such systems with matching starting strings and matching pairs of homomorphisms we can ask whether all "matching" sequences are identical.

Consider

$$(h_1, \ldots, h_n), (h_1', \ldots, h_n') \tag{2.1}$$

where h_i, h_i' are homomorphisms $\Sigma^* \to \Sigma^*$, for $i = 1, \ldots, n$.

Conjecture 2.9: (DTOL sequence equivalence). Given strings $w, w' \in \Sigma^*$ and homomorphisms (2.1) it is decidable whether

$$h_{i_1}(h_{i_2}(\ldots h_{i_k}(w)\ldots)) = h'_{i_1}(h'_{i_2}(\ldots h'_{i_k}(w')\ldots))$$

for all $i_1 i_2 \ldots i_k$ in $\{1, \ldots, n\}^*$.

Lemma 2.10: [20] Conjecture 2.9 holds if it holds for $n = 2$ (two tables).

Later we show another conjecture equivalent to Conjecture 2.9 (Theorem 3.4).

Note that the DTOL language equivalence problem has been shown undecidable in [39] and recently [46] it has been shown that it becomes decidable if only one system is a DTOL system and the other is DOL. This is a strengthening of the decidability of DOL language equivalence.

All the decidable problems mentioned in this section, as well as some other problems in L-systems (see e.g. [22]) have been shown decidable by reducing them to the DOL sequence equivalence problem (Theorem 2.1). Another problem shown decidable in the same way has been the equivalence problem for simple single loops programs with respect to symbolic evaluation [33].

3. Homomorphism Equivalence on a Language

The problems discussed in this section originated in a simple observation in the proof of decidability of DOL equivalence problem [8,15]. The first step in the proof was that given homomorphisms $g,h:\Sigma^* \to \Sigma^*$ and w in Σ^* the following two conditions are clearly equivalent.

(i) $g^n(w) = h^n(w)$ for all $n \geqslant 0$;
(ii) $g(u) = h(u)$ for all u in L, $L = \{g^n(w) : n \geqslant 0\}$.

So, the testing of iterative equivalence of two homomorphisms g,h can be reduced to the testing of string by string equivalence of g and h on a certain language, namely the language generated by g from the "starting string" w. It is natural and also very useful (cf. Theorems 3.10 and 3.11) to attempt such testing also for other types of languages.

The problem to test whether two homomorphisms agree (string by string) on a given language from family L is called the *homomorphic equivalence problem for L* [20]. Its decidability for regular sets was already implicitly contained in [15]. The following is the main result from [20].

Theorem 3.1: [20] (Homomorphism equivalence for CFL). Given a context free language $L \subseteq \Sigma^*$ and homomorphisms $h,g:\Sigma^* \to \Delta^*$, it is decidable whether $h(x) = g(x)$ for each $x \in L$.

The decidability of homomorphic equivalence is open for all families of languages between DOL and indexed. In particular we have the following:

Conjecture 3.2: (Homomorphism equivalence for DOL languages). Given w in Σ^* and homomorphism $h : \Sigma^* \to \Sigma^*$ and $f,g : \Sigma^* \to \Delta^*$ it is decidable whether

$$f(h^n(w)) = g(h^n(w)) \tag{3.1}$$

for all $n \geqslant 0$.

The following is mentioned in [20].

Theorem 3.3: Conjecture 2.2 is equivalent to Conjecture 3.2, i.e. the HDOL equivalence problem is decidable iff the homomorphism equivalence problem for

DOL languages is decidable.

Proof: 1. To test (3.1) means to compare two HDOL sequences based on the same DOL system. 2. Given $u \in \Sigma^*$, $v \in \Delta^*$ and homomorphisms $g_1 : \Sigma^* \to \Sigma^*$, $g_2 : \Sigma^* \to \Gamma^*$, $h_1 : \Delta^* \to \Delta^*$, $h_2 : \Delta^* \to \Gamma^*$. Assume without loss of generality that $\Sigma \cap \Delta = \phi$ and define homomorphisms f, f_1 and $f_2 : (\Sigma \cup \Delta)^* \to (\Sigma \cup \Delta)^*$ by $f(a) = g_1(a)$ for $a \in \Sigma$, $f(b) = h_1(b)$ for $b \in \Delta$, $f_1(a) = g_2(a)$, $f_2(a) = \epsilon$ for $a \in \Sigma$, $f_1(b) = \epsilon$, $f_2(b) = h_2(b)$ for $b \in \Delta$. Then, clearly, $f_1(f^n(uv)) = f_2(f^n(uv))$ for all n iff $g_2(g_1^n(u)) = h_2(h_1^n(v))$ for all n.

Using similar techniques as in the proof of Theorem 3.2 we also get the following reduction result.

Theorem 3.4: The following three problems are equivalent (and thus all conjectured to be decidable by Conjecture 2.9).

(a) DTOL sequence equivalence problem;
(b) HDTOL sequence equivalence problem;
(c) Homomorphism equivalence problem for DTOL languages.

Proof: We show the reduction (c) to (a); the others are easier.

Let $\overline{\Sigma} = \{\overline{a} \mid a \in \Sigma\}$ and for w in Σ let $\overline{w}$ denote the word obtained from w by "barring" each symbol. Given DTOL system $G = (\Sigma, h_1, h_2, w)$ and homomorphisms g_1, g_2, we construct DTOL systems $G_i = (\Sigma \cup \overline{\Sigma}, h_1', h_2', f_i, w)$ for $i = 1, 2$, where $h_j'(a) = h_j(a)$, $h_j'(\overline{a}) = \epsilon$ for all $a \in \Sigma$ and $j = 1, 2$; $f_i(a) = \overline{g_i(a)}$, $f_i(\overline{a}) = \epsilon$ for all a in Σ and $i = 1, 2$.

Since $h_j'(f_i(u) = \epsilon$ for all $i, j = 1, 2$ and $u \in \Sigma^*$, it is easy to verify that G_1 and G_2 are sequence equivalent iff homomorphisms g_1 and g_2 are equivalent on $L(G)$.

In [20] it has been conjectured that even a much stronger result than Conjecture 3.2 holds. However, in the view of Theorem 3.3 we cannot expect it to be easy to prove the following:

Conjecture 3.5: The homomorphism equivalence problem for indexed languages is decidable.

For the special case of elementary homomorphisms (see Section 4) decidability has been shown using Theorem 4.4.

Theorem 3.6: [49] It is decidable whether two given elementary homomorphisms are equivalent on a given indexed language.

The following is a partial solution of Conjecture 3.5, which is incomparable with Theorem 3.1. It is based on the fact that every homomorphism on a binary alphabet is either elementary or periodic with the same period for each letter (see Section 4), and on Theorem 3.6.

Theorem 3.7: [19] The homomorphism equivalence problem for ETOL languages over a binary alphabet is decidable.

Finally, we have an easy undecidability result:

Theorem 3.8: [20] The homomorphism equivalence problem for (deterministic) context-sensitive languages is undecidable.

We conclude this section with applications of Theorem 3.1 to problems about finite and push-down transducers [12]. All these quite powerful results follow easily from Theorem 3.1. Note, for example, that the equivalence problem for deterministic generalized sequential machines is a very special case of Theorem 3.11.

We call a transducer defining a regular (rational) translation a finite transducer (a-transducer in [28]). In [1] it has been shown that regular (rational) and push down translations can be homomorphically characterized, i.e. each regular or push-down translation t can be expressed as

$$t = \{(g(w), h(w)) : w \in L\}$$

where g,h are homomorphisms and L is regular or context free, respectively. Therefore, we immediately obtain by Theorem 3.1:

Theorem 3.9: [12] Given a finite transducer or a push down transducer it is decidable whether it defines an identity relation restricted to its domain.

From Theorem 3.9 we easily obtain the following:

Theorem 3.10: [12] Given a finite transducer M and a context-free grammar G, it is decidable whether t_M (the relation defined by M) is functional on $L(G)$.

The inverse relation of the restriction of t_M to $L(G)$ is not necessarily equal to the restriction of t_M^{-1} to $t_M(L(G))$. Hence it does not follow as a corollary of Theorem 3.10, as claimed in [12], that it is decidable whether t_M is one-to-one on $L(G)$. Actually this problem has been shown to be undecidable in [30]. However, we can test whether t_M is one-to-one (on its domain).

Among the other consequences of Theorem 3.1 shown in [12] is the decidability of the equivalence problem for functional finite transducers, or the even stronger result which follows, where an unambiguous pushdown transducer is a p.d.t. based on an unambiguous pushdown automaton [32].

Theorem 3.11: [12] (Equivalence between a functional finite transducer and an unambiguous pushdown transducer). Given an unambiguous pushdown transducer P and a functional finite transducer it is decidable whether $t_P = t_M$.

4. Elementary Homomorphisms and Equality Sets

Here we consider a very useful special type of homomorphism first introduced in [23], equality sets for them and equality sets over a binary alphabet.

A homomorphism $h : \Sigma^* \to \Delta^*$ is *elementary* if there is no decomposition of h into homomorphisms f and g, that is $h = gf$:

$$\begin{tikzcd}
\Sigma^* \arrow[rr, "h"] \arrow[dr, "f"'] & & \Delta^* \\
& \Gamma^* \arrow[ur, "g"'] &
\end{tikzcd}$$

such that card $\Gamma <$ card Σ. Similarly, a finite language L is *elementary* if there is no language K such that card $K <$ card L and $L \subseteq K^*$. Let $L_h = \{h(a) : a \in \Sigma\}$. Clearly, a homomorphism $h : \Sigma^* \to \Delta^*$ is elementary iff card L_h = card Σ and L_h is elementary.

Properties of elementary homomorphisms and languages were studied in [23, 24, 41 and 35]. We mention a few of them.

Theorem 4.1: [23] Each elementary homomorphism is injective.

Theorem 4.2: [41] Let $L = \{u_1, \ldots, u_n\}$ be an elementary language over the alphabet Σ. If $u_i x z = u_j y$ for some $i \neq j$, $x, y \in L^*$ and $z \in \Sigma^*$, then $| u_i x | \leqslant | u_1 u_2, \ldots, u_n | - n$.

Corollary 4.3: [41] Every elementary language is a code with bounded delay (both from left to right and from right to left).

The following is an important result. In particular it has made it possible to simplify the proof of the decidability of DOL sequence equivalence.

Theorem 4.4: [24] If homomorphisms g and h are elementary, then the equality set $E(g,h)$ is regular.

This result has been strengthened in [26] for the weaker assumption that at least one of g and h is elementary and then for even weaker assumptions in [35]. No effective proof even for the weakest result is known so we have the following open problem and its even harder versions.

Open Problem 4.5: Given elementary homomorphisms $g, h : \Sigma^* \to \Delta^*$, can the regular set $E(g,h)$ (represented e.g. by a regular expression) be found effectively?

This problem is presently open even for the case of binary alphabets [17]. A positive answer in this special case already implies the validity of Conjecture 5.2, the decidability of PCP restricted to lists of length two.

The fact that every homomorphism on a binary alphabet is either elementary or periodic severely restricts the form of equality sets for homomorphisms over a binary alphabet (on free monoids with two generators). Each equality set is either regular or of the form $\{ w \in \{a,b\}^* \mid \#a(w)/\#b(w) = k \}$ for some rational $k \neq 0$, where $\#a(w)$ is the number of occurrences of letter a in w. In [17] an attempt has been made to fully classify such equality sets. In particular for some words all possible homomorphisms agreeing on them are shown. On the other hand a number of sets or words (singleton sets) are shown to be "periodicity forcing," meaning that only periodic homomorphisms could agree on them. These results support the following:

Conjecture 4.6: Every regular equality set for homomorphisms over a binary alphabet is of the form F^* where F is of cardinality at most two.

