Homomorphism Equivalence on ETOL Languages

K. Culik II
Department of Computer Science
University of Waterloo

and

J. L. Richier
I.R.I.A., Rocquencourt, France

Research Report CS-78-22
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
April 1978

Homomorphism Equivalence on ETOL Languages*

K. Culik II
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada

and

J. L. Richier+
I.R.I.A., Rocquencourt, France

Abstract

The following problems are shown to be decidable. Given an ETOL language L and two homomorphisms h_1, h_2 is the length (Parikh vector) of $h_1(x)$ and $h_2(x)$ equal for each x in L ? If L is over a binary alphabet then we can also test whether $h_1(x) = h_2(x)$ for each x in L .

* This research was supported by the National Research Council of Canada Grant No. A7403.

+ This paper was written during the second author's stay at the University of Waterloo, Waterloo, Ontario, Canada.

0. INTRODUCTION

The basic technique used in the recent solution of the DOL (ultimate) equivalence problem [2, 3, 6] was to check equality of two homomorphisms on all strings from a certain language. This decision problem of "string by string" equivalence of two homomorphisms on a given language turned to be of interest of its own. In [7] this problem was shown decidable for CFL languages and arbitrary homomorphisms. Applications of this result to decision problems about rational and push-down translations were shown in [5]. The sets of all strings of which given two homomorphisms agree so called equality set were studied in [12] and [4]. In [7] it was conjectured that the problem of homomorphic equivalence is decidable for indexed languages. This remains an open problem, in this paper we show a partial result in this direction namely that the problem is decidable for ETOL languages (a subset of indexed languages) over a binary alphabet. We also introduce the natural notions of length and Parikh equivalence of two homomorphisms on a language. We show that length (Parikh) equivalence of two given homomorphisms on a given (arbitrary) ETOL language is decidable.

1. PRELIMINARIES

We assume that the reader is familiar with fundamental theory of formal languages including the basics of L systems, c.f. [11]. For convenience, some definitions are given here.

An ETOL system is a fourtuple $G = (V, T, P, \sigma)$ where

(i) V is a finite set, the alphabet.

(ii) $T \subseteq V$, the terminal alphabet.

(iii) P is a finite set of tables, $P = \{P_1, \ldots, P_n\}$ for $n \geq 1$, where each $P_i \subseteq V \times V^*$. Element (u, v) of P_i, $1 \leq i \leq n$, is called a production and usually written $u \to v$. Every P_i, $1 \leq i \leq n$, satisfies the following condition of completeness: For each $a \in V$ there is $w \in V^*$ so that $(a, w) \in P_i$.

(iv) $\sigma \in V^+$, the axiom.

We write $x \overset{G}{\Rightarrow} y$ if there are $k \geq 0$, $a_1, \ldots, a_k \in V$ and $y_1, \ldots, y_k \in V^*$ so that $x = a_1 \ldots a_k$, $y = y_1 \ldots y_k$ and for some $P_i \in P$, $a_j \to y_j \in P_i$ for $j = 1, \ldots, k$. The transitive and reflexive closure of $\overset{G}{\Rightarrow}$ is denoted by $\overset{G}{\Rightarrow}{}^*$. The language generated by G is denoted $L(G)$ and defined as $L(G) = \{w \in T^* : \sigma \overset{G}{\Rightarrow}{}^* w\}$.

An ETOL system is called a TOL system if $V = T$. Moreover, if for each $i = 1, \ldots, n$ and each $a \in T$ there is exactly one w in T^* so that $(a,w) \in P_i$, then the system is called a DTOL system.

Note that in the case of a DTOL system each table P_i defines a homomorphism h_i. Therefore it is sometimes convenient to write a

DTOL system in the form $G = (T, h_1, \ldots, h_m, \sigma)$.

Let $\Sigma = \{a_1, \ldots, a_n\}$. For x in Σ^* the length of x is denoted by $|x|$, the Parikh vector of x by $[x]$. For the empty string ε , $|\varepsilon| = 0$, $[\varepsilon] = (0, \ldots, 0)$.

Now, we turn to the central notions of this paper, some of them are new.

Definition Given a language $L \subseteq \Sigma^*$ and two homomorphisms $h_1, h_2 : \Sigma^* \to \Delta^*$, we say that h_1 and h_2 are equivalent on L , written $h_1 \overset{L}{\equiv} h_2$, iff $h_1(x) = h_2(x)$ for all x in L . We say that h_1 and h_2 are length (Parikh) equivalent, written $h_1 \equiv_\ell h_2$ ($h_1 \equiv_p h_2$) iff $|h_1(x)| = |h_2(x)|$ ($[h_1(x)] = [h_2(x)]$) for all x in L .

Given (effectively) a family of languages $\mathcal{L}$ and a family of homomorphisms $\mathcal{H}$, the problem of $\mathcal{H}$ (length, Parikh) equivalence for $\mathcal{L}$, is to decide for given $L \in \mathcal{L}$ and $h_1, h_2 \in \mathcal{H}$ whether $h_1 \overset{L}{\equiv} h_2$ ($h_1 \overset{L}{\equiv}_\ell h_2$, $h_1 \overset{L}{\equiv}_p h_2$) . If $\mathcal{H}$ is the family of all homomorphisms we speak simply about homomorphism equivalence for $\mathcal{L}$.

Lemma 1.1 Let $\mathcal{L}$ be an arbitrary family of languages. The decidability of homomorphism equivalence for $\mathcal{L}$ implies the decidability of homomorphism length (Parikh) equivalence for $\mathcal{L}$. The problems of homomorphic length equivalence for $\mathcal{L}$ and homomorphic Parikh equivalence for $\mathcal{L}$ are equivalent.

Proof Let $L \in \mathcal{L}$. To reduce $h_1 \overset{L}{\equiv}_\ell h_2$ to $h_1 \overset{L}{\equiv} h_2$ modify $h_1, h_2 : \Sigma^* \to \Delta^*$ to $h_1', h_2' : \Sigma^* \to \{a\}^*$ by replacing each symbol in Δ by a . Clearly, $h_1 \overset{L}{\equiv}_\ell h_2$ iff $h_1' \overset{L}{\equiv} h_2'$.

To prove the second statement in one direction we do the same as above; in the reverse direction we erase all symbols in Δ but one. In this way we reduce the testing of Parikh equivalence to n tests of length equivalence. □

Theorem 1.1 The problem of homomorphism (length, Parikh) equivalence for the family of regular sets is decidable.

Proof For homomorphism equivalence the result is proved in [6, 7]. The other two results follow by Lemma 1.1.

Definition [12] For two homomorphisms $h_1, h_2 : \Sigma^* \to \Delta^*$ the equality set is denoted by $E(h_1, h_2)$ and defined as
$E(h_1, h_2) = \{w \in \Sigma^* : h_1(w) = h_2(w)\}$. Note that $h_1 \overset{L}{\equiv} h_2$ iff $L \subseteq E(h_1, h_2)$.

2. LENGTH AND PARIKH EQUIVALENCE OF HOMOMORPHISMS ON ETOL LANGUAGES

Our goal in this section is to show that the problem of homomorphic length (Parikh) equivalence for ETOL languages is decidable. We proof this result first for DTOL languages, then for TOL languages and finally we extend it to ETOL languages by expressing an ETOL language as a coding of a TOL language.

Definition Let $G = (\Sigma, h_1, \ldots, h_n, \sigma)$ and $G' = (\Sigma, h_1', \ldots, h_n', \sigma')$ be two DTOL systems. Let h, h' be homomorphisms $\Sigma^* \to \Delta^*$. We say that HDTOL systems (G, h) and (G', h') are growth sequence equivalent if for any $m \geq 0$, $i_1, \ldots, i_m$, $1 \leq i_j \leq n$ we have

$$|h(h_{i_1}(h_{i_2}(\ldots(\sigma)\ldots)))| = |h'(h_{i_1}'(h_{i_2}'(\ldots(\sigma)\ldots)))| .$$

Lemma 2.1 The growth sequence equivalence problem for DTOL systems is decidable.

Proof [13, Theorem 8.1].

Lemma 2.2 The problem of homomorphism length equivalence for DTOL languages is decidable.