This conjecture would imply a simple proof of Theorem 6.2 and also sharpen this theorem, namely it would imply that for $L \subseteq \{a,b\}^*$ there always exists (noneffectively) a test set (see Section 6) of cardinality at most three. Some other implications of Conjecture 4.6 are discussed in [17].

Note that there is no loss of generality in assuming that the range of considered homomorphisms is over a binary alphabet, since a larger alphabet can always be encoded into a binary one, preserving the equality set. This is, of course, not the case for the domain. Hence, we have a rather unusual situation that many problems considered here are much easier for a binary alphabet than in the general case. One such example is the DOL equivalence problem.

5. Homomorphism Compatibility

In Section 3 we were interested in testing whether two given homomorphisms agree "string by string" on a given language. In [20] four kinds of "homomorphism agreements" were considered, namely, compatibility, strong compatibility, ultimate equivalence and equivalence. The last one was considered in Section 3, here we will consider the first one, the other two are omitted since the results for them are similar to the two cases considered.

Homomorphisms g and h are *compatible* on a language L if $g(w) = h(w)$ for some w in L, that is if $L \cap E(g,h) \neq \phi$.

To decide whether homomorphisms g and h are compatible on Σ^+, i.e. whether $E(g,h) - \{\epsilon\} \neq \phi$, is nothing else but the Post Correspondence Problem (PCP). An instance PCP(g,h) is given by two nonerasing homomorphisms $g,h : \{1,\ldots,n\}^* \to \Sigma^*$, traditionally called *lists* of length n.

The problem of homomorphism compatibility for a family of languages $\mathcal{L}$ can be stated as: given L in $\mathcal{L}$ and homomorphisms g and h, to decide whether there is w in L such that $g(w) = h(w)$. Hence, this problem is undecidable for any family containing Σ^+ for arbitrarily large alphabet Σ. Actually, it is known that there is a certain fixed size for Σ, which is sufficient to make the PCP undecidable. However the minimal size is not known. We have the following open problem and conjecture.

Open Problem 5.1: What is the minimal integer n such that PCP with only lists of length n is undecidable?

The results in [17] support the following generally accepted but not yet proven

Conjecture 5.2: The PCP with the restriction to lists of length two, i.e. homomorphic compatibility on Σ^+ for a binary Σ, is decidable.

More difficult to prove would be the following:

Conjecture 5.3: The PCP restricted to instances PCP(g,h) with elementary homomorphisms (lists) g and h is decidable.

Open is also the modification of Conjecture 5.3 obtained by assuming that the homomorphisms are injective rather than elementary.

Theorem 5.4: The positive solution of Problem 4.5 implies the validity of Conjectures 5.2 and 5.3.

Proof: If g and h are elementary, then we can effectively find a regular set $E(g,h)$ and test whether $E(g,h) \neq \{\epsilon\}$. This validates Conjecture 5.3, in the case of Conjecture 5.2. There remains the easy case when at least one of g,h is periodic, see [17].

Obviously, $\#\Sigma^*$, for every alphabet Σ and $\# \notin \Sigma$, is a DTOL language. Hence, the homomorphism compatability problem is clearly undecidable for the family of DTOL languages. However, we have the following:

Open Problem 5.5: (Homomorphism compatibility on DOL languages). Given w in Σ^*, and homomorphisms $h:\Sigma^* \to \Sigma^*$, $g,f:\Sigma^* \to \Delta^*$, is it decidable whether there is an $n \geqslant 0$ such that $g(h^n(w)) = f(h^n(w))$?

Related are the following two problems:

Open Problem 5.6: (Intersecting DOL sequences). Given u,v in Σ^* and homomorphisms $g,h:\Sigma^* \to \Sigma^*$, is it decidable whether $g^n(u) = h^n(v)$ for some $n \geqslant 0$?

Open Problem 5.7: (Intersecting HDOL sequences). Given u in Σ^*, v in Δ^* and homomorphisms

$$g_1:\Sigma^* \to \Sigma^*,$$
$$g_2:\Sigma^* \to \Gamma^*,$$
$$h_1:\Delta^* \to \Delta^*,$$
$$h_2:\Delta^* \to \Gamma^*$$

is it decidable whether $g_2(g_1^n(u)) = h_2(h_1^n(v))$ for some n?

Theorem 5.8: Problems 5.5 and 5.7 are equivalent, the decidability of Problem 5.5 implies the decidability of Problem 5.6.

Proof: Similar, to the proof of Theorem 3.3.

If we modify Problem 5.6 so that only the length of generated strings is compared we obtain the following problem which is shown in [51] to be equivalent to the well known open problem of finding zeros of Z-rational functions.

Open Problem 5.9: (Intersecting DOL growth sequences). Given u in Σ^*, v in Δ^* and $h : \Delta^* \to \Delta^*$ decide whether $|g^n(u)| = |h^n(v)|$ for some n.

There are a large number of results (and open problems) concerning growth (length) and Parikh vector sequences generated by one or more iterative homomorphisms. Mathematically, they belong to the theory of noncommutative formal power series and we refer the interested reader to [51]. We have included Problem 5.9 here because of its strong implications to our other open problems shown in the following theorem. The first part was shown in the terminology of Z-rational functions in [42] the second follows by Theorem 3.3.

Theorem 5.10: The decidability of Problem 5.9 implies

(i) The decidability of the HDOL sequence equivalence problem (Conjecture 2.2).

(ii) The decidability of the homomorphism equivalence problem for DOL languages.

Clearly, the decidability of Problem 5.7 (or 5.5) implies the decidability of Problem 5.9 and therefore, by Theorem 5.10, also the decidability of (i) and (ii) above.

In [17] a problem which can be considered dual to the Post Correspondence Problem is shown to be decidable by reducing it to Makanin's result concerning solvability of equations in free monoids [36].

Theorem 5.11: Given a string w in Σ^*, it is decidable whether there exist two distinct homomorphisms $g,h : \Sigma^* \to \Delta^*$ for some Δ, such that at least one of them is aperiodic and $g(w) = h(w)$.

Note that the problem is trivial if g and h are not required to be distinct or aperiodic.

HOMOMORPHISM COMPATIBILITY FOR DOL LANGUAGES

$\Updownarrow$

INTERSECTION OF HDOL SEQUENCES

$\swarrow\!\!\!\swarrow$ $\qquad$ $\searrow\!\!\!\searrow$

INTERSECTION OF DOL SEQUENCES

INTERSECTION OF DOL GROWTH SEQUENCES

$\Updownarrow$

IMMEDIATE REPETITION IN A DOL GROWTH SEQUENCE

$\Updownarrow$

ZERO OF Z-RATIONAL FUNCTION

$\Downarrow$

HDOL SEQUENCE EQUIVALENCE

$\Updownarrow$

HOMOMORPHISM EQUIVALENCE FOR DOL LANGUAGES

RELATIONS BETWEEN OPEN DECIDABILITY PROBLEMS
(SEE 5.5-5.10, $A \Rightarrow B$ READS THE DECIDABILITY OF A
IMPLIES THE DECIDABILITY OF B)

6. Test Sets and Checking Words

The very interesting "Ehrenfeucht conjecture" (Conjecture 6.1) is at least several years older than the notion of homomorphism equivalence to which it is closely related.

We say that a finite subset F of a language L is a *test set for* L if, for any pair of homomorphisms (g,h), $g(x) = h(x)$ for all x in L if and only if $g(x) = h(x)$ for all x in F, i.e. g and h are equivalent on L iff g and h are equivalent on F.

Conjecture 6.1: For every language there exists a test set.

It immediately follows by Theorem 3.8 that given a context sensitive grammar G a test set for $L(G)$ cannot be effectively constructed, since the effective existence of a test set for a family $\mathcal{L}$ obviously implies the decidability of homomorphism equivalence for $\mathcal{L}$.

The discussion in Section 4 indicates that proving the validity of the Ehrenfeucht conjecture is considerably easier for languages over a binary alphabet. This has actually has been done recently in [21].

Theorem 6.2: [21] For each language $L \subseteq \Sigma^*$, where Σ is a binary alphabet, there exists a test set F, i.e. a finite set F, such that for each pair of homomorphisms g,h, $g(x) = h(x)$ for all x in L iff $g(x) = h(x)$ for all x in F.

It easily follows from the discussion in (15 or 20) that for each regular set there effectively exists a test set. Recently this result has been extended to context free languages.

Theorem 6.3: [2] For each context free language there effectively exists a test set.

As mentioned above, this result immediately implies the decidability of homomorphism equivalence on the CFL (Theorem 3.1).

Actually, a somewhat stronger form of Theorem 6.3 is shown in [2], namely that given a CFG $G = (N,T,P,S)$ with n=card N and m the maximal length of the right side of the a production in P,

$$F = \{w \in L : |w| \leqslant m^{3n+1}\}$$

is a test set for $L(G)$.

This result is then used to obtain also finite "test sets" for CFL with respect to gsm mappings realized by gsm with a uniformly bounded number of states. Despite our reasons for expecting it to be hard to prove Conjecture 3.2 (equivalent to Conjecture 2.2) we venture to make an even stronger one:

Conjecture 6.4: For every indexed language (given by an indexed grammar) there effectively exists a test set.

According to [21] a word in Σ^* is a *checking word* for a language $L \subseteq \Sigma^*$ if, for any pair of homomorphisms (g,h), $g(x) = h(x)$ for all x in L if and only if $g(w) = h(w)$. Observe that it is not required that w be in L, hence $\{w\}$ might not be a test set for L.

A language L is *rich* if two homomorphisms g and h are equivalent on L only in case $g = h$. Somewhat surprisingly it is easier to show the following result than Theorem 6.2.

Theorem 6.5: [21] Every language over a binary alphabet is either rich or possesses a checking word.

7. Representation of Language Families

Equality sets have already been discussed in the previous sections mainly in relation to decidability problems about homomorphisms. They were explicitly introduced in [24] under a different name and their basic properties were studied in [49]. A generalization of equality sets to more than two homomorphisms is considered by [6].

It turned out that equality sets and similarly fixed point languages [26] provide simple representations of the recursively enumerable languages [11, 27, 49] which can also be extended to various time and complexity classes [4, 5, 13, 14]. The first results say that the closure of equality sets under dgsm mappings is the family of recursively enumerable languages.

Theorem 7.1: [27, 49] For every recursively enumerable (r.e.) set L, there exists a pair of homomorphisms (h_1, h_2) and a dgsm mapping g such that $L = g(E(h_1, h_2))$.

This result has been strengthened in [11] by replacing dgsm's by erasings and equality sets by minimal equality sets.

A homomorphism $h : \Sigma^* \rightarrow \Delta^*$ is called an *erasing* if, for each $a \in \Sigma$ either $h(a) = a$ or $h(a) = \epsilon$. For homomorphisms $g, h : \Sigma^* \rightarrow \Delta^*$, *the minimal equality set* is the set $e(g, h) = \{w \in \Sigma^+ \mid g(w) = h(w)$ and if $w = uv$ for $u, v \in \Sigma^+$, then $g(u) \neq h(u)\}$, that is, using the notation of [32],

$$e(g, h) = \min(E(g, h)) - \{\epsilon\}.$$

Theorem 7.2: [11] For every r.e. language L, there exist homomorphisms h_1, h_2 and an erasing h_0 so that $L = h_0(e(h_1, h_2))$.