Proof Let $L = L(G)$ where $G = (\Sigma, h_1, \ldots, h_n, \sigma)$ is a DTOL system and let h_1, h_2 be homomorphisms. We consider two HDTOL systems (G, h_1) and (G, h_2). Clearly, $h_1 \overset{L}{\equiv}_{\ell} h_2$ iff the HDTOL systems (G, h_1) and (G, h_2) are growth sequence equivalent, which is decidable by Lemma 2.1.

Definition With every TOL system $G = (\Sigma, \mathcal{P}, \sigma)$ we associate a DTOL system $G' = (\Sigma, \mathcal{P}', \sigma)$ as follows:

$$\mathcal{P}' = \{P' : P' \subseteq P ,\ P \in \mathcal{P} \text{ and } P' \text{ is deterministic}\} .$$

For example if $\mathcal{P} = (\{a \to a, a \to b, b \to a, b \to b\})$ we have $\mathcal{P}' = (\{a \to a, b \to a\}, \{a \to a, b \to b\}, \{a \to b, b \to a\}, \{a \to b, b \to b\})$. Note that if the order of tables is unessential we view them as a set rather than a sequence.

Lemma 2.3 Let G be a TOL system and G' its associated DTOL system, and let h_1, h_2 be arbitrary homomorphisms. Then $h_1 \overset{L}{\equiv}_\ell h_2$ iff $h_1 \overset{L'}{\equiv}_\ell h_2$, where $L(G) = L$ and $L(G') = L'$.

Proof Using induction on the number of "splits" of tables it is clear that we need only to show the following. If $G_1 = (\Sigma, P \cup \{P_0\}, \sigma)$ is a TOL system and $G_2 = (\Sigma , P \cup \{P_1,P_2\}, \sigma)$ where for some $s \geq 2$, $P_0 = P' \cup \{a \to w_j : 1 \leq j \leq s\}$, $P_1 = P' \cup \{a \to w_1\}$, $P_2 = P' \cup \{a \to w_j : 2 \leq j \leq s\}$, then $h_1 \overset{L_1}{\equiv}_\ell h_2$ iff $h_1 \overset{L_2}{\equiv}_\ell h_2$ where $L_i = L(G_i)$, $i = 1,2$. Since $P_i \subseteq P_0$ for $i = 1,2$ we have $L_2 \subseteq L_1$, and so if $h_1 \overset{L_1}{\equiv}_\ell h_2$, then $h_1 \overset{L_2}{\equiv}_\ell h_2$. Suppose that the converse is false, i.e. $h_1 \overset{L_2}{\equiv}_\ell h_2$ and there exists w in L_1 so that $|h_1(w)| \neq |h_2(w)|$. Clearly we do not change $L(G_i)$ if we add to P the tables P_1, P_2 . So assume that $G_i = (\Sigma, \hat{P}_i, \sigma)$ for $i = 1,2$ where if $P = (P_3, P_4, \ldots, P_t)$, $t \geq 2$, then $\hat{P}_1 = (P_0,P_1,P_2,\ldots,P_t)$ and $\hat{P}_2 = (P_1,P_2,P_3,\ldots,P_t)$. We write $u \xrightarrow[i_1 \ldots i_n]{G_i} v$ if u derives v by the application of tables $P_{i_1},\ldots,P_{i_n}$ in that order. We make a few observations. If $u \in (\Sigma - \{a\})^*$, then $u \underset{P_0}{\overset{G_1}{\Rightarrow}} v$ is equivalent to $u \underset{P_1}{\overset{G_1}{\Rightarrow}} v$ or $u \underset{P_2}{\overset{G_2}{\Rightarrow}} v$, moreover for $i_1,\ldots,i_n$ such that $i_j \neq 0$ for $1 \leq j \leq n$ we have $u \xrightarrow[i_1 \ldots i_n]{G_1} v$ iff $u \xrightarrow[i_1 \ldots i_n]{G_2} v$. Finally, letter a is useful in G_1 iff it is useful in G_2 .

Now if there is w in L such that $|h_1(w)| \neq |h_2(w)|$, then we may assume without loss of generality that w is chosen so that in $\sigma \overset{G_1}{\underset{i_1\ldots i_n}{\Longrightarrow}} w$, n is minimal, and further the largest p such that $i_p = 0$ is minimal (for selected n). If such p does not exist, then $i_j \neq 0$ for $1 \le j \le n$ and $\sigma \overset{G_2}{\Longrightarrow} w$, thus $|h_1(w)| = |h_2(w)|$ a contradiction.