Also we have a representation based on fixed points of dgsm mappings. Let $g : \Sigma^* \rightarrow \Sigma^*$ be a function. The *fixed-point language* $F_p(g)$ *of the function* g is defined to be $F_p(g) = \{w \in \Sigma^* \mid g(w) = w\}$.

Theorem 7.3: [27] For every r.e. language L, there exists a dgsm mapping g and an erasing h such that $L = h(F_p(g))$.

By imposing simple restrictions on the mappings used in Theorems 7.1, 7.2 and 7.3 we obtain a simple "machine independent" characterization of many time and space complexity classes of languages. Similar results have been shown independently in [4, 5] and in [13, 14]. We give here a few of these results.

We say that $\mathcal{C}$ is a class of complexity functions if $\mathcal{C}$ is a class of functions closed under addition of and multiplication by a constant. A language L is of time (space) complexity $\mathcal{C}$ if L is accepted by a nondeterministic multitape on-line Turing machine M which operates within time-bound (space-bound) f, for some f in $\mathcal{C}$, we write $L \in \mathrm{NTIME}(\mathcal{C})$ $(L \in \mathrm{NSPACE}(\mathcal{C}))$.

We generalize the notion of k-limited erasing [32] as follows: For a function f on the integers we say that *an erasing h is f-bounded on a language L* if for each w in L, $w = xyz$ and $h(y) = \epsilon$ implies $|y| \leqslant f(|w|)$, that is at most $f(|w|)$ consecutive symbols of w may be erased. We say that h is $\mathcal{C}$-bounded, for a class $\mathcal{C}$ of complexity functions, if h is f-bounded for some f from $\mathcal{C}$.

We get the following "machine independent" characterization of the time complexity classes of languages.

Theorem 7.4: [5,14] Let $\mathcal{C}$ be a class of complexity functions closed under squaring. Then the following three conditions are equivalent.

(i) $L \in \mathrm{NTIME}(\mathcal{C})$

(ii) $L = h_0(e(h_1, h_2))$ where h_1, h_2 are homomorphisms and h_0 is a $\mathcal{C}$-bounded erasing on $e(h_1, h_2)$.

(iii) $L = h(F_p(g))$ where g is a *dgsm* mapping and h is an erasing on $F_p(g)$.

Corollary: A language L is in NP iff there exist homomorphisms h_0, h_1, h_2 such that $h_0(e(h_1, h_2)) = L$ and h_0 is polynomial-bounded erasing on $e(h_1, h_2)$.

Corollary: A language L is primitive recursive (recursive) iff there exist homomorphisms h_0, h_1, h_2 such that $h_0(e(h_1, h_2)) = L$ and h_0 is primitive recursive- (recursive-) bounded erasing on $e(h_1, h_2)$.

In order to get a characterization of space complexity classes we need to generalize the notion of bounded balance on a language considered in Section 3.

Consider two fixed homomorphisms $g,h : \Sigma^* \to \Delta^*$. Recall that for each w in Σ^*, the balance of w is defined as $B(w) = |g(w)| - |h(w)|$. Now, for a monotone function f on the integers, an erasing π and $L \subseteq \Sigma^*$ we say that the *pair (g,h) has f-bounded balance on language L, with respect to erasing π*, if for each x in L and each prefix w of x we have $|B(w)| \leqslant f(|h(x)|)$. For a complexity class $\mathcal{C}$, we say that (g,h) *has $\mathcal{C}$-bounded balance on L, with respect to π*, if the same holds true for some $f \in \mathcal{C}$.

Theorem 7.5: [14] Let $\mathcal{C}$ be any class of complexity functions. Then $L \in \mathrm{NSPACE}(\mathcal{C})$ iff $L = h_0(e(h_1,h_2))$ where h_0 is an erasing and the pair (h_1,h_2) has $\mathcal{C}$-bounded balance on $e(h_1,h_2)$ with respect to h_0.

Corollary: A language L is context sensitive iff there exist an erasing h_0 and homomorphisms h_1, h_2 such that $L = h_0(e(h_1, h_2))$ and the pair (h_1, h_2) has linear-bounded balance on $e(h_1, h_2)$.

The class NP also has an alternative characterization:

Theorem 7.6: [5] The class NP is the smallest class containing all equality sets $E(h_1, h_2)$ (of nonerasing homomorphisms) with square-root-bounded balance (i.e. with the pair (h_1, h_2) having square-root-bounded balance on $E(h_1, h_2)$ with respect to the identity) and closed under intersection with regular sets and polynomial-bounded erasings.

We also have an alternative necessary condition for PSPACE:

Theorem 7.7: [5] For every language L in PSPACE there is a pair of nonerasing homomorphisms (h_1, h_2) with log n-bounded balance, a regular set R, and an erasing h such that $L = h(E(h_1, h_2) \cap R)$ and for some constants $c > 1$, $k > 0$, h is c^{n^k}-bounded on $E(h_1, h_2) \cap R$.

The study of equality sets has contributed representation theorems of the following form:

Let $\mathcal{L}$ be a family of languages over an alphabet Σ. Then there exist a language L_Σ and an erasing π_Σ such that $L \in \mathcal{L}$ iff $L = \pi_\Sigma(L_\Sigma \cap R_L)$ for some regular set R_L.

Letting $\mathcal{L}$ being the family of context free languages we have the well known Chomsky-Schützenberger theorem with L_Σ being the Dyck language over Σ; for the families of EOL and ETOL languages such representation has been established in [7].

Now, given Σ let $\overline{\Sigma}$ denote the alphabet disjoint from Σ consisting of "barred" symbols, $\overline{\Sigma} = \{\overline{a} \mid a \in \Sigma\}$, and for any word x in Σ^*, let $\overline{x}$ denote the word obtained from x by barring each symbol. To get a representation as above for the r.e. sets, the *twin-shuffle* over Σ has been defined in [27] as

$$L(\Sigma) = \{x \in (\Sigma \cup \overline{\Sigma})^* \mid \overline{\pi_\Sigma(x)} = \pi_{\overline{\Sigma}}(x)\}$$

where $\pi_\Sigma, \pi_{\overline{\Sigma}}$ are erasings on $(\Sigma \cup \overline{\Sigma})^*$, which only preserve the symbols from Σ, $\overline{\Sigma}$ respectively.

Clearly, the twin-shuffle $L(\Sigma)$ is an equality set of two homomorphisms. They cannot be nonerasing as shown in [5].

Theorem 7.8: [27] Let L be an r.e. language over Σ. There exists a regular set $R_L \subseteq (\Sigma \cup \overline{\Sigma} \cup \{0, \overline{0}, 1, \overline{1},\})^*$ such that $L = \pi_\Sigma(L(\Sigma \cup \{0, 1\}) \cap R_L)$, where π_Σ is the erasing which only preserves the symbols from Σ.

This theorem has essentially been shown in [27] using Theorem 7.3. It also follows by Theorem 7.2, see [13, 18].

It is actually not difficult to show that every principal cone has a representation as above. A family of languages $\mathcal{L}$ is a *principal cone* if there is an L in $\mathcal{L}$ such that $\mathcal{L}$ is the closure of $\{L\}$ under the operations of homomorphism, inverse homomorphism and intersection with a regular set, or equivalently $\mathcal{L}$ is the closure of $\{L\}$ under finite transducers (rational relations), see [28].

Theorem 7.9: [13,18] Let Σ be an alphabet and $\mathcal{L}$ a principal cone. There exists a language L_Σ in $\mathcal{L}$ such that for each L in $\mathcal{L}$, $L \subseteq \Sigma^*$, there exists a regular set R_L such that $L = \pi_\Sigma(L_\Sigma \cap R_L)$ where π_Σ is the erasing, which only preserves the symbols in Σ.

We refer to [4,5,13,14,18,27] for a number of additional representation results.

REFERENCES

[1] Aho, A.V. and Ullman, J.D., *The Theory of Parsing, Translation, and Compiling*, Prentice-Hall (1972).

[2] Albert, J. and Culik II, K., Test sets for homomorphism equivalence on context-free languages, University of Waterloo, Research Report CS-79-39.

[3] Blattner, M. The unsolvability of the equality problem for sentential forms of context-free languages, *J. Comp. System Sci.*, vol. 7, pp. 463-468 (1973).

[4] Book, R. and Brandenburg, F.-J., Representing complexity classes by equality sets, Proceedings of the Sixth ICALP, Graz, *Lecture Notes in Computer Science*, vol. 71, Springer Verlag, pp. 49-57 (1979).

[5] Book, R. and Brandenburg, F.-J., Equality sets and complexity classes, *SIAM J. Computing*, to appear.

[6] Brandenburg, F.-J., Multiple equality sets and Post machines, Math. Tech. Rep., University of California, Santa Barbara, (1979).

[7] Culik II, K., On some families of languages related to development systems, *International J. of Computer Math.*, Section A, 4, pp. 31-42 (1974).

[8] Culik II, K., On the decidability of the sequence equivalence problem for DOL-systems, *Theor. Computer Science* vol. 3, pp. 75-84 (1977).

[9] Culik II, K., The decidability of v-local catenativity and of other problems of DOL systems, *Inf. Process. Lett.* vol. 7 (1): pp. 33-35 (1978).

[10] Culik II, K., The ultimate equivalence problem for DOL systems, *Acta Informatica* vol. 10, pp. 69-84 (1978).

[11] Culik II, K., A purely homomorphic characterization of recursively enumerable sets, *JACM* vol. 6, pp. 345-350 (1979).

[12] Culik II, K., Some decidability results about regular and push down translations, *Information Processing Letters* vol. 8, pp. 5-8 (1979).

[13] Culik II, K., On homomorphic characterization of families of languages, Proceedings of the Sixth ICALP, Graz, *Lecture Notes in Computer Science 71*, Springer Verlag, pp. 161-170 (1979).

[14] Culik II, K. and Diamond, N.D., A homomorphic characterization of time and space complexity classes of languages, *Int. J. of Computer Math.*, to appear.

[15] Culik II, K. and Fris, J., The decidability of the equivalence problem for DOL-systems, *Inf. and Control*, vol. 35, pp. 20-39 (1977).

[16] Culik II, K. and Karhumaki, J., Interactive L-systems with almost interactionless behaviour, *Inf. and Control*, vol. 43, pp. 83-100 (1979).

[17] Culik II, K. and Karhumaki, J., On the equality sets for homomorphisms on free monoids with two generators, *R.A.I.R.O.*, to appear, also University of Waterloo Research Report CS-79-17, (1979).

[18] Culik II, K. and Maurer, H.A., On simple representations of languages families, *R.A.I.R.O.*, to appear, also University of Waterloo Research Report CS-78-41, (1978).

[19] Culik II, K. and Richier, J.L., Homomorphism equivalence on ETOL languages, *Int. J. of Computer Math.*, Section A, vol. 7, pp. 43-51 (1979).

[20] Culik II, K. and Salomaa, A., On the decidability of homomorphism equivalence for languages, *J. Comp. System Sci.*, vol. 17, pp. 613-175 (1978).

[21] Culik II, K. and Salomaa, A., Test sets and checking words for homomorphism equivalence, submitted to *J. Comp. System Sci.*, also University of Waterloo Research Report CS-79-04, (1979).

[22] Culik II, K. and Wood, D., Double deterministic tabled OL systems, *Int. J. of Computer and Inf. Science* vol. 8, pp. 335-347 (1979).

[23] Ehrenfeucht, A. and Rozenberg, G., Simplifications of homomorphisms, *Inf. and Control*, vol. 38 pp. 298-308 (1978).

[24] Ehrenfeucht, A. and Rozenberg, G., Elementary homomorphisms and a solution to the DOL sequence equivalence problem, *Theoretical Computer Science* vol. 7, pp. 169-183 (1978).

[25] Ehrenfeucht, A. and Rozenberg, G., Every two equivalent DOL systems have a regular true envelope, *Theoret. Comp. Science*, to appear.

[26] Engelfriet, J. and Rozenberg, G., Equality languages and fixed point languages, *Inf. and Control*, vol. 43, pp. 20-49 (1979).

[27] Engelfriet, J. and Rozenberg, G., Fixed point languages, equality languages and representation of recursively enumerable alanguages, *JACM*, to appear.

[28] Ginsburg, S., *Algebraic and Automata-Theoretic Properties of Formal Languages*, North Holland/American Elsevier, (1975).

[29] Harrison, M.A., *Introduction to Formal Language Theory*, Addison-Wesley, (1978).

[30] Head, T., Unique decipherability relative to a language, Manuscript (1979).

[31] Herman, G.T. and Rozenberg, G., *Developmental Systems and Languages*, (North-Holland, New York,) (1975).

[32] Hopcroft, J.E. and Ullman, J.D., *Formal Languages and Their Relation to Automata*, Addison-Wesley, (1969).

[33] Howden, W.E., Lindenmayer grammars and symbolic testing, *Inf. Processing Letters*, vol. 7, pp. 36-45 (1978).

[34] Jacob, G., Un algorithme calculant le cardinal, fini ou infini, des demi groups de matrices, *Theoret. Comp. Science*, vol. 5, pp. 183-204 (1977).

[35] Karhumaki, J. and Simon, I., A note on elementary homomorphisms and the regularity of equality sets, *Bulletin of the EATCS*, no. 9, pp. 16-24 (Oct. 1979).

[36] Makanin, G.S., The problem of solvabiilty of equations in a free semigroup (in Russian), *Matematiceskij Sbornik* vol. 103 (145), pp. 148-236 (1977).

[37] Mandel, A. and Simon, I., On finite semigroups of matrices, *Theoretical Computer Science*, vol. 5, pp. 101-111 (1977).

[38] Nielsen, M., On the decidability of some equivalence problems for DOL-systems, *Inform. Control*, vol. 25, pp. 166-193 (1974).

[39] Rozenberg, G., The equivalence problem for deterministic TOL systems is indecidable, *Inf. Process. Lett.* 1:201-204, errata, 252 (1972).

[40] Rozenberg, G. and Salomaa, A., The Mathematical Theory of *L* Systems, Chapter 4 in *Advances in Information Systems Science*, vol. 6, J. Tou, ed. (Plenum Press, New York,) (1976).

[41] Rozenberg, G. and Salomaa, A., The mathematical theory of *L* systems, *Academic Press*, (1980).

[42] Ruohonen, K., Zeros of Z-rational Functions and DOL-equivalence, *Theor. Comp. Science*, vol. 3, pp. 283-292 (1976).

[43] Ruohonen, K., On the decidability of the OL-DOL equivalence problem, *Inform. Contr.*, vol. 40, pp. 301-318 (1979).

[44] Ruohonen, K., The decidability of the FOL-DOL equivalence problem, *Inform. Process. Lett.*, vol. 8, pp. 257-260 (1979).

[45] Ruohonen, K., The inclusion problem for DOL languages, *Inform. and Control*, to appear.

[46] Ruohonen, K., The decidability of the DOL-DTOL equivalence problem, manuscript, (1979).

[47] Salomaa, A., *Formal Languages*, Academic Press, New York, (1973).

[48] Salomaa, A., DOL equivalence: The problem of iterated and morphisms, *Bulletin of the European Assoc. for Theoretical Computer Science*, No. 4, pp. 5.12 (1978).

[49] Salomaa, A., Equality sets for homomorphisms of free monoids, *Acta Cybernetica*, vol. 4, pp. 127-139 (1978).

[50] Salomaa, A., DOL language equivalence, *Bulletin of the European Assoc. for Theoretical Computer Science*, No. 8, pp. 4-12 (July 1979).

[51] Salomaa, A. and Soittola, M., *Automata-Theoretic Aspects of Formal Power Series*, Springer-Verlag, (1978).

[52] Vitanyi, P.M.B., Growth of strings in context dependent Lindenmayer systems, In: *L Systems*, edited by Rozenberg, G. and Salomaa, A., *Lecture Notes in Computer Science*, vol. 15, Springer-Verlag, pp. 104-123 (1974).

Replacement
type for cover
of Computer Science
Technical Report
CS-80-02.
Rich Beach

Homomorphisms: Decidability, Equality and Test Sets

K. Culik II

CS-80-02

January 1980

Homomorphism compatibility for DOL languages

$\Updownarrow$

Intersection of HDOL sequences

$\swarrow\!\!\!\swarrow$ $\qquad\qquad$ $\searrow\!\!\!\searrow$

Intersection of DOL sequences

Intersection of DOL growth sequences

$\Updownarrow$

Immediate repetition in a DOL growth sequence

$\Updownarrow$

Zero of Z-rational function

$\Downarrow$

HDOL sequence equivalence

$\Updownarrow$

Homomorphism equivalence for DOL languages

Relations between open decidability problems (see 5.5-5.10, $A \Rightarrow B$ reads the decidability of A implies the decidability of B)

HOMOMORPHISM COMPATIBILITY FOR DOL LANGUAGES

$\Updownarrow$

INTERSECTION OF HDOL SEQUENCES

$\swarrow$ $\searrow$

INTERSECTION OF DOL SEQUENCES

INTERSECTION OF DOL GROWTH SEQUENCES

$\Updownarrow$

IMMEDIATE REPETITION IN A DOL GROWTH SEQUENCE

$\Updownarrow$

ZERO OF Z-RATIONAL FUNCTION

$\Downarrow$

HDOL SEQUENCE EQUIVALENCE

$\Updownarrow$

HOMOMORPHISM EQUIVALENCE FOR DOL LANGUAGES

RELATIONS BETWEEN OPEN DECIDABILITY PROBLEMS
(SEE 5.5-5.10, $A \Rightarrow B$ READS THE DECIDABILITY OF A
IMPLIES THE DECIDABILITY OF B)

Open Problem 5.1: What is the minimal integer n such that PCP with only lists of length n is undecidable?

The results in [17] support the following generally accepted but not yet proven

Conjecture 5.2: The PCP with the restriction to lists of length two, i.e. homomorphic compatibility on Σ^+ for a binary Σ, is decidable.

More difficult to prove would be the following:

Conjecture 5.3: The PCP restricted to instances $PCP(g,h)$ with elementary homomorphisms (lists) g and h is decidable.

Open is also the modification of Conjecture 5.3 obtained by assuming that the homomorphisms are injective rather than elementary.

Theorem 5.4: The positive solution of Problem 4.5 implies the validity of Conjectures 5.2 and 5.3.

Proof: If g and h are elementary, then we can effectively find a regular set $E(g,h)$ and test whether $E(g,h) \neq \{\epsilon\}$. This validates Conjecture 5.3, in the case of Conjecture 5.2. There remains the easy case when at least one of g,h is periodic, see [17].

Obviously, $\#\Sigma^*$, for every alphabet Σ and $\# \notin \Sigma$, is a DTOL language. Hence, the homomorphism compatability problem is clearly undecidable for the family of DTOL languages. However, we have the following:

Open Problem 5.5: (Homomorphism compatibility on DOL languages). Given w in Σ^*, and homomorphisms $h : \Sigma^* \to \Sigma^*$, $g,f : \Sigma^* \to \Delta^*$, is it decidable whether there is an $n \geq 0$ such that $g(h^n(w)) = f(h^n(w))$?

Related are the following two problems:

Open Problem 5.6: (Intersecting DOL sequences). Given u,v in Σ^* and homomorphisms $g,h : \Sigma^* \to \Sigma^*$, is it decidable whether $g^n(u) = h^n(v)$ for some $n \geq 0$?

Open Problem 5.7: (Intersecting HDOL sequences). Given u in Σ^*, v in Δ^* and homomorphisms

$$g_1 : \Sigma^* \to \Sigma^*,$$
$$g_2 : \Sigma^* \to \Gamma^*,$$
$$h_1 : \Delta^* \to \Delta^*,$$
$$h_2 : \Delta^* \to \Gamma^*$$

Homomorphisms: Decidability, Equality and Test Sets

K. Culik II

Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

ABSTRACT

A number of recent results and open problems on homomorphisms on free monoids are discussed. Many of the results and conjectures state that various equivalence problems about homomorphisms are decidable. Also discussed are equality sets, test sets and new representation theorems for families of languages.

Presented at the International Symposium on Formal Language Theory, Santa Barbara, California, December 10-14, 1979.

Homomorphisms: Decidability, Equality and Test Sets

K. Culik II

Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

0. Introduction

We survey a number of recent results and open problems on homomorphisms on free monoids. Except for the last section, dealing with representation of language families, most of the results are decidability results. They were motivated or directly constitute problems in L-systems theory. However all of them are basic problems about free monoids and as such are not only of purely mathematical interest but also, since they are all simply formulated decidability problems, are of fundamental interest for theoretical computer science.

Whenever possible we give an algebraic formulation of each problem so that reading, not only the whole paper, but even a particular problem or theorem does not require any specialized knowledge. Open problems are specifically of interest, which makes us stress some topics. The only new results in this paper are some relations among the open problems (conjectures) mostly very easily shown.

In section 2 we deal with iterations of one or more homomorphisms (DOL, HDOL, DTOL systems) and some generalizations thereof. The next section is about "homomorphism equivalence on languages," i.e. the problem whether two given homomorphisms agree "string by string" on a given language, and its applications to transducers.

In Section 4 we consider elementary homomorphisms and questions about equality sets, in particular over a binary alphabet. In the next section we consider "homomorphism compatibility on languages," i.e. the problem whether there exists a string in given languages on which two given homomorphisms agree, in particular various restricted forms of the Post Correspondence Problem.

In Section 6 we discuss the Ehrenfeucht conjecture: Each language possesses a finite subset such that any two homomorphisms which agree (string by string) on the subset agree also on the whole language. Some partial solutions are discussed. Finally in the last section we list some new representation theorems for language families based on equality sets and related phenomena.

Research supported in part by the Natural Science and Engineering Research Council of Canada under Grant No. A-7403. Preparation of this paper was supported in part by the National Science Foundation under Grant No. MCS79-04012.

1. Preliminaries

We consider homomorphisms $\Sigma^* \to \Delta^*$, where Σ^*, Δ^* are free monoids generated by finite alphabets Σ, Δ. The monoid unit (empty word) is denoted by ϵ, the length of a word w in Σ^* by $|w|$. We also use $|n|$ to denote the absolute value of number n. The cardinality of set S is denoted by card S.

An alphabet Σ, homomorphism $h : \Sigma^* \to \Sigma^*$ and an (initial) word w in Σ^* form a *DOL system* $G = (\Sigma, h, w)$. *The sequence generated by* G, denoted $E(G)$, is defined by $E(G) = w, h(w), h^2(w), \ldots$; *the language generated by* G, denoted $L(G)$, *is defined by* $L(G) = \{h^n(w) \mid n \geqslant 0\}$.