So we have $\sigma \overset{G_1}{\underset{i_1\ldots i_{p-1}}{\Longrightarrow}} Z \overset{G_1}{\underset{i_p\ldots i_n}{\Longrightarrow}} w$ with $i_p = 0$. Let

$Z = u_1 a \ldots u_\ell a u_{\ell+1}$, where $u_i \in (\Sigma - \{a\})^*$. We can write w in the form $w = x_1 y_1 \ldots x_\ell y_\ell x_{\ell+1}$ so that $u_k \overset{G_2}{\underset{1 i_{p+1}\ldots i_n}{\Longrightarrow}} x_k$ (also $u_k \overset{G_2}{\underset{2 i_{p+1}\ldots i_n}{\Longrightarrow}} x_k$) and $a \overset{G_2}{\underset{r_k i_{p+1}\ldots i_n}{\Longrightarrow}} y_k$ where $r_k = 1$ if the production $a \to w_1$ was used at the p-th step to replace the k-th (from left) occurrence of an otherwise $r_k = 2$.

Since a is useful in G_1 , it is useful in G_2 , so there exist $s_1,\ldots,s_m$, $m \ge 0$, $1 \le s_j \le t$ for $j = 1,\ldots,m$, $q > 0$, $\alpha_j \in (\Sigma - \{a\})^*$ for $1 \le j \le q$, such that

$\sigma \overset{G_2}{\underset{s_1\ldots s_m}{\Longrightarrow}} \alpha_1 a \alpha_2 \ldots \alpha_q a \alpha_{q+1}$. Let $\beta_j \in \Sigma^*$ for $j = 1,\ldots,q$ such that $\alpha_j \overset{G_2}{\underset{1 i_{p+1}\ldots i_n}{\Longrightarrow}} \beta_j$ and therefore also $\alpha_j \overset{G_2}{\underset{2 i_{p+1}\ldots i_n}{\Longrightarrow}} \beta_j$. For each $k = 1,\ldots,\ell$ we have $\sigma \overset{G_2}{\underset{s_1\ldots s_m r_k i_{p+1}\ldots i_n}{\Longrightarrow}} \beta_1 y_k \beta_2 \ldots \beta_\ell y_k \beta_{\ell+1}$, so since $h_1 \overset{L_2}{\equiv_\ell} h_2$, we have

$$|h_1(y_k)| - |h_2(y_k)| = 1/q\ (|h_2(\beta_1\ldots\beta_{\ell+1})| - |h_1(\beta_1\ldots\beta_{\ell+1})|)$$

which is a constant independent of k . Now let

$\gamma = x_1 y_1 x_2 y_1 \ldots x_\ell y_1 x_{\ell+1}$, since $|h_1(w)| - |h_2(w)|$ does not change when we replace y_i by y_1 and since $|h_1(w)| \neq |h_2(w)|$ we have $|h_1(\gamma)| \neq |h_2(\gamma)|$. However, this contradicts the choice of w since $\sigma \xrightarrow[i_1 \ldots i_{p-1} r_1 i_{p+1} \ldots i_n]{G_1} \gamma$. Hence, we conclude that $h_1 \overset{L_2}{\equiv_\ell} h_2$ implies $h_1 \overset{L_1}{\equiv_\ell} h_2$. □

<u>Theorem 2.1</u> The problem of homomorphism length equivalence for ETOL languages is decidable, i.e. given an ETOL language L and arbitrary homomorphisms h_1, h_2 it is decidable whether $|h_1(w)| = |h_2(w)|$ for all $w \in L$.

<u>Proof</u> Let $L \subseteq \Sigma^*$ be an ETOL language. If $L = \phi$, then $h_1 \overset{L}{\equiv_\ell} h_2$, otherwise by [9] there exists a letter-to-letter homomorphism h (coding) and a TOL language L' (both effectively) so that $L = h(L')$. We have $h_1 \overset{L}{\equiv_\ell} h_2$ iff $h_1' \overset{L'}{\equiv_\ell} h_2'$ where h_i' is the homomorphism defined by $h_i'(w) = h_i(h(w))$ for each $w \in \Sigma^*$. So the proof is completed by Lemma 2.3. □

<u>Corollary</u> The problem of homomorphism Parikh equivalence for ETOL languages is decidable.