A DOL system G and another homomorphism g form an HDOL system $K = \langle G, g \rangle$. It generates the sequence

$$E(K) = g(w), g(h(w)), g(h^2(w)), \ldots$$

and the language

$$L(K) = g(L(G)) = \{g(h^n(w)) \mid n \geqslant 0\}.$$

A DTOL system G is a tuple $(\Sigma, h_1, \ldots, h_n, w)$ where $h_i : \Sigma^* \to \Sigma^*$ for $i = 1, \ldots, n$. It generates a set of sequences

$$\{w, h_{i_1}(w), h_{i_1}(h_{i_2}(w)), \ldots \mid i_1, i_2, \ldots \in \{1, \ldots, n\}\}$$

and the language

$$L(G) = \{h_{i_1}(h_{i_2}(\ldots h_{i_k}(w)\ldots)) \mid i_1, \ldots, i_k \in \{1, \ldots, n\}\}.$$

For homomorphisms $g, h : \Sigma^* \to \Delta^*$ the *equality set for the pair* (g,h) is denoted by $E(g,h)$ and defined by $E(g,h) = \{x \in \Sigma^* \mid g(x) = h(x)\}$.

A *deterministic generalized sequential mapping* (*dgsm mapping*) is a mapping defined by deterministic generalized sequential machine with accepting states (dgsm) as in [32].

For other standard definitions and notations we refer the reader to [40, 41 or 47].

2. Iterated Homomorphisms

We will discuss a number of decision problems about iterative homomorphisms. The following problem and techniques used in its proof stimulated most of the research reported in this paper.

Theorem 2.1 [15] (DOL sequence equivalence problem). Given two homomorphisms $g, h : \Sigma^* \to \Sigma^*$ and w in Σ^* it is decidable whether $g^n(w) = h^n(w)$ for all $n \geqslant 0$.

The strategy of the solution of this problem is to show that any two (normal) equivalent systems must behave in certain "similar" ways and then to show the decidability for similar systems only. Here a pair of DOL systems is similar if the pair (g,h) has "bounded balance" on the language $\{g^n(w) \mid n \geqslant 0\}$

The balance of a string w in Σ^* with respect to a pair of homomorphisms g,h on Σ^* is defined as

$$B(w) = |g(w)| - |h(w)|$$

The pair (g,h) is said to have *bounded balance* on language L if there is a $C > 0$ so that $|B(w)| \leqslant C$ for each prefix of every word in L.

A property of a pair of DOL systems $G_1 = (\Sigma,g,w)$ and $G_2 = (\Sigma,h,w)$ equivalent to "bounded balance" is introduced in [15]. The pair (G_1, G_2) is said to have a *true envelope* R if $L(G_1) \cup L(G_2) \subseteq R \subseteq E(g,h)$. Obviously, if a pair (G_1,G_2) has a true envelope, then G_1 and G_2 are sequence equivalent. It is shown in [15] that a pair of equivalent DOL systems (G_1,G_2) has a regular true envelope iff the pair of homomorphisms (g,h) has bounded balance on $L(G_1)$, and consequently that each pair of equivalent normal DOL systems has a regular true envelope. The latter result is extended in [25] to all pairs of equivalent DOL systems.

The "bounded balance technique" is also useful when testing homomorphism equivalence discussed in Section 3. (See [20]). The same holds also for another technique introduced in [8], the "shifting argument". Roughly speaking, it is used to show that if homomorphisms g,h agree on two words of the form xwy and uwv, i.e. with a common subword w, where w is "sufficiently long" and $|B(x) - B(u)|$ "sufficiently small", then either $B(x) = B(u)$ or $g(w)$ and $h(w)$ are periodic.

The bounded balance technique is not helpful in proving the following generalization of the DOL sequence equivalence problem.

Conjecture 2.2: (HDOL equivalence problem). Given four homomorphisms $g_1 : \Sigma_1^* \to \Sigma_1^*$, $g_2 : \Sigma_1^* \to \Sigma_2^*$, $h_1 : \Delta_1^* \to \Delta_1^*$, $h_2 : \Delta_1^* \to \Delta_2^*$ and strings $u \in \Sigma_1^*$, $v \in \Delta_1^*$. It is decidable whether $g_2(g_1^n(u)) = h_2(h_1^n(v))$.

We show later a problem equivalent to the HDOL equivalence problem (Theorem 3.3). There are two other interesting extensions of DOL equivalence which have been shown decidable by reducing them to DOL equivalence. (Theorems 2.3 and 2.6). The proof of the following theorem also uses results about monoids generated by integer matrices obtained by [34] and by [37].

Theorem 2.3: [10] (Ultimate sequence equivalence). Given two homomorphisms $g,h : \Sigma^* \to \Sigma^*$ and u,v in Σ^* it is decidable whether there exists $n \geqslant 0$ such that $g^k(u) = h^k(v)$ for all $k \geqslant n$.

It is natural to ask whether sequence equivalence remains decidable for more complicated mappings than homomorphisms, in particular for mappings defined symbol by symbol but in a context dependent manner. This is also strongly biologically motivated since such mappings abstract developmental systems of higher level where individual cells interact, i.e. their behaviour is context dependent. The simplest case is dependence on one symbol at the left, the so called D1L system. The sequence equivalence has been shown undecidable even for propagating (nonerasing) version of these systems.

Theorem 2.4: [52] The PD1L sequence equivalence problem is undecidable.

In the view of the last theorem it is rather surprising that the equivalence problem becomes decidable when the rewriting of a letter might depend on one neighbour from each side but only when the letter is being rewritten by at least two new letters. That is any letter-to-letter rewriting must be context free (no erasing is allowed). A deterministic system based on this type of rewriting is introduced in [16] and called an e-GD2L system. Two main results of [16] are that e-GD2L systems have essentially context-free behaviour and that the sequence equivalence for them is decidable. The former result could be compared to "Baker's Theorem" (29, Theorem 10.2.1) giving a condition under which context-sensitive grammar generates a context-free language.

Theorem 2.5: If the sequence $s_0, s_1, \ldots$ is generated by an e-GD2L system, then there exist a nonerasing homomorphism h and a letter-to-letter homomorphism (coding) g so that $s_n = g(h^n(s_0))$ for all $n \geqslant 0$.

Theorem 2.6: [16] The sequence equivalence problem for e-GD2L systems is decidable.

We are not directing our attention here to the languages generated by various parallel rewriting systems, but for completeness of the decidability results we mention the following two theorems. The DOL language equivalence had already been reduced to DOL sequence equivalence in [38] before the latter was shown to be decidable. Recently even the inclusion problem has been shown decidable.

Theorem 2.7: [45] The inclusion problem for DOL languages is decidable.

In the nondeterministic case we have the following result which follows from the undecidability of the equality problem for sentential forms of context free languages.

Theorem 2.8: [3] The equivalence problem for OL (even POL) languages is undecidable.

Another biologically important generalization of DOL systems is obtained when several starting strings and several homomorphisms (tables) are considered. Given two such systems with matching starting strings and matching pairs of homomorphisms we can ask whether all "matching" sequences are identical.

Consider

$$(h_1, \ldots, h_n), (h_1', \ldots, h_n') \tag{2.1}$$

where h_i, h_i' are homomorphisms $\Sigma^* \to \Sigma^*$, for $i = 1, \ldots, n$.

Conjecture 2.9: (DTOL sequence equivalence). Given strings $w, w' \in \Sigma^*$ and homomorphisms (2.1) it is decidable whether

$$h_{i_1}(h_{i_2}(\ldots h_{i_k}(w)\ldots)) = h'_{i_1}(h'_{i_2}(\ldots h'_{i_k}(w')\ldots))$$

for all $i_1 i_2 \ldots i_k$ in $\{1, \ldots, n\}^*$.

Lemma 2.10: [20] Conjecture 2.9 holds if it holds for $n = 2$ (two tables).

Later we show another conjecture equivalent to Conjecture 2.9 (Theorem 3.4).

Note that the DTOL language equivalence problem has been shown undecidable in [39] and recently [46] it has been shown that it becomes decidable if only one system is a DTOL system and the other is DOL. This is a strengthening of the decidability of DOL language equivalence.

All the decidable problems mentioned in this section, as well as some other problems in L-systems (see e.g. [22]) have been shown decidable by reducing them to the DOL sequence equivalence problem (Theorem 2.1). Another problem shown decidable in the same way has been the equivalence problem for simple single loops programs with respect to symbolic evaluation [33].

3. Homomorphism Equivalence on a Language

The problems discussed in this section originated in a simple observation in the proof of decidability of DOL equivalence problem [8,15]. The first step in the proof was that given homomorphisms $g,h:\Sigma^*\to\Sigma^*$ and w in Σ^* the following two conditions are clearly equivalent.

(i) $g^n(w) = h^n(w)$ for all $n \geqslant 0$;
(ii) $g(u) = h(u)$ for all u in L, $L = \{g^n(w) : n \geqslant 0\}$.

So, the testing of iterative equivalence of two homomorphisms g,h can be reduced to the testing of string by string equivalence of g and h on a certain language, namely the language generated by g from the "starting string" w. It is natural and also very useful (cf. Theorems 3.10 and 3.11) to attempt such testing also for other types of languages.

The problem to test whether two homomorphisms agree (string by string) on a given language from family L is called the *homomorphic equivalence problem for L* [20]. Its decidability for regular sets was already implicitly contained in [15]. The following is the main result from [20].

Theorem 3.1: [20] (Homomorphism equivalence for CFL). Given a context free language $L \subseteq \Sigma^*$ and homomorphisms $h,g:\Sigma^*\to\Delta^*$, it is decidable whether $h(x) = g(x)$ for each $x \in L$.

The decidability of homomorphic equivalence is open for all families of languages between DOL and indexed. In particular we have the following:

Conjecture 3.2: (Homomorphism equivalence for DOL languages). Given w in Σ^* and homomorphism $h:\Sigma^*\to\Sigma^*$ and $f,g:\Sigma^*\to\Delta^*$ it is decidable whether

$$f(h^n(w)) = g(h^n(w)) \tag{3.1}$$

for all $n \geqslant 0$.

The following is mentioned in [20].

Theorem 3.3: Conjecture 2.2 is equivalent to Conjecture 3.2, i.e. the HDOL equivalence problem is decidable iff the homomorphism equivalence problem for

DOL languages is decidable.

Proof: 1. To test (3.1) means to compare two HDOL sequences based on the same DOL system. 2. Given $u \in \Sigma^*$, $v \in \Delta^*$ and homomorphisms $g_1 : \Sigma^* \to \Sigma^*$, $g_2 : \Sigma^* \to \Gamma^*$, $h_1 : \Delta^* \to \Delta^*$, $h_2 : \Delta^* \to \Gamma^*$. Assume without loss of generality that $\Sigma \cap \Delta = \phi$ and define homomorphisms f, f_1 and $f_2 : (\Sigma \cup \Delta)^* \to (\Sigma \cup \Delta)^*$ by $f(a) = g_1(a)$ for $a \in \Sigma$, $f(b) = h_1(b)$ for $b \in \Delta$, $f_1(a) = g_2(a)$, $f_2(a) = \epsilon$ for $a \in \Sigma$, $f_1(b) = \epsilon$, $f_2(b) = h_2(b)$ for $b \in \Delta$. Then, clearly, $f_1(f^n(uv)) = f_2(f^n(uv))$ for all n iff $g_2(g_1^n(u)) = h_2(h_1^n(v))$ for all n.

Using similar techniques as in the proof of Theorem 3.2 we also get the following reduction result.

Theorem 3.4: The following three problems are equivalent (and thus all conjectured to be decidable by Conjecture 2.9).

(a) DTOL sequence equivalence problem;
(b) HDTOL sequence equivalence problem;
(c) Homomorphism equivalence problem for DTOL languages.

Proof: We show the reduction (c) to (a); the others are easier.

Let $\overline{\Sigma} = \{\overline{a} \mid a \in \Sigma\}$ and for w in Σ let $\overline{w}$ denote the word obtained from w by "barring" each symbol. Given DTOL system $G = (\Sigma, h_1, h_2, w)$ and homomorphisms g_1, g_2, we construct DTOL systems $G_i = (\Sigma \cup \overline{\Sigma}, h_1', h_2', f_i, w)$ for $i = 1, 2$, where $h_j'(a) = h_j(a)$, $h_j'(\overline{a}) = \epsilon$ for all $a \in \Sigma$ and $j = 1, 2$; $f_i(a) = \overline{g_i(a)}$, $f_i(\overline{a}) = \epsilon$ for all a in Σ and $i = 1, 2$.

Since $h_j'(f_i(u)) = \epsilon$ for all $i, j = 1, 2$ and $u \in \Sigma^*$, it is easy to verify that G_1 and G_2 are sequence equivalent iff homomorphisms g_1 and g_2 are equivalent on $L(G)$.

In [20] it has been conjectured that even a much stronger result than Conjecture 3.2 holds. However, in the view of Theorem 3.3 we cannot expect it to be easy to prove the following:

Conjecture 3.5: The homomorphism equivalence problem for indexed languages is decidable.

For the special case of elementary homomorphisms (see Section 4) decidability has been shown using Theorem 4.4.

Theorem 3.6: [49] It is decidable whether two given elementary homomorphisms are equivalent on a given indexed language.

The following is a partial solution of Conjecture 3.5, which is incomparable with Theorem 3.1. It is based on the fact that every homomorphism on a binary alphabet is either elementary or periodic with the same period for each letter (see Section 4), and on Theorem 3.6.

Theorem 3.7: [19] The homomorphism equivalence problem for ETOL languages over a binary alphabet is decidable.

Finally, we have an easy undecidability result:

Theorem 3.8: [20] The homomorphism equivalence problem for (deterministic) context-sensitive languages is undecidable.

We conclude this section with applications of Theorem 3.1 to problems about finite and push-down transducers [12]. All these quite powerful results follow easily from Theorem 3.1. Note, for example, that the equivalence problem for deterministic generalized sequential machines is a very special case of Theorem 3.11.

We call a transducer defining a regular (rational) translation a finite transducer (a-transducer in [28]). In [1] it has been shown that regular (rational) and push down translations can be homomorphically characterized, i.e. each regular or push-down translation t can be expressed as

$$t = \{(g(w), h(w)) : w \in L\}$$

where g, h are homomorphisms and L is regular or context free, respectively. Therefore, we immediately obtain by Theorem 3.1:

Theorem 3.9: [12] Given a finite transducer or a push down transducer it is decidable whether it defines an identity relation restricted to its domain.

From Theorem 3.9 we easily obtain the following:

Theorem 3.10: [12] Given a finite transducer M and a context-free grammar G, it is decidable whether t_M (the relation defined by M) is functional on $L(G)$.

The inverse relation of the restriction of t_M to $L(G)$ is not necessarily equal to the restriction of t_M^{-1} to $t_M(L(G))$. Hence it does not follow as a corollary of Theorem 3.10, as claimed in [12], that it is decidable whether t_M is one-to-one on $L(G)$. Actually this problem has been shown to be undecidable in [30]. However, we can test whether t_M is one-to-one (on its domain).

Among the other consequences of Theorem 3.1 shown in [12] is the decidability of the equivalence problem for functional finite transducers, or the even stronger result which follows, where an unambiguous pushdown transducer is a p.d.t. based on an unambiguous pushdown automaton [32].

Theorem 3.11: [12] (Equivalence between a functional finite transducer and an unambiguous pushdown transducer). Given an unambiguous pushdown transducer P and a functional finite transducer it is decidable whether $t_P = t_M$.

4. Elementary Homomorphisms and Equality Sets

Here we consider a very useful special type of homomorphism first introduced in [23], equality sets for them and equality sets over a binary alphabet.

A homomorphism $h : \Sigma^* \to \Delta^*$ is *elementary* if there is no decomposition of h into homomorphisms f and g, that is $h = gf$:

[redacted]

such that card $\Gamma <$ card Σ. Similarly, a finite language L is *elementary* if there is no language K such that card $K <$ card L and $L \subseteq K^*$. Let $L_h = \{h(a) : a \in \Sigma\}$. Clearly, a homomorphism $h : \Sigma^* \to \Delta^*$ is elementary iff card L_h = card Σ and L_h is elementary.

Properties of elementary homomorphisms and languages were studied in [23, 24, 41 and 35]. We mention a few of them.

Theorem 4.1: [23] Each elementary homomorphism is injective.

Theorem 4.2: [41] Let $L = \{u_1, ..., u_n\}$ be an elementary language over the alphabet Σ. If $u_i xz = u_j y$ for some $i \neq j$, $x, y \in L^*$ and $z \in \Sigma^*$, then $|u_i x| \leqslant |u_1 u_2 \ldots u_n| - n$.

Corollary 4.3: [41] Every elementary language is a code with bounded delay (both from left to right and from right to left).

The following is an important result. In particular it has made it possible to simplify the proof of the decidability of DOL sequence equivalence.

Theorem 4.4: [24] If homomorphisms g and h are elementary, then the equality set $E(g,h)$ is regular.

This result has been strengthened in [26] for the weaker assumption that at least one of g and h is elementary and then for even weaker assumptions in [35]. No effective proof even for the weakest result is known so we have the following open problem and its even harder versions.

Open Problem 4.5: Given elementary homomorphisms $g, h : \Sigma^* \to \Delta^*$, can the regular set $E(g,h)$ (represented e.g. by a regular expression) be found effectively?

This problem is presently open even for the case of binary alphabets [17]. A positive answer in this special case already implies the validity of Conjecture 5.2.

the decidability of PCP restricted to lists of length two.

The fact that every homomorphism on a binary alphabet is either elementary or periodic severely restricts the form of equality sets for homomorphisms over a binary alphabet (on free monoids with two generators). Each equality set is either regular or of the form $\{w \in \{a,b\}^* \mid \#a(w)/\#b(w) = k\}$ for some rational $k \neq 0$, where $\#a(w)$ is the number of occurrences of letter a in w. In [17] an attempt has been made to fully classify such equality sets. In particular for some words all possible homomorphisms agreeing on them are shown. On the other hand a number of sets or words (singleton sets) are shown to be "periodicity forcing," meaning that only periodic homomorphisms could agree on them. These results support the following:

Conjecture 4.6: Every regular equality set for homomorphisms over a binary alphabet is of the form F^* where F is of cardinality at most two.

This conjecture would imply a simple proof of Theorem 6.2 and also sharpen this theorem, namely it would imply that for $L \subseteq \{a,b\}^*$ there always exists (noneffectively) a test set (see Section 6) of cardinality at most three. Some other implications of Conjecture 4.6 are discussed in [17].

Note that there is no loss of generality in assuming that the range of considered homomorphisms is over a binary alphabet, since a larger alphabet can always be encoded into a binary one, preserving the equality set. This is, of course, not the case for the domain. Hence, we have a rather unusual situation that many problems considered here are much easier for a binary alphabet than in the general case. One such example is the DOL equivalence problem.

5. Homomorphism Compatibility

In Section 3 we were interested in testing whether two given homomorphisms agree "string by string" on a given language. In [20] four kinds of "homomorphism agreements" were considered, namely, compatibility, strong compatibility, ultimate equivalence and equivalence. The last one was considered in Section 3, here we will consider the first one, the other two are omitted since the results for them are similar to the two cases considered.

Homomorphisms g and h are *compatible* on a language L if $g(w) = h(w)$ for some w in L, that is if $L \cap E(g,h) \neq \phi$.

To decide whether homomorphisms g and h are compatible on Σ^+, i.e. whether $E(g,h)-\{\epsilon\} \neq \phi$, is nothing else but the Post Correspondence Problem (PCP). An instance PCP(g,h) is given by two nonerasing homomorphisms $g,h : \{1,\ldots,n\}^* \to \Sigma^*$, traditionally called *lists* of length n.

The problem of homomorphism compatibility for a family of languages $\mathbf{L}$ can be stated as: given L in $\mathbf{L}$ and homomorphisms g and h, to decide whether there is w in L such that $g(w) = h(w)$. Hence, this problem is undecidable for any family containing Σ^+ for arbitrarily large alphabet Σ. Actually, it is known that there is a certain fixed size for Σ, which is sufficient to make the PCP undecidable. However the minimal size is not known. We have the following open problem and conjecture.

is it decidable whether $g_2(g_1^n(u)) = h_2(h_1^n(v))$ for some n?

Theorem 5.8: Problems 5.5 and 5.7 are equivalent, the decidability of Problem 5.5 implies the decidability of Problem 5.6.

Proof: Similar, to the proof of Theorem 3.3.

If we modify Problem 5.6 so that only the length of generated strings is compared we obtain the following problem which is shown in [51] to be equivalent to the well known open problem of finding zeros of Z-rational functions.

Open Problem 5.9: (Intersecting DOL growth sequences). Given u in Σ^*, v in Δ^* and $h : \Delta^* \to \Delta^*$ decide whether $|g^n(u)| = |h^n(v)|$ for some n.

There are a large number of results (and open problems) concerning growth (length) and Parikh vector sequences generated by one or more iterative homomorphisms. Mathematically, they belong to the theory of noncommutative formal power series and we refer the interested reader to [51]. We have included Problem 5.9 here because of its strong implications to our other open problems shown in the following theorem. The first part was shown in the terminology of Z-rational functions in [42] the second follows by Theorem 3.3.

Theorem 5.10: The decidability of Problem 5.9 implies

(i) The decidability of the HDOL sequence equivalence problem (Conjecture 2.2).

(ii) The decidability of the homomorphism equivalence problem for DOL languages.

Clearly, the decidability of Problem 5.7 (or 5.5) implies the decidability of Problem 5.9 and therefore, by Theorem 5.10, also the decidability of (i) and (ii) above.

In [17] a problem which can be considered dual to the Post Correspondence Problem is shown to be decidable by reducing it to Makanin's result concerning solvability of equations in free monoids [36].

Theorem 5.11: Given a string w in Σ^*, it is decidable whether there exist two distinct homomorphisms $g,h : \Sigma^* \to \Delta^*$ for some Δ, such that at least one of them is aperiodic and $g(w) = h(w)$.

Note that the problem is trivial if g and h are not required to be distinct or aperiodic.

6. Test Sets and Checking Words

The very interesting "Ehrenfeucht conjecture" (Conjecture 6.1) is at least several years older than the notion of homomorphism equivalence to which it is closely related.

We say that a finite subset F of a language L is a *test set for L* if, for any pair of homomorphisms (g,h), $g(x) = h(x)$ for all x in L if and only if $g(x) = h(x)$ for all x in F, i.e. g and h are equivalent on L iff g and h are equivalent on F.

Conjecture 6.1: For every language there exists a test set.

It immediately follows by Theorem 3.8 that given a context sensitive grammar G a test set for $L(G)$ cannot be effectively constructed, since the effective existence of a test set for a family $\mathbf{L}$ obviously implies the decidability of homomorphism equivalence for $\mathbf{L}$.