<u>Proof</u> By Lemma 1.1. □

3. $\mathcal{H}$-SMOOTH FAMILIES AND HOMOMORPHISM EQUIVALENCE FOR ETOL LANGUAGES OVER A BINARY ALPHABET

The main goal of this section is to show that the problem of homomorphism equivalence for ETOL languages over a binary alphabet is decidable.

Definition Given a family of homomorphisms $\mathcal{H}$, we say that a family of languages $\mathcal{L}$ (effectively given) is $\mathcal{H}$-smooth if:

(i) The emptiness problem for $\mathcal{L}$ is decidable.

(ii) Family $\mathcal{L}$ is effectively closed under the intersection with a regular set.

(iii) For all $h_1, h_2 \in \mathcal{H}$, $L \in \mathcal{L}$; if $h_1 \overset{L}{\equiv} h_2$, then there is a regular set R so that $L \subseteq R \subseteq E(h_1, h_2)$.

We mention a result which will not be used here but might be useful for other applications. It was essentially proved in [6] in more details then in [7] and [12].

Theorem 3.1 The condition (iii) in the above definition is equivalent to condition:

(iv) For all $h_1, h_2 \in \mathcal{H}$, $L \in \mathcal{L}$; if $h_1 \overset{L}{\equiv} h_2$, then the pair (h_1, h_2) has bounded balance on L , that is $\| h_1(x)| - |h_2(x)\|$ is bounded for all prefixes x of words of L .

Now, we show a result which was already proved but not formulated in the general form in [6]. In somewhat weaker form it was explicitly stated in [7] and [12].

<u>Theorem 3.2</u> Let L be a family of languages and H a family of homomorphisms. If L is H-smooth, then the problem of H equivalence for L is decidable.

<u>Proof</u> We give two semiprocedures, one testing the equivalence, the other the nonequivalence. Their correctness follows from Theorem 1.1 and the definition of H-smoothness.

For the equivalence, if $L \subseteq \Sigma^*$, effectively enumerate all the regular sets $R \subseteq \Sigma^*$, if such R is found that $L \subseteq R \subseteq E(h_1, h_2)$, then stop.

For the nonequivalence, effectively enumerate all x in L, which is possible since properties (i) and (ii) of H-smoothness imply that the membership problem for L is decidable. If x in L is found such that $h_1(x) \neq h_2(x)$, then stop.

It was shown in [8] that the equality set for elementary homomorphisms is always regular.

<u>Definition</u> A homomorphism h $\Sigma^* \to \Gamma^*$ is <u>simplifiable</u> if there exists an alphabet Δ, with cardinality smaller than Γ, and two homomorphisms $f : \Sigma^* \to \Delta^*$ and $g : \Delta^* \to \Gamma^*$, such that $h = gf$, i.e. $h(x) = g(f(x))$ for each $x \in \Sigma^*$. If h is not simplifiable, then h is called <u>elementary</u>. Let E be the family of elementary homomorphisms.

<u>Theorem 3.3</u> If $h_1, h_2 \in E$, then $E(h_1, h_2)$ is a regular set.

<u>Proof</u> [8].

From Theorem 3.2 and Theorem 3.3 it immediately follows that the equivalence of two elementary homomorphisms on a DOL sequence (or language which is the same) is decidable a result which was reproved in [8]. Actually, we can formulate the following general result essentially also given in [11].

Theorem 3.4 Let L be a family of (effectively given) languages such that

(i) The emptiness problem for L is decidable.

(ii) L is effectively closed under intersection with a regular set.

Then the problem of E equivalence for L is decidable.

Proof By the conditions above and Theorem 3.3 L is E-smooth. Hence the problem is decidable by Theorem 3.2. □

In [1] the index languages are defined and it is shown they satisfy the assumptions of Theorem 3.4. Similarly for uc-programmed languages defined by unconditional transfer programmed grammars in [10]. So we have the following

Corollary 3.1 The problem of E equivalence for the families of indexed and uc-programmed languages is decidable.

Now we are prepared for the one main result.

Theorem 3.5 Let L be a family of languages (effectively given) over a binary alphabet such that

(i) The emptiness problem for $\mathcal{L}$ is decidable.