The discussion in Section 4 indicates that proving the validity of the Ehrenfeucht conjecture is considerably easier for languages over a binary alphabet. This has actually has been done recently in [21].

Theorem 6.2: [21] For each language $L \subseteq \Sigma^*$, where Σ is a binary alphabet, there exists a test set F, i.e. a finite set F, such that for each pair of homomorphisms g,h, $g(x) = h(x)$ for all x in L iff $g(x) = h(x)$ for all x in F.

It easily follows from the discussion in (15 or 20) that for each regular set there effectively exists a test set. Recently this result has been extended to context free languages.

Theorem 6.3: [2] For each context free language there effectively exists a test set.

As mentioned above, this result immediately implies the decidability of homomorphism equivalence on the CFL (Theorem 3.1).

Actually, a somewhat stronger form of Theorem 6.3 is shown in [2], namely that given a CFG $G = (N,T,P,S)$ with n=card N and m the maximal length of the right side of the a production in P,

$$F = \{w \in L : |w| \leqslant m^{3n+1}\}$$

is a test set for $L(G)$.

This result is then used to obtain also finite "test sets" for CFL with respect to gsm mappings realized by gsm with a uniformly bounded number of states. Despite our reasons for expecting it to be hard to prove Conjecture 3.2 (equivalent to Conjecture 2.2) we venture to make an even stronger one:

Conjecture 6.4: For every indexed language (given by an indexed grammar) there effectively exists a test set.

According to [21] a word in Σ^* is a *checking word* for a language $L \subseteq \Sigma^*$ if, for any pair of homomorphisms (g,h), $g(x) = h(x)$ for all x in L if and only if $g(w) = h(w)$. Observe that it is not required that w be in L, hence $\{w\}$ might not be a test set for L.

A language L is *rich* if two homomorphisms g and h are equivalent on L only in case $g = h$. Somewhat surprisingly it is easier to show the following result than Theorem 6.2.

Theorem 6.5: [21] Every language over a binary alphabet is either rich or possesses a checking word.

7. Representation of Language Families

Equality sets have already been discussed in the previous sections mainly in relation to decidability problems about homomorphisms. They were explicitly introduced in [24] under a different name and their basic properties were studied in [49]. A generalization of equality sets to more than two homomorphisms is considered by [6].

It turned out that equality sets and similarly fixed point languages [26] provide simple representations of the recursively enumerable languages [11, 27, 49] which can also be extended to various time and complexity classes [4, 5, 13, 14]. The first results say that the closure of equality sets under dgsm mappings is the family of recursively enumerable languages.

Theorem 7.1: [27, 49] For every recursively enumerable (r.e.) set L, there exists a pair of homomorphisms (h_1, h_2) and a dgsm mapping g such that $L = g(E(h_1, h_2))$.

This result has been strengthened in [11] by replacing dgsm's by erasings and equality sets by minimal equality sets.

A homomorphism $h : \Sigma^* \to \Delta^*$ is called an *erasing* if, for each $a \in \Sigma$ either $h(a) = a$ or $h(a) = \epsilon$. For homomorphisms $g, h : \Sigma^* \to \Delta^*$, *the minimal equality set* is the set $e(g, h) = \{w \in \Sigma^+ \mid g(w) = h(w)$ and if $w = uv$ for $u, v \in \Sigma^+$, then $g(u) \neq h(u)\}$, that is, using the notation of [32],

$$e(g, h) = \min(E(g, h)) - \{\epsilon\}.$$

Theorem 7.2: [11] For every r.e. language L, there exist homomorphisms h_1, h_2 and an erasing h_0 so that $L = h_0(e(h_1, h_2))$.

Also we have a representation based on fixed points of dgsm mappings. Let $g : \Sigma^* \to \Sigma^*$ be a function. The *fixed-point language $F_p(g)$ of the function g* is defined to be $F_p(g) = \{w \in \Sigma^* \mid g(w) = w\}$.

Theorem 7.3: [27] For every r.e. language L, there exists a dgsm mapping g and an erasing h such that $L = h(F_p(g))$.

By imposing simple restrictions on the mappings used in Theorems 7.1, 7.2 and 7.3 we obtain a simple "machine independent" characterization of many time and space complexity classes of languages. Similar results have been shown independently in [4, 5] and in [13, 14]. We give here a few of these results.

We say that $\mathbf{C}$ is a class of complexity functions if $\mathbf{C}$ is a class of functions closed under addition of and multiplication by a constant. A language L is of time (space) complexity $\mathbf{C}$ if L is accepted by a nondeterministic multitape on-line Turing machine M which operates within time-bound (space-bound) f, for some f in $\mathbf{C}$, we write $L \in \mathrm{NTIME}(\mathbf{C})$ $(L \in \mathrm{NSPACE}(\mathbf{C}))$.

We generalize the notion of k-limited erasing [32] as follows: For a function f on the integers we say that *an erasing h is f-bounded on a language L*, if for each w in L, $w = xyz$ and $h(y) = \epsilon$ implies $|y| \leqslant f(|w|)$, that is at most $f(|w|)$ consecutive symbols of w may be erased. We say that h is $\mathbf{C}$-bounded, for a class $\mathbf{C}$ of complexity functions, if h is f-bounded for some f from $\mathbf{C}$.

We get the following "machine independent" characterization of the time complexity classes of languages.

Theorem 7.4: [5,14] Let $\mathbf{C}$ be a class of complexity functions closed under squaring. Then the following three conditions are equivalent.

(i) $L \in \mathrm{NTIME}(\mathbf{C})$

(ii) $L = h_0(e(h_1, h_2))$ where h_1, h_2 are homomorphisms and h_0 is a $\mathbf{C}$-bounded erasing on $e(h_1, h_2)$.

(iii) $L = h(F_p(g))$ where g is a *dgsm* mapping and h is an erasing on $F_p(g)$.

Corollary: A language L is in NP iff there exist homomorphisms h_0, h_1, h_2 such that $h_0(e(h_1, h_2)) = L$ and h_0 is polynomial-bounded erasing on $e(h_1, h_2)$.

Corollary: A language L is primitive recursive (recursive) iff there exist homomorphisms h_0, h_1, h_2 such that $h_0(e(h_1, h_2)) = L$ and h_0 is primitive recursive- (recursive-) bounded erasing on $e(h_1, h_2)$.

In order to get a characterization of space complexity classes we need to generalize the notion of bounded balance on a language considered in Section 3.

Consider two fixed homomorphisms $g,h : \Sigma^* \to \Delta^*$. Recall that for each w in Σ^*, the balance of w is defined as $B(w) = |g(w)| - |h(w)|$. Now, for a monotone function f on the integers, an erasing π and $L \subseteq \Sigma^*$ we say that the *pair (g,h) has f-bounded balance on language L, with respect to erasing π*, if for each x in L and each prefix w of x we have $|B(w)| \leqslant f(|h(x)|)$. For a complexity class $\mathbf{C}$, we say that *(g,h) has $\mathbf{C}$-bounded balance on L, with respect to π*, if the same holds true for some $f \in \mathbf{C}$.

Theorem 7.5: [14] Let $\mathbf{C}$ be any class of complexity functions. Then $L \in \mathrm{NSPACE}(\mathbf{C})$ iff $L = h_0(e(h_1,h_2))$ where h_0 is an erasing and the pair (h_1,h_2) has $\mathbf{C}$-bounded balance on $e(h_1,h_2)$ with respect to h_0.

Corollary: A language L is context sensitive iff there exist an erasing h_0 and homomorphisms h_1, h_2 such that $L = h_0(e(h_1, h_2))$ and the pair (h_1, h_2) has linear-bounded balance on $e(h_1, h_2)$.

The class NP also has an alternative characterization:

Theorem 7.6: [5] The class NP is the smallest class containing all equality sets $E(h_1, h_2)$ (of nonerasing homomorphisms) with square-root-bounded balance (i.e. with the pair (h_1, h_2) having square-root-bounded balance on $E(h_1, h_2)$ with respect to the identity) and closed under intersection with regular sets and polynomial-bounded erasings.

We also have an alternative necessary condition for PSPACE:

Theorem 7.7: [5] For every language L in PSPACE there is a pair of nonerasing homomorphisms (h_1, h_2) with log n-bounded balance, a regular set R, and an erasing h such that $L = h(E(h_1, h_2) \cap R)$ and for some constants $c > 1$, $k > 0$, h is c^{n^k}-bounded on $E(h_1, h_2) \cap R$.

The study of equality sets has contributed representation theorems of the following form:

Let $\mathbf{L}$ be a family of languages over an alphabet Σ. Then there exist a language L_Σ and an erasing π_Σ such that $L \in \mathbf{L}$ iff $L = \pi_\Sigma(L_\Sigma \cap R_L)$ for some regular set R_L.

Letting $\mathbf{L}$ being the family of context free languages we have the well known Chomsky-Schützenberger theorem with L_Σ being the Dyck language over Σ; for the families of EOL and ETOL languages such representation has been established in [7].

Now, given Σ let $\overline{\Sigma}$ denote the alphabet disjoint from Σ consisting of "barred" symbols, $\overline{\Sigma} = \{\overline{a} \mid a \in \Sigma\}$, and for any word x in Σ^*, let $\overline{x}$ denote the word obtained from x by barring each symbol. To get a representation as above for the r.e. sets, the *twin-shuffle* over Σ has been defined in [27] as

$$L(\Sigma) = \{x \in (\Sigma \cup \overline{\Sigma})^* \mid \overline{\pi_\Sigma(x)} = \pi_{\overline{\Sigma}}(x)\}$$

where $\pi_\Sigma, \pi_{\overline{\Sigma}}$ are erasings on $(\Sigma \cup \overline{\Sigma})^*$, which only preserve the symbols from Σ, $\overline{\Sigma}$ respectively.

Clearly, the twin-shuffle $L(\Sigma)$ is an equality set of two homomorphisms. They cannot be nonerasing as shown in [5].

Theorem 7.8: [27] Let L be an r.e. language over Σ. There exists a regular set $R_L \subseteq (\Sigma \cup \overline{\Sigma} \cup \{0, \overline{0}, 1, \overline{1}, \}^*$ such that $L = \pi_\Sigma(L(\Sigma \cup \{0, 1\}) \cap R_L)$, where π_Σ is the erasing which only preserves the symbols from Σ.

This theorem has essentially been shown in [27] using Theorem 7.3. It also follows by Theorem 7.2, see [13, 18].

It is actually not difficult to show that every principal cone has a representation as above. A family of languages $\mathbf{L}$ is a *principal cone* if there is an L in $\mathbf{L}$, such that $\mathbf{L}$ is the closure of $\{L\}$ under the operations of homomorphism, inverse homomorphism and intersection with a regular set, or equivalently $\mathbf{L}$ is the closure of $\{L\}$ under finite transducers (rational relations), see [28].

Theorem 7.9: [13,18] Let Σ be an alphabet and $\mathbf{L}$ a principal cone. There exists a language L_Σ in $\mathbf{L}$ such that for each L in $\mathbf{L}$, $L \subseteq \Sigma^*$, there exists a regular set R_L such that $L = \pi_\Sigma(L_\Sigma \cap R_L)$ where π_Σ is the erasing, which only preserves the symbols in Σ.

We refer to [4,5,13,14,18,27] for a number of additional representation results.

REFERENCES

[1] Aho, A.V. and Ullman, J.D., *The Theory of Parsing, Translation, and Compiling*, Prentice-Hall (1972).

[2] Albert, J. and Culik II, K., Test sets for homomorphism equivalence on context-free languages, University of Waterloo, Research Report CS-79-39.

[3] Blattner, M. The unsolvability of the equality problem for sentential forms of context-free languages, *J. Comp. System Sci.*, vol. 7, pp. 463-468 (1973).

[4] Book, R. and Brandenburg, F.-J., Representing complexity classes by equality sets, Proceedings of the Sixth ICALP, Graz, *Lecture Notes in Computer Science*, vol. 71, Springer Verlag, pp. 49-57 (1979).

[5] Book, R. and Brandenburg, F.-J., Equality sets and complexity classes, *SIAM J. Computing*, to appear.

[6] Brandenburg, F.-J., Multiple equality sets and Post machines, Math. Tech. Rep., University of California, Santa Barbara, (1979).

[7] Culik II, K., On some families of languages related to development systems, *International J. of Computer Math.*, Section A, 4, pp. 31-42 (1974).

[8] Culik II, K., On the decidability of the sequence equivalence problem for DOL-systems, *Theor. Computer Science* vol. 3, pp. 75-84 (1977).

[9] Culik II, K., The decidability of v-local catenativity and of other problems of DOL systems, *Inf. Process. Lett.* vol. 7 (1): pp. 33-35 (1978).

[10] Culik II, K., The ultimate equivalence problem for DOL systems, *Acta Informatica* vol. 10, pp. 69-84 (1978).

[11] Culik II, K., A purely homomorphic characterization of recursively enumerable sets, *JACM* vol. 6, pp. 345-350 (1979).

[12] Culik II, K., Some decidability results about regular and push down translations, *Information Processing Letters* vol. 8, pp. 5-8 (1979).

[13] Culik II, K., On homomorphic characterization of families of languages, Proceedings of the Sixth ICALP, Graz, *Lecture Notes in Computer Science 71*, Springer Verlag, pp. 161-170 (1979).

[14] Culik II, K. and Diamond, N.D., A homomorphic characterization of time and space complexity classes of languages, *Int. J. of Computer Math.*, to appear.

[15] Culik II, K. and Fris, J., The decidability of the equivalence problem for DOL-systems, *Inf. and Control*, vol. 35, pp. 20-39 (1977).

[16] Culik II, K. and Karhumaki, J., Interactive L-systems with almost interactionless behaviour, *Inf. and Control*, vol. 43, pp. 83-100 (1979).

[17] Culik II, K. and Karhumaki, J., On the equality sets for homomorphisms on free monoids with two generators, *R.A.I.R.O.*, to appear, also University of Waterloo Research Report CS-79-17, (1979).

[18] Culik II, K. and Maurer, H.A., On simple representations of languages families, *R.A.I.R.O.*, to appear, also University of Waterloo Research Report CS-78-41, (1978).

[19] Culik II, K. and Richier, J.L., Homomorphism equivalence on ETOL languages, *Int. J. of Computer Math.*, Section A, vol. 7, pp. 43-51 (1979).

[20] Culik II, K. and Salomaa, A., On the decidability of homomorphism equivalence for languages, *J. Comp. System Sci.*, vol. 17, pp. 613-175 (1978).

[21] Culik II, K. and Salomaa, A., Test sets and checking words for homomorphism equivalence, submitted to *J. Comp. System Sci.*, also University of Waterloo Research Report CS-79-04, (1979).

[22] Culik II, K. and Wood, D., Double deterministic tabled OL systems, *Int. J. of Computer and Inf. Science* vol. 8, pp. 335-347 (1979).

[23] Ehrenfeucht, A. and Rozenberg, G., Simplifications of homomorphisms, *Inf. and Control*, vol. 38 pp. 298-308 (1978).

[24] Ehrenfeucht, A. and Rozenberg, G., Elementary homomorphisms and a solution to the DOL sequence equivalence problem, *Theoretical Computer Science* vol. 7, pp. 169-183 (1978).

[25] Ehrenfeucht, A. and Rozenberg, G., Every two equivalent DOL systems have a regular true envelope, *Theoret. Comp. Science*, to appear.

[26] Engelfriet, J. and Rozenberg, G., Equality languages and fixed point languages, *Inf. and Control*, vol. 43, pp. 20-49 (1979).

[27] Engelfriet, J. and Rozenberg, G., Fixed point languages, equality languages and representation of recursively enumerable alanguages, *JACM*, to appear.

[28] Ginsburg, S., *Algebraic and Automata-Theoretic Properties of Formal Languages*, North Holland/American Elsevier, (1975).

[29] Harrison, M.A., *Introduction to Formal Language Theory*, Addison-Wesley, (1978).

[30] Head, T., Unique decipherability relative to a language, Manuscript (1979).

[31] Herman, G.T. and Rozenberg, G., *Developmental Systems and Languages*, (North-Holland, New York.) (1975).

[32] Hopcroft, J.E. and Ullman, J.D., *Formal Languages and Their Relation to Automata*, Addison-Wesley, (1969).

[33] Howden, W.E., Lindenmayer grammars and symbolic testing, *Inf. Processing Letters*, vol. 7, pp. 36-45 (1978).

[34] Jacob, G., Un algorithme calculant le cardinal, fini ou infini, des demi groups de matrices, *Theoret. Comp. Science*, vol. 5, pp. 183-204 (1977).

[35] Karhumaki, J. and Simon, I., A note on elementary homomorphisms and the regularity of equality sets, *Bulletin of the EATCS*, no. 9, pp. 16-24 (Oct. 1979).

[36] Makanin, G.S., The problem of solvabiilty of equations in a free semigroup (in Russian), *Matematiceskij Sbornik* vol. 103 (145), pp. 148-236 (1977).

[37] Mandel, A. and Simon, I., On finite semigroups of matrices, *Theoretical Computer Science*, vol. 5, pp. 101-111 (1977).

[38] Nielsen, M., On the decidability of some equivalence problems for DOL-systems, *Inform. Control*, vol. 25, pp. 166-193 (1974).

[39] Rozenberg, G., The equivalence problem for deterministic TOL systems is indecidable, *Inf. Process. Lett.* 1:201-204, errata, 252 (1972).

[40] Rozenberg, G. and Salomaa, A., The Mathematical Theory of *L* Systems, Chapter 4 in *Advances in Information Systems Science*, vol. 6, J. Tou, ed. (Plenum Press, New York.) (1976).

[41] Rozenberg, G. and Salomaa, A., The mathematical theory of *L* systems, *Academic Press*, (1980).

[42] Ruohonen, K., Zeros of Z-rational Functions and DOL-equivalence, *Theor. Comp. Science*, vol. 3, pp. 283-292 (1976).

[43] Ruohonen, K., On the decidability of the OL-DOL equivalence problem, *Inform. Contr.*, vol. 40, pp. 301-318 (1979).

[44] Ruohonen, K., The decidability of the FOL-DOL equivalence problem, *Inform. Process. Lett.*, vol. 8, pp. 257-260 (1979).

[45] Ruohonen, K., The inclusion problem for DOL languages, *Inform. and Control*, to appear.

[46] Ruohonen, K., The decidability of the DOL-DTOL equivalence problem, manuscript, (1979).

[47] Salomaa, A., *Formal Languages*, Academic Press, New York, (1973).

[48] Salomaa, A., DOL equivalence: The problem of iterated and morphisms, *Bulletin of the European Assoc. for Theoretical Computer Science*, No. 4, pp. 5.12 (1978).

[49] Salomaa, A., Equality sets for homomorphisms of free monoids, *Acta Cybernetica*, vol. 4, pp. 127-139 (1978).

[50] Salomaa, A., DOL language equivalence, *Bulletin of the European Assoc. for Theoretical Computer Science*, No. 8, pp. 4-12 (July 1979).

[51] Salomaa, A. and Soittola, M., *Automata-Theoretic Aspects of Formal Power Series*, Springer-Verlag, (1978).

[52] Vitanyi, P.M.B., Growth of strings in context dependent Lindenmayer systems, In: *L Systems*, edited by Rozenberg, G. and Salomaa, A., *Lecture Notes in Computer Science*, vol. 15, Springer-Verlag, pp. 104-123 (1974).

[17.] Culik II, K. and Karhumaki, J., On the equality sets for homomorphisms on free monoids with two generators, *R.A.I.R.O.*, to appear, also University of Waterloo Research Report CS-79-17, (1979).

[18.] Culik II, K. and Maurer, H.A., On simple representations of languages families, *R.A.I.R.O.*, to appear, also University of Waterloo Research Report CS-78-41, (1978).

Culik II, K. and Richier, J.L., Homomorphism equivalence on ETOL languages, *Int. J. of Computer Math.*, Section A, vol. 7, pp. 43-51 (1979).

[20.] Culik II, K. and Salomaa, A., On the decidability of homomorphism equivalence for languages, *J. Comp. System Sci.*, vol. 17, pp. 613-175 (1978).

[21.] Culik II, K. and Salomaa, A., Test sets and checking words for homomorphism equivalence, submitted to *J. Comp. System Sci.*, also University of Waterloo Research Report CS-79-04, (1979).

[22.] Culik II, K. and Wood, D., Double deterministic tabled OL systems, *Int. J. of Computer and Inf. Science* vol. 8, pp. 335-347 (1979).

[23.] Ehrenfeucht, A. and Rozenberg, G., Simplifications of homomorphisms, *Inf. and Control*, vol. 38 pp. 298-308 (1978).

[24.] Ehrenfeucht, A. and Rozenberg, G., Elementary homomorphisms and a solution to the DOL sequence equivalence problem, *Theoretical Computer Science* vol. 7, pp. 169-183 (1978).

[25.] Ehrenfeucht, A. and Rozenberg, G., Every two equivalent DOL systems have a regular true envelope, *Theoret. Comp. Science*, to appear.

[26.] Engelfriet, J. and Rozenberg, G., Equality languages and fixed point languages, *Inf. and Control*, vol. 43, pp. 20-49 (1979).

[27.] Engelfriet, J. and Rozenberg, G., Fixed point languages, equality languages and representation of recursively enumerable alanguages, *JACM*, to appear.

[28.] Ginsburg, S., *Algebraic and Automata-Theoretic Properties of Formal Languages*, North Holland/American Elsevier, (1975).

[29.] Harrison, M.A., *Introduction to Formal Language Theory*, Addison-Wesley, (1978).

[30.] Head, T., Unique decipherability relative to a language, Manuscript (1979).

[31.] Herman, G.T. and Rozenberg, G., *Developmental Systems and Languages*, (North-Holland, New York,) (1975).

[32.] Hopcroft, J.E. and Ullman, J.D., *Formal Languages and Their Relation to Automata*, Addison-Wesley, (1969).

[33.] Howden, W.E., Lindenmayer grammars and symbolic testing, *Inf. Processing Letters*, vol. 7, pp. 36-45 (1978).

[34.] Jacob, G., Un algorithme calculant le cardinal, fini ou infini, des demi groups de matrices, *Theoret. Comp. Science*, vol. 5, pp. 183-204 (1977).

[35.] Karhumaki, J. and Simon, I., A note on elementary homomorphisms and the regularity of equality sets, *Bulletin of the EATCS*, no. 9, pp. 16-24 (Oct. 1979).

1. please [illegible] PH [illegible] diagr[illegible]
[illegible] of text [illegible]
2. [illegible] [illegible] [illegible]
page 11.
→ cut [illegible] [illegible] for page 11
and paste [illegible] [illegible]
pages 12-[illegible] [illegible]
the [illegible] [illegible] [illegible]
[illegible] will [illegible] to make a [illegible]
[illegible] page 11.
[illegible] [illegible]