(ii) $\mathcal{L}$ is effectively closed under intersection with a regular set.

(iii) The problem of homomorphic length equivalence for $\mathcal{L}$ is decidable.

Then the problem of (arbitrary) homomorphism equivalence for $\mathcal{L}$ is decidable.

<u>Proof</u> Consider $L \in \mathcal{L}$, $L \subseteq \Sigma^*$ and homomorphisms $h_1, h_2 : \Sigma^* \to \Delta^*$. Let $\Sigma = \{a_1, a_2\}$. First test whether h_1 and h_2 are elementary which, clearly, is decidable. If both h_1 and h_2 are elementary, then $h_1 \overset{L}{\equiv} h_2$ is decidable by Theorem 3.4. Otherwise assume that h_1 is simplifiable. Since Σ has only two symbols that implies there is α in Δ^* and $n_i \geq 0$ such that $h_1(a_i) = \alpha^{n_i}$ for $i = 1,2$. Hence $h_1 \overset{L}{\equiv} h_2$ iff $L \subseteq h_2^{-1}(\alpha^*)$ and $h_1 \overset{L}{\equiv}_{\ell} h_2$. Indeed, if $h_1 \overset{L}{\equiv} h_2$, then $h_1 \overset{L}{\equiv}_{\ell} h_2$ and $h_2(L) = h_1(L)$, hence $h_2(L) \subseteq \alpha^*$ and $L \subseteq h_2^{-1}(\alpha^*)$. Conversely, if $L \subseteq h_2^{-1}(\alpha^*)$, then $h_2(L) \subseteq \alpha^*$, so for $x \in L$ we have $h_1(x) \in \alpha^*$ and $h_2(x) \in \alpha^*$, hence $h_1(x) = h_2(x)$ iff $|h_1(x)| = |h_2(x)|$.

By assumption (i) and (ii) we can test whether $L \cap R = \phi$ for R being the complement of regular set $h_2^{-1}(\alpha^*)$. By (iii) we can test whether $h_1 \overset{L}{\equiv}_{\ell} h_2$. □

<u>Corollary 3.2</u> The problem of (arbitrary) homomorphism equivalence for the family of ETOL over a binary alphabet is decidable.

Proof The family of ETOL languages satisfies the assumptions of Theorem 3.5, the property (iii) was proved in Theorem 2.1. □

References

1. A. V. Aho, Indexed Grammars, An Extension of Context-free Grammar. JACM 15 (1968), 647-671.

2. K. Culik II, On the Decidability of the Sequence Equivalence Problem for DOL-systems. Theoretical Computer Science 3 (1977), 75-84.

3. K. Culik II, The Ultimate Equivalence Problem for DOL-systems, Acta Informatica, to appear.

4. K. Culik II, A Purely Homomorphic Characterization of Recursively Enumerable Sets, Dept. of Computer Science, University of Waterloo Research Report CS-78-13 (1978).

5. K. Culik II, Some Decidability Results about Regular and Push Down Translations, Dept. of Computer Science, University of Waterloo Research Report CS-78-09 (1978).

6. K. Culik II and I. Fris, The Decidability of the Equivalence Problem for DOL-systems. Information and Control 35 (1977), 20-39.

7. K. Culik II and A. Salomaa, On the Decidability of Homomorphism Equivalence for Languages. J. Comput. System Sci., to appear.

8. A. Ehrenfeucht and G. Rozenberg, Elementary Homomorphisms and a Solution of the DOL Sequence Equivalence Problem, Theoretical Computer Science, to appear.

9. A. Ehrenfeucht and G. Rozenberg, Nonterminals Versus Homomorphisms in Defining Languages for Some Classes of Rewriting Systems, Acta Informatica 3 (1974), 265-283.

10. D. J. Rozenkrantz, Programmed Grammars and Classes of Formal Languages, JACM 16 (1969), 107-131.

11. A. Salomaa, Formal Languages. Academic Press, New York (1973).

12. A. Salomaa, Equality Sets for Homomorphisms of Free Monoids, Manuscript 1978.

13. A. Salomaa and M. Soittola, Automata Theoretic Aspects of Formal Power Series, Springer Verlag, to appear.